

We gratefully acknowledge our sponsors:

SIEMENS Healthineers

A FRIEND OF THE CT MEETING

GE HealthCare

Canon
CANON MEDICAL RESEARCH USA, INC.

PHILIPS

VAREX IMAGING

Welcoming address

Dear Colleagues,

It was a pleasure to welcome the attendees in Salt Lake City for the ninth edition of *The International Conference on Image Formation in X-ray Computed Tomography*, also known as "The CT Meeting".

The idea for this meeting came in 2010, following a wish from many of us to have a venue where CT scientists could meet to discuss in depth all aspects that impact the image formation process in CT. These aspects include dose evaluation and reduction strategies, image quality assessment, non-linearity effects and compensation schemes for these effects, image reconstruction algorithms, spectral decomposition, dynamic effects, geometrical calibration, phase-contrast CT. They further include the utilization of photon-counting detectors, which have now become commercially available.

Pleasantly, the meeting this year remained very strong, as it attracted more than 100 submissions as well as more than 150 attendees. The breadth of the topics being covered was amazing. We are grateful to all authors for submitting their work. Similarly to the previous meetings, each oral presentation was allocated a significant block of time to allow for in-depth material description and related discussion. Also, two comfortable poster sessions were planned, and attractive locations were selected to promote discussion during the meals.

The meeting could not have been a success without a great scientific committee. We would like to take the opportunity here to once again thank the following scientists for their support, as well as for their help with the evaluation of submitted abstracts, which was critical in shaping the scientific program:

Adam Wang	Stanford University, USA
Andreas Maier	Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
Bruno De Man	GE Healthcare, USA
Charlotte Hagen	University College London, United Kingdom
Craig Abbey	University of California in Santa Barbara, USA
Cristian Badea	Duke University, USA
Cyril Riddell	GE HealthCare, France
Emil Sidky	University of Chicago, USA
Ge Wang	Rensselaer Polytechnic Institute, USA
Grace Gang	University of Pennsylvania, USA
Guang-Hong Chen	University of Wisconsin-Madison, USA
Hao Zhang	Stanford University, USA
Hengyong Yu	University of Massachusetts Lowell, USA
Hiroyuki Kudo	University of Tsukuba, Japan
Hui Hu	Nautilus X-ray, USA
Irene Hernandez-Giron	University College Dublin, Ireland
Jeff Siewerdsen	The University of Texas MD Anderson Cancer Center, USA

Jingyan Xu	Johns Hopkins University, USA
Jing Wang	University of Texas Southwestern Medical Center, USA
John Marian Hoffman	University of California in Los Angeles, USA
Jongduk Baek	Yonsei University, South Korea
Josh Star-Lack	Varex Imaging Corporation, USA
Karl Stierstorfer	Siemens Healthcare AG, Germany
Katia Parodi	Ludwig-Maximilians-Universität München, Germany
Ke Li	The University of Texas MD Anderson Cancer Center, USA
Ken Taguchi	Johns Hopkins University, USA
Kirsten Boedeker	Canon Medical Systems Corporation, USA
Koen Michielsen	Radboud University Medical Center, The Netherlands
Kyoung-Yong Lee	Samsung Electronics, South Korea
Jerome Lesaint	ESRF – The European Synchrotron Radiation Facility, France
Lifeng Yu	Mayo Clinic, USA
Ludwig Ritschl	Siemens Healthineers, Germany
Maria Magnusson	Linköping University
Marc Kachelrieß	German Cancer Research Center, Germany
Mats Persson	KTH Royal Institute of Technology, Sweden
Monica Abella	Universidad Carlos III de Madrid, Spain
Nicole Maass	Siemens Healthineers, Germany
Pascal Paysan	Varian Medical Systems Imaging Lab, Switzerland
Paul Segars	Duke University, USA
Paulo R. Costa	University of São Paulo, Brazil
Peter Noël	University of Pennsylvania, USA
Rongping Zeng	Food and Drug Administration, USA
Scott Hsieh	Mayo Clinic, USA
Seungryong Cho	Korea Advanced Institute of Science and Technology, South Korea
Simon Rit	CNRS, France
Srinivasan Vedantham	University of Arizona, USA
Stanislav Zabic	Philips, USA
Thomas Köhler	Philips GmbH Innovative Technologies, Hamburg, Germany Johns
Web Stayman	Hopkins University, USA
Xiangyang Tang	Emory University, USA
Xiaochuan Pan	University of Chicago, USA
Yuxiang Xing	Tsinghua University, China
Zhicong Yu	University of California in Santa Cruz, USA
Zhou Yu	Canon Medical Research USA, Inc.

The scientific program was further enhanced by two refresher courses. The first course, entitled “Photon-counting CT: physics, clinical use and future directions” was graciously and jointly offered by Dr. Scott Hsieh from the Mayo Clinic and Dr. Guang-Hong Chen from the University of Wisconsin-Madison. The second course entitled “Diffusion models and applications” was graciously offered by Dr. Jingyan Xu from Johns Hopkins University.

For the first time, the meeting was preceded by a 2-hour “Workshop on CT-omics” chaired by Dr. Jerome Liang from Stony Brook University. A summary of this workshop is provided hereafter, and we look forward to future instances.

Finally, as mentioned on the cover, we were fortunate to receive generous support from

- ✓ Siemens Healthineers
- ✓ A Friend of the CT Meeting
- ✓ GE Healthcare
- ✓ Philips
- ✓ Canon Medical Research USA, Inc
- ✓ Varex Imaging Corporation, USA

Their financial support was essential to accommodate a lower registration fee for graduate students, as well as to keep the registration fee affordable for all others. This year again, more than one third of the attendees were PhD students who will shape the future of CT image formation.

We hope that all readers of these proceedings will find them inspiring!

We also look forward seeing you all at the next edition of the CT meeting which will take place in scenic Oberstdorf (Germany) on 10 July 2028, with updates available on the conference website https://ct-meeting.org.

Frédéric Noo, PhD
Conference Chair
Department of Radiology and Imaging Sciences
University of Utah

Larry Zeng, PhD
Conference Co-Chair
Department of Computer Science
Utah Valley University

Workship on CT-omics

Decoding *In-Vivo* Tissue Biology for Diagnosis of Tissue Pathology and Prediction of Tissue Healthy IQ

Organizer: Jerome Liang, Stony Brook University

Moderator: Larry Zeng, Utah Valley University

CT is a non-invasive medical imaging modality that captures how *in-vivo* tissues respond to X-ray energies. These responses are encoded in CT image contrasts and contain biologically relevant information about tissue conditions. Decoding this biological information has great potential to characterize tissues' pathology and health conditions. This panel discussion focuses on the following interpretations of these responses:

The interactions of *in-vivo* tissues with X-ray energies are complex. The responses captured by CT only reflect part of these interactions. The captured responses are quantitatively represented as linear attenuation coefficients through the X-ray energy field. The coefficient at a point in the X-ray energy field is tissue-specific and energy-dependent.

Intuitively, the responses should ideally be sampled at each point over both time and energy to provide a more complete description of tissues' biological characteristics, such as pathology and health conditions.

Tissues' pathology and health (including aging) conditions are mainly characterized by tissues' functionalities or the underlying living cells' activities. These activities are encoded into the variations of attenuation coefficients at a point and the neighbors in the tissue space. Different groups of living cells can alter the attenuation coefficients within an image voxel and among neighboring voxels in many ways (or at different tissue conditions) and, therefore, change the spatial patterns of the attenuation coefficients at different times and energies.

Based on the above interpretations, hypotheses and methods for testing them were proposed for discussion.

(1). CT acquires (energy integrated) data from the higher energy portion of the full X-ray energy spectrum, which is measurable but incomplete data. The data from the lower energy portion of the spectrum is not measurable but can be estimated by incorporation of prior knowledge about X-ray Physics and Tissue Biology into consideration of the acquired measurable data. Full energy spectrum (complete) data or tissues monoenergetic images (TMIs) provides the necessary but not sufficient warrant information to achieve CT-based tissue characterization tasks. AI/Machine Learning (ML) using full energy spectrum TMIs would increase or maximize the CT diagnostic and prediction power.

(2). While both the time and energy samplings are lacking in CT data acquisition, the spatial sampling across continuous tissues encodes cells' activities into the variation of spatial patterns of TMIs in terms of Hounsfield Units. The variation of spatial patterns can be described by derivatives of TMIs and used to characterize *in-vivo* tissues.

In short, clinical data experiments show the feasibility of CT-omics. In other words, leveraging prior knowledge about how CT images are made in AI-ML architecture design will help interpret the changing of attenuation coefficients at different times and energies and, therefore, maximize CT diagnostic and prediction power.

Non-invasive measurements are frequently incomplete to describe *in-vivo* tissues. While more descriptive “data” may not be measurable but can provide deep information for AI-ML observers to interpret the underlying *in-vivo* tissues. Only experimental outcomes will validate the effectiveness of the AI-ML observers.

The conference records are arranged in the same order as they were scheduled for presentation as shown below.

	Novel CT imaging concepts	**Session chairs: M. Kachelrieß, H. Zhang**
Page #	**Oral session M1** **Monday 8:10 – 9:50**	**Authors**
20	Multisource C-arm CT : Initial Results with a 2 x 2 Carbon Nanotube Source Array	Hojin Jung (Yonsei University); Jongduk Baek (Yonsei University)*
25	One-Dimensional Anti-Scatter Grids for Static CT	Andreas Heinkele (German Cancer Research Center)*; Julien Erath (Siemens Healthineers AG); Lukas Hennemann (German Cancer Research Center); Johan Sunnegaardh, Christian Hofmann, Martin Petersilka, Karl Stierstorfer (Siemens Healthineers AG); Marc Kachelrieß (German Cancer Research Center)
29	Foveated-imaging geometry CT architecture and prior-guided diffusion model enabling global super-resolution image reconstruction	Wenxin Mo (Tsinghua University)*; Yinxian Xia, Hao Zhou, Hewei Gao (Tsinghua University)
33	PINN-Driven Scatter Correction for Multi-source Linear-array CT	Yankun Wang (Southern Medical University)*; Yule Li, Junhao Wu, Yuanxiang Gao, Haomin Hou, Deyang Kong (Southern Medical University); Limin Li (Xi'an Jiaotong University); Jingyi Liao, Hao Wang (Southern Medica University); Jianhua Ma (Xi'an Jiaotong University); Dong Zeng (Southern Medical University)
38	Upright Cone-Beam Breast CT: Towards Breast Cancer Screening using Short-Scan and Sparse-View Acquisitions	Hsin Wu Tseng (University of Arizona)*; Jing-Tzyh Chiang, Srinivasan Vedantham (University of Arizona)
	More on this in poster session I **Monday 13:30 – 15:30**	**Authors**
42	Analytical Weighted Reconstruction for Longitudinal Multiple X-ray Source Array CT (MXA-CT)	Jiaming Liu (Johns Hopkins University)*; Gengxin Shi (Johns Hopkins University); Andrew M. Hernandez (University of California Davis); Alejandro Sisniega (Sedecal Corporation, Madrid); Richard E. Colbeth, Vance Robinson (Varex Imaging Corp.); Jeffrey H. Siewerdsen (The University of Texas MD Anderson Cancer Cancer); Craig K. Abbey (University of California, Santa Barbara); Paul Schwoebel (University of New Mexico); John M. Boone (University of California Davis); Wojciech Zbijewski (Johns Hopkins University)

47	Bayesian regulators to promote Sharp Image Edges in Limited Angle Tomography	Larry Zeng (Utah Valley University)*
51	Breaking the G-Force Barrier in CT: A Sixth-Generation Architecture Using a Stationary Magnetically Steered X-ray Source	Hui Hu (Nautilus Xray, Inc.)*; Vitaliy Ziskin, Andrew Cross, Hong Chen (Nautilus Xray, Inc.)
55	A Static Dual-Ring CT System with Integrated Cold-Cathode X-Ray Source and Detector Arrays: A Simulation Study	Bonghong Ye (Southern Medical University)*; Yule Li, Yankun Wang, Hao Wang, Jingyi Liao (Southern Medical University); Limin Li (Xi'an Jiaotong University); Zhaoying Bian (Southern Medical University); Zhipeng Zhang (Sun Yat-sen University); Jiahua Ma (Xi'an Jiaotong University); Dong Zeng (Southern Medical University)
59	New directions for old detectors	Scott Hsieh (Mayo Clinic)*
63	Practical Geometric Decoupling for Projection Unmixing in Multi-Source Linear-Array CT	Qiule Chen (Southern Medical University)*; Yule Li, Junhao Wu, Yankun Wang, Jingyi Liao, Yuanxiang Gao (Southern Medical University); Jianhua Ma (Xi'an Jiaotong University); Hao Wang, Dong Zeng (Southern Medical University)

	Top Four Abstracts	**Session chairs: E. Sidky, A. Wang**
Page #	**Oral session M2** **Monday 10:10 – 11:30**	**Authors**
67	Understanding Radiologist Performance and Missed Lesion Detection	Parmita Mondal (Mayo Clinic)*; Akitoshi Inoue, Rickey Carter, Andrew D Missert, Joel G. Fletcher, Hao Gong, Shuai Leng, Lifeng Yu, Cynthia H. McCollough, Scott S. Hsieh (Mayo Clinic)
71	A Theoretical Prediction of CT-based Radiomics Feature Variance to Improve the Performance Classification Models	Huay Din (University of Pennsylvania)*; Yijie Yuan (Johns Hopkins University); Grace Hyun Kim, Michael McNitt-Gray (UCLA); J. Webster Stayman, Grace J. Gang (Johns Hopkins University)
77	Focal Spot Design Optimization for High-Resolution CT	Yue Fan (Johns Hopkins University)*; Xiao Jiang (Johns Hopkins University); Grace Gang (University of Pennsylvania); Joseph Stayman (Johns Hopkins University)
81	Scatter Characteristic and LBTE-based Correction in Multi-source Dual-ring Stationary CT	Guoxi Zhu (Tsinghua University)*; Hewei Gao, Li Zhang, Yuxiang Xing, Changyu Chen, Zhiqiang Chen (Tsinghua University)

	Handling non-ideal conditions (incomplete data, metal, motion, rings, scatter)	**Session chairs: K. Stierstorfer, S. Zabic**
Page #	**Oral session M3** **Monday 16:00 – 17:40**	**Authors**
86	Measurement-based Deep Scatter Estimation	Lukas Hennemann (German Cancer Research Center)*; Julien Erath (Siemens Healthineers AG); Andreas Heinkele (German Cancer Research Center); Eric Fournié, Martin Petersilka, Karl Stierstorfer (Siemens Healthineers AG); Marc Kachelrieß (German Cancer Research Center)
91	Preliminary investigation of motion-artifact suppression in image-guided radiation therapy of prostate cancer	Zheng Zhang (The University of Chicago)*; Bulent Aydogan, Erik Pearson, Dan Xia, Emil Sidky, Xiaochuan Pan (The University of Chicago)
95	Dynamic CT Chest Respiratory Motion Phantom with Lifelike 3D-Printed Deformable Lungs	Jessica Im (University of Pennsylvania)*; Michael Geagan (University of Pennsylvania); Sebastian Meyer (Memorial Sloan Kettering Cancer Center); Kai Mei, Peter Noel (University of Pennsylvania)
100	Tunable Metal-Artifact CT Phantoms via Dual-Filament 3D Printing	Pouyan Pasyar (University of Pennsylvania)*; Kai Mei, Jessica Im, Leonid Roshkovan, Michael Geagan, Peter Noel (University of Pennsylvania)
104	A Site-Agnostic Dual-Domain Deep Learning Reconstruction Framework for Nonstop Gated CBCT	Mitchell Yu, Noah Silverberg, Sean Berry, Wendy Harris, Jean Moran (Memorial Sloan Kettering Cancer Center); Lei Xing (Stanford University); Hao Zhang (Stanford University)*
	More on this in poster session I **Monday 13:30 – 15:30**	**Authors**
108	Differentiable Forward and Back Projectors for Gradient-Based Deformable Motion Estimation	Xiao Jiang (Johns Hopkins University)*; Grace Gang (University of Pennsylvania); Joseph Stayman (Johns Hopkins University)
113	Segmentation-free automatic geometry calibration method using a neural-network prediction model in CBCT	Sungho Yun (Korea Advanced Institute of Science and Technology)*, Seungryong Cho (Korea Advanced Institute of Science and Technology)
117	Image Reconstruction for Helical CBCT With a Laterally Shifted Flat-Panel Detector	Zhicong Yu (University of California, Santa Cruz)*, Amit Jain, Chuanyong Bai (formerly with Accuray Inc.), Jacob Shea (Accuray Inc.), Daniel Cagnon (formerly with Accuray Inc.)
121	Normalized Metal Artifact Reduction using Physics-Inspired Deep Prior Images	Anton Kabelac (German Cancer Research Center)*; Marc Kachelrieß (German Cancer Research Center); Andreas Specovius (Siemens Healthineers)

126	Self-operating Geometry Calibration for Cone-beam Computed Tomography using Bayesian Optimization	Jiyong Shim (Yonsei University)*; Hyosung Cho (Yonsei University)
131	Auxiliary Orbit Compensation for Cone-Beam Artifacts in CBCT	Zeyu Zhou (Carl Zeiss X-ray Microscropy, Inc)*; Andriy Andreyev, Faguo Yang (Carl Zeiss X-ray Microscropy, Inc)
135	Comparison of Local Regression-Based Method for Removing Ring Artifacts in CT Images	Xiangyuan Liu (Southern University of Science and Tech)*; Xudong Ru (Beijing Normal University); Shusen Zhao (Southern University of Science and Tech)
140	DOCTOR: Physics-Constrained Flow Matching Method for Cone-Beam CT Reconstruction Under a Variety of Non-Ideal Imaging Conditions	Qiang Tang, Li Chen, Haoxiang Hu (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences); Jun Xiang (X-Ray Department, United Imaging Healthcare Limited Company); Dong Liang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences); Yinsheng Li (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)*
144	LAB-PhD: A Learnable Physics-Informed Diffusion Model for Spectral Cone Beam CT Imaging	Li Chen, Ruifeng Chen (Southeast University); Qiang Tang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, and University of Chinese Academy of Sciences); Juan Feng, Jun Xiang (United Imaging Healthcare LC); Xu Ji, Yang Chen (Southeast University); Dong Liang (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Yinsheng Li (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)*

	Photon-counting CT	**Session chairs: J. Star-Lack, M. Persson**
Page #	**Oral session T1** **Tuesday 8:50 – 10:30**	**Authors**
148	Dual-Contrast K-Edge Imaging on a Clinical PCCT: Influence of K-Edge Spacing	Martin Rybertt (University of Pennsylvania)*; Katie Villasenor, Nicole Mirzaian, Ryan Fair (University of Pennsylvania); Pooyan Sahbaee, Tristan Nowak (Siemens Healthineers); Leening Liu, Ali Dhanaliwala, David Cormode, Peter Noël (University of Pennsylvania)
152	Enable K-Edge Material Decomposition Without K-Edge Material Calibration: A Feasibility Study	Yirong Yang (United Imaging Healthcare North America)*; Wenying Wang (United Imaging Healthcare North America)
157	A Digital Twin of Photon-Counting CT System	Ruiran Lai (University of Wisconsin-Madison)*; Ran Zhang, Linying Zhan (University of Wisconsin-Madison); Roy Nilsen, Tyler Curtis (GE Healthcare); Xiaopeng Yu (University of Wisconsin-Madison); Brian Yanoff, Bruno De Man (GE Healthcare); Ke Li (University of Texas MD Anderson Cancer Center); Guang-Hong Chen (University of Wisconsin-Madison)

162	Combining Poisson random variables to generate and analyze the statistics of photon counting detectors	Karl Stierstorfer (Siemens Healthineers)*
166	Spectral performance comparison of counting schemes for photon counting x-ray detectors with anti-scatter grids	Katsuyuki Taguchi (Johns Hopkins University)*; Patrick Riehl (Analog Devices); Scott Hsieh (Mayo Clinic)
	More on this in poster session I **Monday 13:30 – 15:30**	**Authors**
169	Extension of an Open and Vendor-neutral DICOM-CT-PD Format for Photon-Counting-Detector CT	Zhongxing Zhou (Mayo Clinic)*; Scott Hsieh, Jarod Wellinghoff, Solveig Narum, Cynthia McCollough, Lifeng Yu (Mayo Clinic)
174	Design and Characterization of the Pyxis™ X-Ray Detection Platform for CdTe- and CZT-based Photon-Counting Computed Tomography	Josh Star-Lack (Varex Imaging Corp)*, Mattias Urech, Devang Savaliya, Simo Veijola, Henrik Lohman, Fredrik Reinholdsen, Kyle Holland, Kevin Holt, Steve Hoelzer, Alex Stewart, Mark Dinh, Thomas Frach, Ivan Mollov, Tuomas Pantsar, Christer Ullberg, Richard Colbeth (Varex Imaging Corp)
178	Material-Aware Pulse Pile-Up And Pixel Gain Correction for Spectral Photon Counting Detectors	Simo Veijola, Kyle Holland, Henrik Lohman, Fredrik Reinholdsen, Mattias Urech, Alex Stewart, Devang Savaliya, Richard Colbeth (Varex Imaging Corp); Josh Star-Lack (Varex Imaging Corp)*
182	Inter-institutional consistency of spectral results from a clinical photon-counting CT: effects of tube voltage and radiation dose level	Leening Liu (University of Pennsylvania)*; Martin Rybertt, Peter Noël (University of Pennsylvania); Liqiang Ren, Xinhui Duan (UT Southwestern Medical Center); Kishore Rajendran (Mayo Clinic); Di Zhang, Grace Kim, Michael McNitt-Gray (University of California, Los Angeles)
186	Quantitative Gadolinium K-edge Imaging of Hepatobiliary Excretion Using Clinical PCCT: An In Vivo Study	Martin Rybertt (University of Pennsylvania)*; Leening Liu, George McClung (University of Pennsylvania); Pooyan Sahbaee (Siemens Healthineers); Tristan Nowak (Siemens Healthineers AG); Manoj Mathew, Peter Noël, Ali Dhanaliwala (University of Pennsylvania)
190	Water Beam Hardening and Pulse Pileup Correction with Small Neural Networks	Jannik Dettmer (Siemens Healthineers)*; Julien Erath, Matthias Baer-Beck (Siemens Healthineers); Juliane Szkitsak (University Hospital Erlangen, Radation Clinic); Karl Stierstorfer (Siemens Healthineers); Christoph Bert (University Hospital Erlangen, Radiation Clinic)
195	Comparison of SIRT and FBP Reconstruction Methods for Material Characterization in Photon-Counting CT: a Monte Carlo Study	Gustavo Silva, Renan Santos (University of Sao Paulo); Alessandra Tomal (University of Campinas); Gisell Boiset (University of Sao Paulo); Paulo Costa (University of Sao Paulo)*

200	Physics-Augmented Simulation of Deep Silicon Photon Counting Detector	Haomin Hou (Southern Medical University)*; Jiabing Sheng (Sun Yat-sen University Cancer Center); Yankun Wang, Yule Li, Junhao Wu, Deyang Kong, Jingyi Liao, Hao Wang, Zhaoying Bian (Southern Medical University); Jianhua Ma (Xi'an Jiaotong University); Dong Zeng (Southern Medical University)
205	Bias and Noise Problems in Spectral Photon-counting CT Imaging Task at Low-count Conditions	Jia Wei (Johns Hopkins University)*; Karl Stierstorfer, George Fung, Christoph Polster (Siemens Healthineers); Katsuyuki Taguchi (Johns Hopkins University School of Medicine)

	Beyond absorption X-ray imaging	**Session chairs: X. Tang, P. Noel**
Page #	**Oral session T2** **Tuesday 10:50 – 11:50**	**Authors**
210	Neighbor2Inverse: Self-Supervised Denoising for Low-Dose Region-of-Interest Phase Contrast CT	Johannes Thalhammer (Technical University of Munich)*; Lorenzo D'Amico, Lucy Costello (Monash University); Sebastian Peterhansl, Tina Dorosti, Florian Schaff (Technical University Munich); Jannis Ahlers (Monash University); Ronan Smith (University of Adelaide); Marcus Kitchen (Monash University); Franz Pfeiffer (Technical University Munich); Martin Donnelley (University of Adelaide); Daniela Pfeiffer (TUM Klinikum); Kaye Morgan (Monash University)
214	Bridging the Simulation-to-Reality Gap in Snapshot CAXRDT: A Confidence-Guiding Fine-Tuning Approach	Shengzi Zhao (Tsinghua University)*; Donghang Miao, Yuxiang Xing (Tsinghua University)
218	Geometric Calibration for X-ray Fluorescence Tomography using Multi-Support Consistency	Meo van Duijnen Montijn (ESRF – The European Synchrotron Radiation Facility, France)*; Jérôme Lesaint (ESRF, France); Jean Michel Létang, Simon Rit (INSA Lyon, France)
	More on this in poster session I **Monday 13:30 – 15:30**	**Authors**
223	X-ray Dark-field Imaging of Iron Oxide Nanorods for Instantaneous DSA (iDSA)	Xiangyang Tang (Emory University School of Medicine)*; Yi Yang, Duhee Joen, Huiqiao Xie, Hui Mao (Emory University School of Medicine)
227	Structured Loss Amplification for U-Net-based Human-scale Dark-field CT Streak Reduction	Daniel Frey (TUM)*; Tina Dorosti, Johannes Thalhammer, Josepha Hilmer, Paulina Bleuel, Theresa Hiu, Sebastian Peterhansl, Julian McGinnis (TUM); Thomas Koehler (Philips GmbH Innovative Technologies); Daniela Pfeiffer, Franz Pfeiffer, Daniel Rueckert, Florian Schaff (TUM)

232	Beam Hardening Correction for Dark-Field Imaging in a Human-Scale CT Prototype	Lennard Kayser (Technical University of Munich)*; Daniel Frey, Josepha Hilmer, Sofia Demianova, Tina Dorosti, Maximilian Lochschmidt (Technical University of Munich); Thomas Koehler (Philips Innovative Technologies); Daniela Pfeiffer, Franz Pfeiffer (Technical University of Munich)
237	Simulation Study of Scatter Rejection in Time-of-Flight CT	Altea Lorenzon (Johns Hopkins University)*; Xiao Jiang, J. Webster Stayman (Johns Hopkins University)

	Image quality, radiomics & dose	**Session chairs: P. Costa, S. Leng**
Page #	**Oral session T3** **Tuesday 14:20 – 16:00**	**Authors**
241	Incorporating Tube Current Latency in Organ Risk-Based CT Tube Current Modulation	Catherine Knobloch (DKFZ)*; Philipp Oelze (DKFZ); Andreas Prokein, Michael Grasruck, Matthias Baer-Beck (Siemens Healthineers AG); Marc Kachelrieß (DKFZ)
246	Low Contrast Detectability Performance in a Realistic Anthropomorphic 3D Printed Phantom vs standard QA phantoms	Daniel Shin (Canon Medical Systems Corporation)*; Leening Liu, Jessica Im (University of Pennsylvania); Tomoe Hagio, Bernice Hoppel, Amar Dhanantwari (Canon Medical Systems USA); Peter Noel (University of Pennsylvania); Kirsten Boedeker (Canon Medical Systems Corporation)
252	Risk-Minimizing Tube Current Modulation Compared to Clinical Protocols	Philipp Oelze (German Cancer Research Center)*; Aniol Serra Juhe, Daniel Mosig (University of Erlangen-Nuremberg); Michael Knaup (German Cancer Research Center); Michael Lell (Klinikum Nürnberg); Andreas Maier (University of Erlangen-Nuremberg); Marc Kachelrieß (German Cancer Research Center)
256	The Information Cliff: Testing the Physical Limits of Quantitative Imaging	John Hoffman (UCLA Radiology)*
261	Mean Squared Error, Structure Similarity, and Channelized Hotelling Observer Study in Machine Learning	Larry Zeng (Utah Valley University)*
	More on this in poster session I **Monday 13:30 – 15:30**	**Authors**
265	Task-Based Modeling of Pulmonary Nodule Detectability in CT	Elsa Pimenta, Gisell Boiset (University of Sao Paulo); Luuk Oostveen, Ioannis Sechopoulos (Radboud University Medical Center); Alessandra Tomal (University of Campinas); Paulo Costa (University of Sao Paulo)*

269	Deep Learning Enhancement of Ultralow-Dose 3D CT Scouts for Organ Localization and Dose	Liyan Sun (Stanford University)*; Sen Wang, Maria Jose Medrano, Kaylee Fang (Stanford University); Grant Stevens (GE Healthcare); Justin Tse, Adam Wang (Stanford University)
274	Physics-Informed Scalable CycleGAN for CT Reconstruction Kernel Conversion and Harmonization	Seoyoung Lee (Mayo Clinic)*; Aeden Davis, Shravani Kharat, Wilson Gonsalves, Shuai Leng, Lifeng Yu, Cynthia McCollough, Francis Baffour, Hao Gong (Mayo Clinic)
278	A practical model observer for a task-based image quality assessment based on a local hypothesis test	Okkyun Lee (DGIST)*
282	GrainNet: An Unsupervised Framework for Frequency-Dependent Higher-Order Noise Modeling in CT	Ou Li (Radboud University Medical Center)*; Kirsten Boedeker, Tomoe Hagio, Daniel Shin (Canon Medical Systems Corporation); Koen Michielsen, Luuk Oostveen, Ioannis Sechopoulos (Radboud University Medical Center)
287	CT Reconstruction from Dual-View Depth Images and Topograms for Prospective Risk Minimization	Aniol Serra Juhé (Pattern Recognition Lab, FAU Erlangen-Nuremberg)*; Daniel Mosig (Pattern Recognition Lab, FAU Erlangen-Nuremberg); Chang Liu (Erlangen National High Performance Computing Center, Germany); Philipp Oelze (German Cancer Research Center); Michael Lell (Radiology and Nuclear Medicine, Klinikum Nuremberg, Paracelsus Medical University); Marc Kachelrieß (German Cancer Research Center); Andreas Maier (Pattern Recognition Lab, FAU Erlangen-Nuremberg)

	NDT & Dental Imaging	**Session chairs: S. Cho, J. Lesaint**
Page #	**Oral session T4** **Tuesday 16:20 – 17:40**	**Authors**
291	In-house nano-xCT with an x-ray spectrometer and a hybrid photon counting detector for integrated circuit metrology	Jordan Fonseca (National Institute of Standards and Technology)*
295	DBP-based geometry-robust artifact removal for rotational CL	Wuliang Shi (Tsinghua University)*; Yuxiang Xing, Changyu Chen (Tsinghua University)
299	Simulation of tire phantom based on X-ray computed tomography	Jiaying Zhang (University of California, Santa Cruz); Zhencheng Lin (University of California, San Diego); Zhicong Yu (University of California, Santa Cruz)*
305	An iterative filtered-backprojection algorithm to exactly and stably solve the long object problem in cone-beam CT with source positions evenly distributed on a cylinder	Murdock Grewar (Australian National University)*, Frédéric Noo, (University of Utah)

	More on this in poster session II **Wednesday 13:30 – 15:30**	**Authors**
310	Practical Dental CBCT Reconstruction under Truncated Projections	Kiwan Jeon (National Institute for Mathematical Sciences)*; Hyoung Suk Park (National Institute for Mathematical Sciences)
314	Two-Stage Cascaded Caries Detection on Panoramic X-Ray via Slicing-Aided Hyper Inference	Subin Park (Yonsei University)*; Yoonha Eo, Jongduk Baek (Yonsei University)
319	X-ray CT based Anomaly Detection by Pseudo-Volumetric Reconstruction with Channel-Disentangled Memory	Yunkun Bai (Tsinghua University); Xiang Zou, Yuxiang Xing (Tsinghua University)*
323	Pulsed X-ray micro-CT for minimizing non-contributory irradiation during scanning	Takayuki Okamoto (Chiba University)*; Yuki Morimoto, Hideaki Haneishi (Chiba University)
327	Quantitative measurement of 3D resolution in cone beam CT using touching spheres	Matthew Andrew (Zeiss)*; Konrad Lingel, Johannes Schuele (Zeiss)
331	SynthCaries: Dental Caries Generation on Panoramic Radiograph using Diffusion	Yoonha Eo (Yonsei University); Eunmin Lee (Yonsei University)*; Jongduk Baek (Yonsei University)
335	Gradient Step Plug-and-Play Model for Dental Cone-Beam CT Reconstruction	Idris Tatachak (CREATIS, Acteon Group)*; Luis Kabongo (Acteon Group); Nicolas Papadakis (Univ. Bordeaux, IMB); Xavier Ripoche (Acteon Group); Simon Rit (CREATIS)
340	Super-Resolution µCT of Complete Pathological Tarsal Bones	Sascha Senck (University of Applied Sciences Upper Austria)*; Lukas Nepelius, Johann Kastner, Jonathan Glinz (University of Applied Sciences Upper Austria); Klemens Trieb (Paracelsus Medical University); Elena Kranioti (University of Crete); Bernhard Plank, Patrick Weinberger (University of Applied Sciences Upper Austria)

	Angiographic, cardiac & high-resolution imaging	**Session chairs: K. Taguchi, J. Maier**
Page #	**Oral session W1** **Wednesday 8:10 – 9:50**	**Authors**
345	Clinical Feasibility of Fully Autonomous Cardiac CT Using Pulsed Mode Projections	Pengwei Wu (GE HealthCare)*; Eri Haneda, Isabelle Heukensfeldt Jansen, Stephen Araujo (GE Healthcare); Albert Hsiao, Elliot McVeigh (University of California San Diego); Bruno De Man (GE Healthcare)

350	Mixed Prior Diffusion Posterior Sampling for High Resolution CT Reconstruction	Peiqing Teng (Johns Hopkins University)*; Xin Wang (Johns Hopkins University); Jian Zhou, Liang Cai, Tzu-Cheng Lee, Ruoqiao Zhang (Canon Medical Research USA, Inc.); J. Webster Stayman (Johns Hopkins University)
355	Diffusion Posterior Sampling for Angiographic Reconstruction using Limited Angle and Sparse View data for 4D-DSA	Shudong Li (Johns Hopkins University)*; Suyu Liao (University of Pennsylvania); Peiqing Teng, Webster Stayman (Johns Hopkins University); Grace Gang (University of Pennsylvania)
359	Deep Learning Deblooming in Coronary CT Angiography with CatSim Training Data	Muhan Shao (GE HealthCare)*; Christopher Wiedeman, Lin Fu (GE HealthCare); Ge Wang, (Department of Biomedical Engineering, Rensselaer Polytechnic Institute, New York, USA), Bruno De Man (GE HealthCare)
364	Cardiac Coronary CT Angiography Image Quality Enhancement with Multi-Vendor Generalizability using Deep Learning-based High-resolution and Multi-energy Imaging	Shaojie Chang (Mayo Clinic)*; Maryam Sadeghian, Benjamin Wilson, Eric Williamson, Thomas Foley, Cynthia McCollough, Shuai Leng (Mayo Clinic)
	More on this in poster session II **Wednesday 13:30 – 15:30**	**Authors**
369	An Accurate and Real-Time 3D/2D Coronary Artery Registration Method for Depth-Guided Coronary Intervention Navigation	Ruifeng Chen (Shenzhen Institutes of Advanced Technology Chinese Academy of Sciences); Zhongwei Sun, Lei Song (Fuwai Hospital); Mengxuan Yuan (Radiology, Jiangdu People' s Hospital of Yangzhou); Jun Xiang (United Imaging Healthcare Limited Company); Yang Chen (Southeast University); Dong Liang (Shenzhen Institutes of Advanced Technology Chinese Academy of Sciences); Yinsheng Li (Shenzhen Institutes of Advanced Technology Chinese Academy of Sciences)*
373	Diffusion Model-Based Motion Compensation for Coronary CT Angiography in Dual Wide-Detector CT	Zhiyang Fu (United Imaging Healthcare, North America)*; Jiao Tian, Xi Zhang, Yifu Mao, Wenjing Cao, Guotao Quan (Shanghai United Imaging Healthcare); Wenying Wang (United Imaging Healthcare North America)
377	NERVE: A Near-Real-Time Three-Dimensional Reconstruction Approach for Interventional Devices and Surrounding Vessels Using a Monoplane C-arm Cone-Beam CT	Yang Liu, Zexi Sun (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences); Bingshuai Zhao (United Imaging Healthcare Limited Company); Yikun Zhang, Yang Chen (Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications, Southeast University); Yinsheng Li (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences)*

381	Coronary Mask–Guided Registration for 4D Cardiac CT Dataset Construction	Yuang Wang (Tsinghua University)*; Shuo Wang, Changyu Chen, Le Shen (Tsinghua University); Yunqiang An, Yang Gao, Bin Lu (Fuwai Hospital); Dongrui Dai, Muge Du, Li Zhang, Zhiqiang Chen (Tsinghua University)
385	PC-FLOW: A Physics-Constrained Conditional Flow Matching Model For High Pitch Helical Three Dimensional Digital Subtraction Angiography	Yi Liu, Ruifeng Chen, Zexi Sun (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences); Canzhen Ma, Jun Xiang (United Imaging Healthcare Limited Company); Dong Liang (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences); Yinsheng Li (Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences)*
389	A Framework for Ultra-High-Resolution CT via Data Synthesis and Latent Diffusion Bridge	Sun Yang (Tsinghua University)*; Zhang ShaoYang, Zhang Li (Tsinghua University)
393	Investigations on bone representation in voxelized patient models for estimation of scatter profiles in cone-beam CT	Frédéric Noo (University of Utah)*, Zijia Guo, Mehmet Bugrahan Oktay (University of Utah), Michael Manhart (Siemens Healthcare, Germany), Julie DiNitto (Siemens Medical Solutions, USA)
397	Deep Low-Rank Subspace Learning for 5D Myocardial Perfusion Photon-Counting CT Denoising	Meng Xiao (Biomedical Engineering, Southern Medical University)*; Xiao Meng (Biomedical Engineering, Southern Medical University); Jiangjun Peng (Mathematics and Statistics, Northwestern Polytechnical University); Jingyi Liao, Yue Liu, Hao Wang, Zhaoying Bian (Biomedical Engineering, Southern Medical University); Jianhua Ma (Life Science and Technology, Xi'anJiaotong University); Dong Zeng (Biomedical Engineering, Southern Medical University)

	Limited angle & sparse view imaging	**Session chairs : S. Vedantham, K.Michielsen**
Page #	**Oral session W2** **Wednesday 10:10 – 11:30**	**Authors**
402	Accelerating Low-Frequency Convergence for Limited-Angle DBT via Two-Channel Fidelity in PDHG	Taro Iyadomi, Ricardo Parada (DNE); Anna Kim (UCSD); Lily Jiang (UCLA); Emil Sidky (University of Chicago); William Chang (UCLA)*
406	Unified Analysis of Uniqueness and Stability for Interior and Exterior Problems in Tomographic Image reconstruction	Katsuya Fujii (Kio University)*; Hiroyuki Kudo (University of Tsukuba)
411	Anisotropic Total Variation Regularization for Artifact Reduction in Contrast-Enhanced Digital Breast Tomosynthesis using PDHG Optimization	Yixin Li (Stony Brook University)*; Xiangyi Wu (Stony Brook University); Emil Sidky (The University of Chicago); Wei Zhao (Stony Brook University)

416	Basis material image reconstruction for digital breast tomosynthesis using a single kV scan	JP Phillips (The University of Chicago)*; Emil Sidky, Ingrid Reiser, Zheng Zhang, Buxin Chen, Xiaochuan Pan (The University of Chicago)
	More on this in poster session II **Wednesday 13:30 – 15:30**	**Authors**
420	Robust Limited-Angle CT Reconstruction via Geometry-Conditioned Multi-Environment Learning	Shujie Jin (University of Wisconsin-madison)*; Zilin Jiang, Ran Zhang, Guang-Hong Chen (University of Wisconsin-madison)
424	Efficient Projection Correction Approach for High-Density Object Artifact Reduction in Digital Breast Tomosynthesis	Subong Hyun (KAIST)*; Seoyoung Lee (Mayo Clinic); Seungryong Cho (KAIST)
428	Prior-Embedded Implicit Neural Representation and Dual-Domain Ordered-Subsets for High-Resolution Limited-Angle Non-Coplanar CBCT Imaging	Siqi Ye (Mayo Clinic)*; Yu Gao, Hao Zhang, Lei Xing (Stanford University)
434	Correction Strategies for Material Decomposition in Dual-Energy Digital Breast Tomosynthesis for Cyst-Solid Mass Differentiation	Martina Nassi (Radboud University Medical Center)*; Koen Michielsen, Ioannis Sechopoulos (Radboudumc)
439	Memory-efficient optimization of implicit neural representations for CT reconstruction	Mahrokh Najaf (Marquette University); Greg Ongie (Marquette University)*
443	Limited-Angle CT Reconstruction via Conditional Adversarial Diffusion with Dynamic Multi-Scale Injection	Yutong He (National Center for Applied Mathematics Shenzhen, Southern University of Science and Technology)*; Dimeng Xia, Shusen Zhao (National Center for Applied Mathematics Shenzhen, Southern University of Science and Technology; Detection Institute for Advanced Technology Longhua-Shenzhen)
448	Ensemble learning of different diffusion priors for sparse view CT reconstruction	Sho Ozaki (Hirosaki University)*; Shizuo Kaji (Kyoto University); Toshikazu Imae (University of Tokyo Hospital); Kanabu Nawa (Osaka Medical and Pharmaceutical University); Hideomi Yamashita, Keiichi Nakagawa (University of Tokyo Hospital)

	Addressing noise, low dose & material decomposition	**Session chairs : F. Jolivet, Y. Xing**
Page #	**Oral session W3** **Wednesday 16:00 – 17:40**	**Authors**
452	Addressing noise and resolution non-uniformities in fully 3D axial CT reconstruction	Stanislav Zabic (Philips)*; Kevin M. Brown (Philips)
456	MeanFlow Posterior Sampling for Low-Dose Head CT Reconstruction	Dufan Wu (The Ohio State University)*; Yuang Wang (Tsinghua University); Matthew Tivnan, Quanzheng Li (Massachusetts General Hospital and Harvard Medical School)
460	Median2Median: Zero-shot Suppression of Structured Noise in CT Images	Jianxu Wang (Rensselaer Polytechnic Institute)*; Wenjun Xia (Rensselaer Polytechnic Institute); Mannudeep Kalra (Massachusetts General Hospital); Ge Wang (Rensselaer Polytechnic Institute)
466	Joint Spectrum Estimation and Material Decomposition for One-kVp-switching Dual-energy CBCT: A Pre-clinical Study	Yu Gao (UT Southwestern Medical Center); Tingliang Zhuang, You Zhang (UT Southwestern Medical Center)*
470	Enhancing Pediatric Dual-Layer Spectral CT Quantification Using K-Edge Filtration: Experimental Study	Yinglin Ge (University of Pennsylvania)*; Olivia Sandvold, Roland Proksa (University of Pennsylvania); Amy Perkins (Philips Healthcare); Thomas Koehler, Heiner Daerr (Philips Innovative Technologies); Yannan Jin, Kevin Brown (Philips Healthcare); Leening Liu, Peter Noël (University of Pennsylvania)
	More on this in poster session II **Wednesday 13:30 – 15:30**	**Authors**
475	Dual-Domain Posterior Mean-Guided Diffusion Bridge Model for Sparse-View CT Reconstruction	Xun Zhang (Tsinghua University)*; Hewei Gao, Li Zhang, Yuxiang Xing, Changyu Chen, Zhiqiang Chen (Tsinghua University)
479	Causal Statistical Foundations and Causal Invariance Learning for Material Basis Estimations from Single-kV CT Images	Zilin Jiang (University of Wisconsin-Madison)*; Ran Zhang, Yijing Wu, Shujie Jin (University of Wisconsin-Madison); Zhihua Qi (Henry Ford Health); Ke Li (MD Anderson Cancer Center); Guang-Hong Chen (University of Wisconsin-Madison)
483	Data-Driven Linearization of the Spectral CT Forward Model using Gaussian Mixture Models	Gauthier Multari (CREATIS - THALES)*; Guillaume Delorme (THALES); Frédéric Jolivet (THALES); Simon Rit (CREATIS - CNRS)
487	The Sufficiency of Material Bases in Polychromatic CT: A Quantitative Effective Energy Perspective	Xiaopeng Yu (University of Wisconsin-Madison)*; Ran Zhang, Guang-Hong Chen (University of Wisconsin-Madison)

491	Physics-Informed Adaptive Filtering Based on VVBP τ-Coordinate Properties for Low-Dose CT Image Recovery	Junhao Wu (Southern Medical University)*; Yuanxiang Gao, Haomin Hou, Deyang Kong, Yankun Wang, Yule Li, Jingyi Liao, Hao Wang, Zahoying Bian (Southern Medical University); Jianhua Ma (Xi'an Jiaotong University); Dong Zeng (Southern Medical University)
497	WICD: Wavelet-Informed Conditional Diffusion Model for Low-Dose CT Denoising	Changsheng Fang (University of Massachusetts Lowell)*; Bahareh Morovati, Shuo Han, Yu Shi, Li Zhou, Shuyi Fan, Dayang Wang, Hengyong Yu (University of Massachusetts Lowell)
502	Noise and Anatomy Adaptive Convolution for Efficient Low-Dose CT Denoising	Minwoo Yu (Yonsei University)*; Jongduk Baek (Yonsei University); Adam Wang (Stanford University)
507	Lag-Induced Spectral Contamination in Spectral CBCT: A Comparative Study of Fast kV-Switching and Dual-Layer Flat-Panel Detector Technologies	Yuting Chen (Tsinghua University)*; Hao Zhou, Hewei Gao (Tsinghua University)
511	Low-Dose CT Image Denoising Using Data-Coupled Flow Matching	Bangyan Huang, Jinyi Qi (University of California, Davis); Jian Zhou (Canon Medical Research USA)*
515	Application of UNet-based Network for Projection-domain Material Decomposition in CBCT, A Simulation Study	Yuedong Yuan (STIMULATE, Otto-von-Guericke University)*, Daniel Punzer, Georg Rose (STIMULATE, Otto-von-Guericke University)
519	Enhancing Reproducibility for CT-based Radiomics: A Pragmatic Approach to Mitigate the Impact of Reconstruction Kernel Variations	Yun-Huan Lyu (University of Washington)*; Wei Wu (University of Washington); Yuzheng Zhang (Fred Hutchinson Cancer Center); Paul Kinahan (University of Washington)
524	Streak Artifact Reduction in Human-scale Dark-field CT Using 3D Gaussian Splatting	Tina Dorosti (TUM)*; Daniel Frey, Johannes Thalhammer, Josepha Hilmer, Paulina Bleuel, Sebastian Peterhansl, Julian McGinnis, Daniela Pfeiffer, Franz Pfeiffer, Daniel Rueckert, Florian Schaff (TUM)

Multisource C-arm CT: Initial Results with a 2×2 Carbon Nanotube Source Array

Hojin Jung, Jungsu Kang, Youngkwang Kim, Youngjun Roh, and Jongduk Baek

***Abstract*—C-arm computed tomography (CT) systems are widely used in intraoperative imaging due to their compact and flexible design. However, the achievable field of view (FOV) is inherently limited by geometric constraints, often resulting in truncated anatomy and incomplete coverage. While multi-source C-arm CT systems have been explored to extend in-plane coverage, extension along the axial direction remains challenging. In this work, we propose a 2×2 multi-source C-arm CT system to extend coverage along the axial direction. Carbon nanotube (CNT)-based X-ray sources are employed to allow rapid sequential activation of the four sources while sharing a single flat-panel detector. To validate the proposed concept, we designed and constructed a prototype multi-source C-arm CT system and implemented a complete imaging pipeline, including geometric calibration, projection data correction, source-wise reconstruction, and fusion of reconstructed volumes. Phantom experiments demonstrate that stable reconstructions can be obtained from individual source channels and that multi-source fusion enables natural axial coverage extension and reduced image noise compared to single-source reconstructions. These results confirm the feasibility of the proposed 2×2 multi-source C-arm CT system and demonstrate its ability to extend the effective field of view in compact C-arm CT systems.**

***Index Terms*—C-arm CT system, Multi-source CT.**

I. Introduction

C-ARM computed tomography (CT) systems provide volumetric imaging capabilities within a compact and mobile design, enabling targeted examination of specific anatomical regions depending on surgical requirements [1]. These characteristics make C-arm CT particularly well suited for interventional and intraoperative applications, where flexible imaging and rapid deployment are essential. By supporting surgical planning and real-time anatomical assessment, C-arm CT has significantly improved procedural accuracy and overall clinical workflow. C-arm CT systems are inherently constrained by design considerations intended to minimize interference with the surgical environment [2]. The compact geometry and limited gantry size restrict the achievable field of view (FOV), which limits anatomical coverage and can reduce practical usability, motivating research on effective FOV extension strategies.

Several prior studies have investigated CT system designs that extend the field of view (FOV) using multiple X-ray sources and corresponding reconstruction strategies [3], [4]. Among these approaches, inverse-geometry CT (IGCT) is a representative example that employs distributed source arrays to achieve extended in-plane coverage [5], [6]. A recent study proposed a multi-source C-arm CT system in which three X-ray sources are positioned to extend the in-plane field of view beyond what is achievable with a conventional single-source design [7]. The system employs carbon nanotube (CNT)-based X-ray sources, which can be rapidly switched on and off using digital control signals [8], [9]. This fast electronic switching enables projection data from individual sources to be acquired sequentially within a single system rotation, allowing multiple source views to be captured by a single detector without hardware duplication.

In this work, we implement a multi-source C-arm CT system by employing a 2×2 multi-source C-arm CT configuration with CNT-based X-ray sources positioned at the vertices of a rectangular configuration spanning both in-plane and axial directions. We constructed a prototype system and validated its feasibility through real data acquisition and reconstruction experiments. The following sections describe the proposed system configuration, acquisition strategy, and reconstruction framework, along with experimental validation using real data.

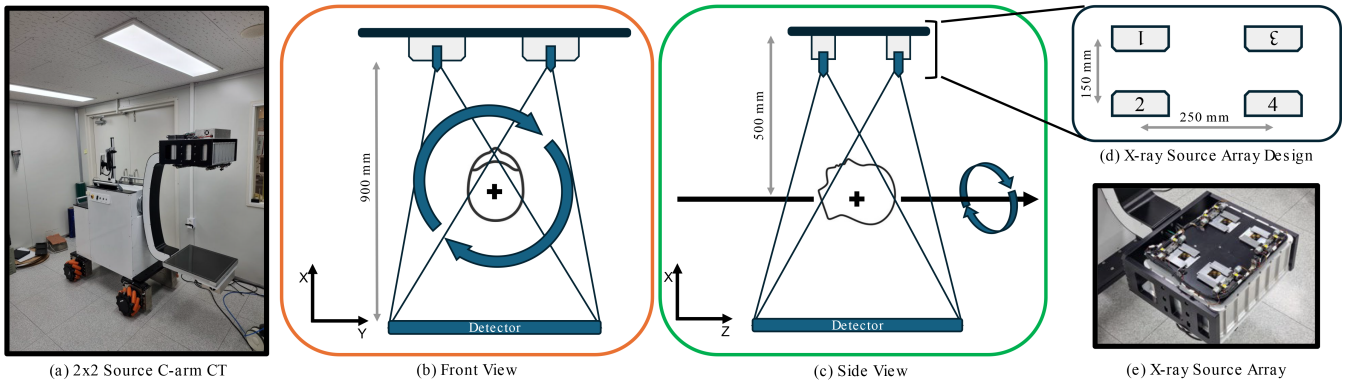
II. System Configuration

Fig. 1 (a) shows the proposed 2×2 multi-source C-arm CT system, including the constructed prototype and the arrangement of multiple X-ray sources. As illustrated in Fig. 1 (b) and Fig. 1 (c), four CNT-based X-ray sources are arranged at the vertices of a rectangular 2×2 configuration with respect to a single flat-panel detector. The front view in Fig. 1 (b) shows the lateral arrangement of the sources, while the side view in Fig. 1 (c) shows the axial separation of the sources along the axial direction.

All X-ray sources share a single flat-panel detector, and the detector specifications are summarized in Table I. The relative layout and separation distances of the four sources are illustrated in Fig. 1 (d), and a photograph of the implemented CNT-based X-ray source array is shown in Fig. 1 (e).

Hojin Jung and Jongduk Baek are with the Department of Artificial Intelligence, Yonsei University, Seoul 03722, Republic of Korea, and also with Bareunex Imaging, Inc., Seoul, Republic of Korea (e-mail: jongdukbaek@yonsei.ac.kr).

Jungsu Kang, Youngkwang Kim, and Youngjun Roh are with KohYoung Technology, Yongin, Republic of Korea.



(a) 2x2 Source C-arm CT (b) Front View (c) Side View (d) X-ray Source Array Design (e) X-ray Source Array

Fig. 1. Overview of the proposed 2×2 multi-source C-arm CT system and source configuration. (a) Photograph of the implemented 2×2 multi-source C-arm CT prototype system. (b) Front view illustrating in-plane coverage extension achieved by lateral separation of multiple X-ray sources. (c) Side view showing coverage extension along the axial direction enabled by axial source arrangement. (d) Schematic of the 2×2 X-ray source array layout, indicating relative source positions and separation distances. (e) Photograph of the carbon nanotube (CNT)-based X-ray source array used in the prototype system. All X-ray sources are sequentially activated via electronic switching, and projection data from multiple source positions are acquired using a single flat-panel detector at each projection angle.

During each projection angle, the X-ray sources are sequentially activated, and projection data corresponding to different source positions are acquired within a single angular view. Each source is operated with a 15 ms exposure followed by an 85 ms idle interval before switching to the next source. This acquisition scheme is enabled by the use of CNT-based X-ray sources, which support rapid electronic switching between sources.

As the proposed system is implemented as an early-stage prototype, the scope of this study is focused on axial coverage extension in order to reduce complexity in power control and system integration. Accordingly, axial extension is investigated by sequentially switching the axially separated sources, while the effectiveness of in-plane multi-source acquisition has been demonstrated in prior studies.

III. Geometric Calibration

Geometric calibration is an essential step in computed tomography reconstruction because inaccurate system geometry directly leads to artifacts and incorrect CT numbers in reconstructed images. In the proposed multi-source C-arm CT system, projection data from multiple X-ray sources are fused during reconstruction, and accurate calibration is therefore required to properly align data acquired from different source positions.

For geometric calibration, we use a calibration phantom embedded with ball bearing (BB) markers [10], as shown in Fig. 2(a). The BBs provide distinct point features that are visible across projections acquired from all source positions. As shown in Fig. 2(b), the BBs are projected to different locations on the detector depending on the source position. The projected BB locations are used as feature points for calibration.

TABLE I
System specifications of the proposed 2×2 multi-source C-arm CT.

Parameter	**Specification**
Tube array	2×2 CNT X-ray sources
Tube pitch	250 mm $\times$ 150 mm
Source-to-detector distance (SDD)	900 mm
Source-to-object distance (SOD)	500 mm
Detector size	430.1 mm $\times$ 430.1 mm
Pixel pitch	140 μm
Pixel matrix	1536 $\times$ 1536 (2×2 binning)
CT scan range	$360 \times 1^{\circ}$
Reconstructed FOV	351(X) $\times$ 351(Y) $\times$ 306(Z) mm

The calibration parameters describe the relative geometry between the X-ray sources, the detector, and the phantom. Their initial values are determined from the nominal system configuration summarized in Table I. All parameters are defined in a global coordinate system centered at the isocenter, with the gantry rotation axis aligned with the z-axis. The y-axis is defined along the line connecting the detector center and the nominal center of the source panel. The estimated parameters include source offsets, detector offsets, detector rotations, phantom offsets, and phantom rotations, as illustrated in Fig. 3.

We constructed a projection function defined by the system geometry, which computes the projected locations of each BB on the detector as a function of the calibration parameters. The calibration parameters are estimated by minimizing a least-squares objective function defined on the differences between the measured and predicted detector coordinates. Let $\mathbf{u}_i(\theta) \in \mathbb{R}^2$ denote the detected image coordinates of the i-th bead at view angle θ, and let $\hat{\mathbf{u}}_i(\phi, \theta)$ denote the corresponding projected location computed by the projection

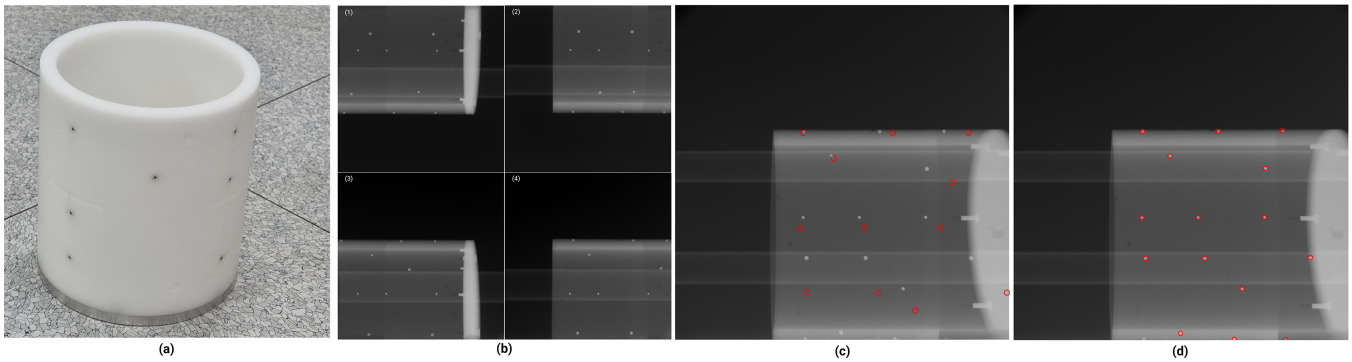


Fig. 2. Calibration phantom and representative projection images. (a) Photograph of the calibration phantom embedded with ball bearings (BBs). (b) Example projection images acquired from different source channels. (c) Simulated projected BB locations overlaid on the projection image with initial setting for optimization. Red circles are simulated projected BB location. (d) Simulated projected BB location after geometric calibration.

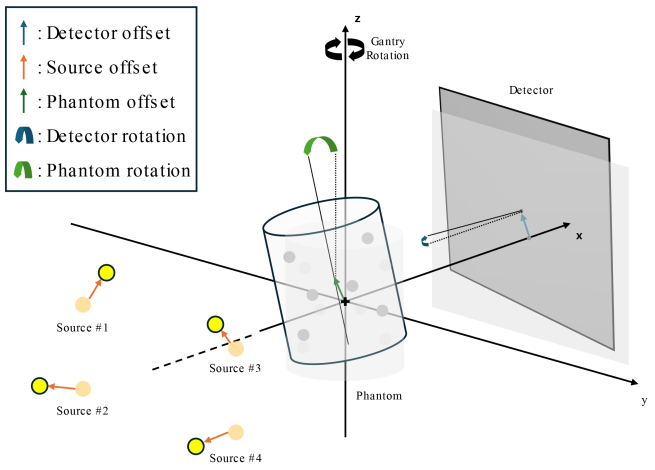


Fig. 3. Illustration of calibration parameters, including source, phantom, and detector offsets, phantom and detector rotations.

function with calibration parameters ϕ. The objective function is defined as

$$\mathcal{L}(\phi) = \sum_{\theta} \sum_{i} \|\mathbf{u}_i(\theta) - \hat{\mathbf{u}}_i(\phi, \theta)\|_2^2 . \tag{1}$$

Geometric calibration is performed sequentially across source channels. Calibration is first conducted on the first source channel, and the resulting detector and phantom parameters are fixed and reused for the remaining channels to maintain a consistent global geometry, while only the source offsets are optimized to compensate for relative misalignment between sources.

IV. Preprocessing of Raw Projection Data

Prior to reconstruction, the acquired raw projection data are preprocessed using standard correction steps. Initially, dead pixels in the detector are identified and corrected using neighboring pixel values. Dark-field images are acquired with the X-ray source turned off, and air-field images are acquired by imaging air with the X-ray source turned on and no object in the field of view.

Let I_{raw} denote the raw measured projection intensity, I_{dark} the dark-field image, and I_{air} the air-field image. Dark-field and air-field corrections followed by a logarithmic transformations are applied as

$$p = -\log\left(\frac{I_{\text{raw}} - I_{\text{dark}}}{I_{\text{air}} - I_{\text{dark}}}\right), \tag{2}$$

where p represents the log-transformed projection data proportional to line integrals of attenuation coefficients.

Beam hardening and detector lag corrections are additionally applied during preprocessing to compensate for spectral hardening effects and temporal signal persistence, respectively [11], [12]. All preprocessing steps are applied independently to the raw projection data from each source channel.

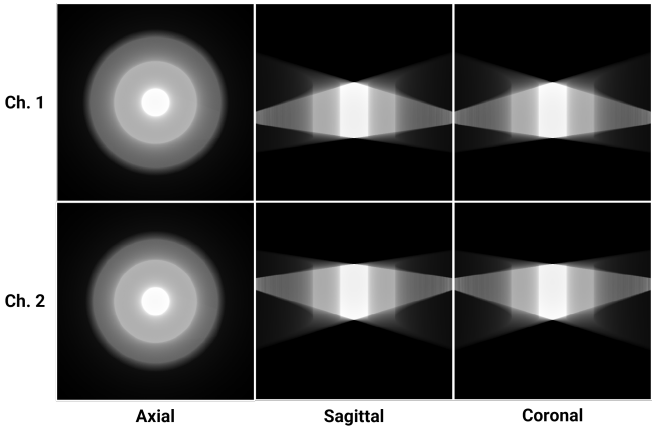


Fig. 4. Examples of weight maps in channel 1 and 2. The weight maps are reconstructed in an image space centered at the isocenter, spanning $[819.2, 819.2, 819.2]$ mm. The figure shows weight maps obtained from each individual source. Axial, coronal, and sagittal views are displayed from left to right.

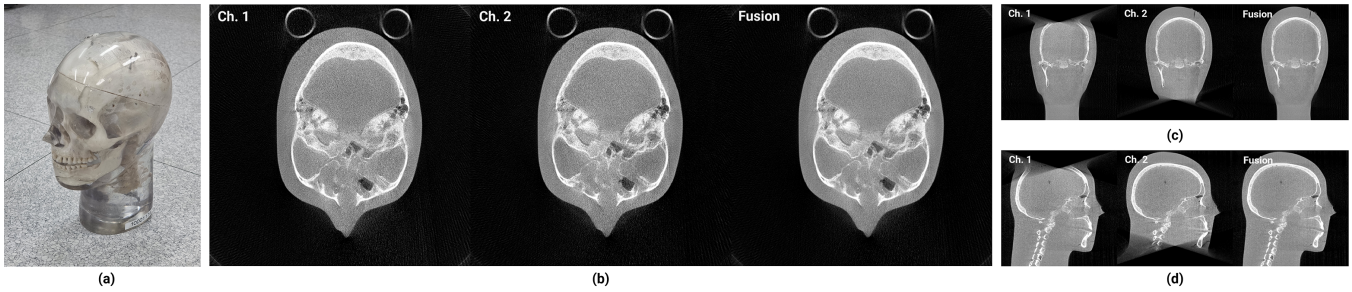


Fig. 5. Skull phantom reconstruction results. (a) Photograph of the skull phantom used in the experiment. (b) Axial slices reconstructed from source channel 1 (Ch. 1), source channel 2 (Ch. 2), and their fused reconstruction. (c) Coronal slices reconstructed from Ch. 1, Ch. 2, and the fused result. (d) Sagittal slices reconstructed from Ch. 1, Ch. 2, and the fused result. The reconstruction grid size was $1290 \times 1290 \times 1290$ with a voxel size of $0.2\ \mathrm{mm}$.

V. Reconstruction and Multi-Source Fusion

Reconstruction is performed independently for each X-ray source, followed by voxel-wise fusion of the reconstructed volumes. For each source, projection data are reconstructed using the Feldkamp-Davis-Kress (FDK) [13] algorithm based on the calibrated system geometry. All reconstructions are performed in the same image coordinate system centered at the isocenter.

Although the reconstructed volumes share the same image space, they exhibit spatially varying effective coverage due to the distinct acquisition geometries. To combine the source-specific reconstructions, a voxel-wise weight map is computed for each source. The weight map is obtained by backprojecting a projection tensor of all ones using the calibrated system geometry. This procedure provides a geometry-based measure of voxel-wise sampling density and is used to normalize coverage differences when combining reconstructions. Fig. 4 visualizes representative weight maps obtained by the proposed method. The resulting weight map is used to fuse source-specific reconstructions in regions with overlapping coverage. The final volume is obtained by weighted summation of the individual reconstructions.

VI. Results

Reconstruction results were evaluated using phantom experiments. Fig. 5 shows representative reconstructions of a skull phantom acquired with the proposed multi-source C-arm CT prototype using source channels 1 and 2. Reconstructions from individual source channels show visually consistent image quality and consistent anatomical structures within the field of view of each source, indicating that each source independently provides reliable reconstruction in its covered region. When reconstructions from multiple sources are combined using the geometry-based weight maps, the fused volumes exhibit seamless integration without visible discontinuities between source contributions. In particular, multi-source fusion enables natural extension of the reconstructed volume along the axial direction. Regions that are truncated or weakly covered in single-source reconstructions are complemented by additional source views, resulting in extended anatomical coverage along the z-axis. In addition to coverage extension, the fused reconstructions exhibit reduced image noise compared to single-source results, which can be attributed to the increased amount of projection data contributing to each voxel. These results demonstrate that the proposed multi-source configuration allows effective fusion of independently reconstructed volumes in a compact C-arm geometry.

VII. Conclusion

In this work, we implemented a 2×2 multi-source C-arm CT system and present an initial reconstruction result. The proposed design builds upon previously reported multi-source configurations and introduces an additional source arrangement to enable axial coverage extension within a compact geometry. By extending the source layout along the axial axis, the system provides complementary sampling that cannot be achieved with in-plane multi-source configurations alone.

To validate the proposed concept, we designed and constructed a prototype multi-source C-arm CT system using CNT-based X-ray sources and a shared flat-panel detector. A complete imaging pipeline was implemented, including geometric calibration, data correction, source-wise reconstruction, and fusion of reconstructed volumes. Phantom experiments demonstrate that independently reconstructed volumes from each source can be consistently fused, resulting in natural extension of the reconstructed volume along the axial axis and reduced image noise due to increased data availability.

These results indicate that the proposed 2×2 source configuration provides a practical means of extending axial coverage in compact C-arm CT systems. The presented prototype and experimental results offer a concrete foundation for further development of multi-source C-arm CT systems and future studies under clinically relevant imaging conditions.

ACKNOWLEDGMENT

This work was supported in part by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIT) under Grant RS-2025-00553670, RS-2023-00219019, and RS-2023-00240135. (Corresponding author: Jongduk Baek.)

REFERENCES

[1] S. Schafer, S. Nithiananthan, D. Mirota, A. Uneri, J. Stayman, W. Zbijewski, C. Schmidgunst, G. Kleinszig, A. Khanna, and J. Siewerdsen, "Mobile c-arm cone-beam ct for guidance of spine surgery: image quality, radiation dose, and integration with interventional guidance," *Medical physics*, vol. 38, no. 8, pp. 4563–4574, 2011.

[2] H. Kageyama, S. Yoshimura, K. Uchida, and T. Iida, "Advantages and disadvantages of multi-axis intraoperative angiography unit for percutaneous pedicle screw placement in the lumbar spine," *Neurologia medico-chirurgica*, vol. 57, no. 9, pp. 481–488, 2017.

[3] V. B. Neculaes, P. M. Edic, M. Frontera, A. Caiafa, G. Wang, and B. De Man, "Multisource x-ray and ct: Lessons learned and future outlook," *Ieee Access*, vol. 2, pp. 1568–1585, 2014.

[4] Z. Wang, Y. Liu, S. Wang, X. Bian, Z. Li, and J. Cui, "Analytical reconstructions of full-scan multiple source-translation computed tomography under large field of views," *Journal of X-Ray Science and Technology*, vol. 31, no. 6, pp. 1245–1262, 2023.

[5] J. Baek, B. De Man, J. Uribe, R. Longtin, D. Harrison, J. Reynolds, B. Neculaes, K. Frutschy, L. Inzinna, A. Caiafa *et al.*, "A multi-source inverse-geometry ct system: initial results with an 8 spot x-ray source array," *Physics in Medicine & Biology*, vol. 59, no. 5, p. 1189, 2014.

[6] T. G. Schmidt, R. Fahrig, N. J. Pelc, and E. G. Solomon, "An inverse-geometry volumetric ct system with a large-area scanned source: a feasibility study," *Medical physics*, vol. 31, no. 9, pp. 2623–2627, 2004.

[7] S. Moon, J. Kang, Y. Kim, E. Jo, P. Jeong, Y. Roh, and J. Baek, "Carbon nanotube-based multiple source c-arm ct system: feasibility study with prototype system," *Optics Express*, vol. 31, no. 26, pp. 44 772–44 797, 2023.

[8] J.-W. Lee, J.-W. Jeong, S. Park, J.-W. Kim, J.-T. Kang, K. N. Yun, E. Go, H. Jeon, Y. Ahn, J.-H. Yeon *et al.*, "Adaptively variable frame-rate fluoroscopy with an ultra-fast digital x-ray tube based on carbon nanotube field electron emitters," in *Medical Imaging 2020: Physics of Medical Imaging*, vol. 11312. SPIE, 2020, pp. 860–865.

[9] H. Sugie, M. Tanemura, V. Filip, K. Iwata, K. Takahashi, and F. Okuyama, "Carbon nanotubes as electron source in an x-ray tube," *Applied physics letters*, vol. 78, no. 17, pp. 2578–2580, 2001.

[10] Varian Medical Systems, Inc., "PaxScan CBCT software tools: Cone beam CT tool kit," Product brochure, Document No. 76827 Rev C 02/16, 2016, IsoCal geometric calibration package and accompanying calibration phantoms.

[11] J. Hsieh, "Computed tomography: principles, design, artifacts, and recent advances," 2003.

[12] J. Starman, J. Star-Lack, G. Virshup, E. Shapiro, and R. Fahrig, "A nonlinear lag correction algorithm for a-si flat-panel x-ray detectors," *Medical Physics*, vol. 39, no. 10, pp. 6035–6047, 2012.

[13] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Journal of the Optical Society of America A*, vol. 1, no. 6, pp. 612–619, 1984.

One-Dimensional Anti-Scatter Grids for Static CT

Andreas Heinkele, Julien Erath, Lukas Hennemann, Johan Sunnegaardh, Christian Hofmann, Martin Petersilka, Karl Stierstorfer, and Marc Kachelrieß

***Abstract*—Static CT is a CT design consisting of a stationary detector ring and a stationary source ring. This design may offer several advantages over 3rd-generation scanners, including a potentially more compact geometry, reduced mechanical complexity, and very fast electronic switching speeds between sources, enabling extremely rapid acquisitions. Therefore, in the future, static CT may serve as a complementary system for specific use cases. However, because the design inherently prevents the use of conventional anti-scatter grids (ASGs), image degradation caused by scatter remains a major challenge. In this work, we propose the application of one-dimensional ASGs, which may reduce scatter artifacts and image noise and thereby allow for patient dose savings. Our results indicate that such ASGs could be particularly effective when scanning large objects, building scanners with high z-coverage, or operating multiple sources simultaneously.**

***Index Terms*—scatter, static CT, stationary CT, anti-scatter grid, ASG.**

I. Introduction

STATIC CT is a novel CT system architecture that has gained attention in research during the past years. Opposed to 3rd-generation scanners, they are stationary-stationary systems, meaning that they do not have a rotating gantry. Instead, they need an array of X-ray sources that enables the acquisition of a sufficient number of projections, ideally distributed over at least 180° plus fan angle. One possible technology to realize such a source array could be based on carbon nanotubes (CNTs). While clinical CT scanners today can produce very good images, they are space-consuming, expensive, and their temporal resolution is limited by the rotation speed of the gantry. Static CT systems, on the other hand, may be built in a more compact manner due to the simpler mechanical design, making them suitable for use cases with limited space such as use in mobile stroke units [1], [2]. The temporal resolution of static CT is mainly determined by source power and the electronic switching speed between sources, enabling potentially very high temporal resolution. Prospective systems might therefore also have application in cardiac imaging. Moreover, any static CT design inherently avoids ring artifacts because it is not equipped with rotary components [3], [4]. Given these advantages, static CT may serve as a complementary system to 3rd-generation scanners in the future, making it preferable in some use cases. However, the stationary design limits the use of anti-scatter grids (ASGs) since any stationary ASG could only be focused on one point in space and therefore would not be able to follow the switching source. This causes challenges regarding scatter artifacts. These artifacts can be reduced using software-based correction algorithms [5], [6]. While software-based scatter correction is part of ongoing research and recent deep learning-based methods [7]–[12] promise high accuracy, such algorithms can only account for the expectation value of the scatter, but not for quantum noise that is added by scattered radiation. The increased quantum noise leads to lower signal-to-noise ratios in the projection domain, which propagates into image domain, leading to noisier images. This image degradation could be reduced by the application of hardware-based scatter rejection. In this work, we simulate and evaluate the benefit of using one-dimensional ASGs in z-direction with regard to SNR, image noise and potential dose savings.

Andreas Heinkele and Lukas Hennemann are with the Division of X-ray Imaging and CT of the German Cancer Research Center (DKFZ), Heidelberg, Germany, the Computed Tomography Division of Siemens Healthineers AG, Forchheim Germany, and the Medical Faculty Heidelberg of the Heidelberg University, Heidelberg Germany.

Dr. Julien Erath, Dr. Johan Sunnegaardh, Dr. Martin Petersilka and Dr. Karl Stierstofer are with the Computed Tomography Division of Siemens Healthineers AG, Forchheim, Germany.

Prof. Dr. Marc Kachelrieß is with the Division of X-ray Imaging and CT of the German Cancer Research Center (DKFZ), Heidelberg, Germany, and the Medical Faculty Heidelberg of the Heidelberg University, Heidelberg Germany.

Corresponding author: Andreas Heinkele (andreas.heinkele@dkfz.de)

II. Methodology

A. Static CT Geometry

The simulated geometry is illustrated by Figure 1.

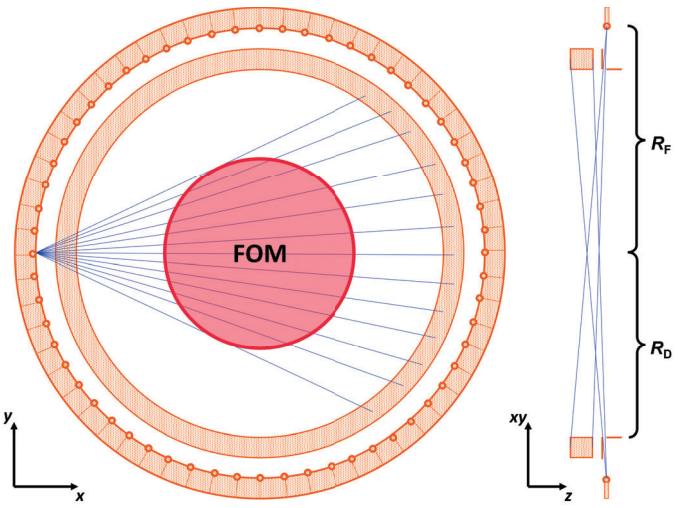


Fig. 1. Geometry of the simulated static CT in the x-y-plane and in y-z-plane. The outer ring represents the source array, the inner ring the detector. Source- and detector ring are offset in z-direction. The collimator is ring-shaped as well. The geometry is not drawn to scale.

Detector- and source-ring need to be offset in z-direction due to the ring-shaped nature. We simulate two scenarios:

1) 16-row detector: corresponds to about 1.2 cm z-coverage in the isocenter.

2) 64-row detector: corresponds to about 4.8 cm z-coverage in the isocenter. The z-shift between source and detector was increased for the 64-row geometry to enable the placement of the collimator.

B. ASG Simulations

Scatter simulations were done using our in-house Monte-Carlo (MC) software [13], [14]. Figure 2 visualizes the simulated one-dimensional ASGs. Our simulations comprise five different detector scenarios: No ASG, 5 mm ASG, 10 mm ASG, 20 mm ASG, and an extreme ASG with 40 mm height. We simulate 10^9 photons per projection. The MC estimates were then smoothed using a Gaussian filter to obtain noise-free scatter distributions. For centrally placed cylindrical water phantoms, we only simulate one projection and make use of the rotational symmetry. For more complex phantoms, we simulate projections in angular steps of 5° and interpolate in between to obtain missing projections for reconstruction. The primary signals were simulated analytically using our in-house forward projection software.

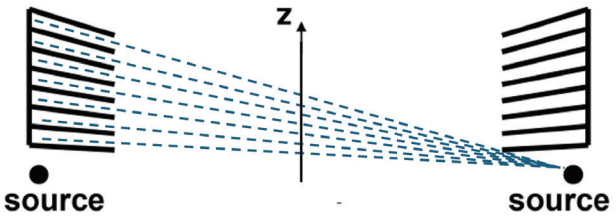


Fig. 2. One-dimensional ASG with lamellae in z-direction, illustrated in the x-z-plane. Each row within the ASG is focused towards the source 180° opposite of the detector element. This was achieved by aligning each row to a different point on the z-axis. Not drawn to scale.

C. Noise Generation

To assess the benefit of using the one-dimensional ASGs, we add noise before reconstruction. We simulate the following scenarios, with $\mathcal{N}$ being an operator that adds Poisson-distributed noise:

1) $\mathcal{N}(\mathrm{P})$: Noise is added to the scatter-free primaries P. This may be seen as a reference scenario of a 3rd-generation scanner with a very efficient ASG.
2) $\mathcal{N}(\mathrm{P{+}S})$: Noise is added to the total signal, consisting of the sum of primary radiation P and scattered radiation S.
3) $\mathcal{N}(\mathrm{P{+}S}) - \mathrm{S}$: Noise is added to the total signal. Then, the noise-free scatter is subtracted. This corresponds to an optimal software-based correction.

D. Evaluation Metrics

1) Relative Scatter-to-Primary Ratio: We define the relative scatter-to-primary ratio (SPR) as

$$\mathrm{RelativeSPR} = \frac{\mathrm{SPR}_{\text{with ASG}}}{\mathrm{SPR}_{\text{no ASG}}}.$$

2) Signal-to-Noise Ratio Improvement Factor: In an ideal case, the ASG blocks scattered radiation while being entirely translucent to primary rays. This leads to an improved SNR since quantum noise caused by scatter is avoided while maintaining the desired primary signal. For static CT, however, such an ideal ASG is not feasible. While the simulated one-dimensional ASG is perfectly aligned at the detector element 180° opposite to the active source, the alignment becomes worse for increasing angle within the X-ray fan. To account for this misalignment, we calculate an SNR improvement factor

$$\mathrm{SNR}_{\mathrm{if}} = \frac{\mathrm{SNR}_{\text{with ASG}}}{\mathrm{SNR}_{\text{no ASG}}}$$

with the SNR being defined as

$$\mathrm{SNR} = \frac{T_{\mathrm{P}} I_{\mathrm{P}}}{\sqrt{T_{\mathrm{P}} I_{\mathrm{P}} + T_{\mathrm{S}} I_{\mathrm{S}}}}$$

with T_{P} being the transmission of primary photons, I_{P} being the primary intensity, T_{S} being the transmission of scattered photons, and I_{S} being the scatter intensity. Since all these terms are functions of the fan angle and z-position at the detector, the SNR and $\mathrm{SNR}_{\mathrm{if}}$ are as well.

3) Evaluation in Image Domain: The evaluation of reconstructions allows assessing the benefit of using one-dimensional ASGs with regard to:

1) Mean absolute error (MAE): We calculate the MAE over the entire phantoms to quantify the artifacts that are introduced by scatter. This error correlates directly with the SPR. Reduction of the SPR by using an ASG will lead to reduced image artifacts.
2) Image noise σ: The image noise correlates with the SNR. Lower image noise at a constant dose is therefore equivalent to a dose saving. This correlation arises from Poisson statistics. The relative dose increase that would be needed to obtain reconstructions with an equivalent noise level is given as

$$\mathrm{RelativeDose} = \frac{\sigma^2_{\text{no ASG}}}{\sigma^2_{\text{with ASG}}}.$$

A relative dose of e.g. 180% means that 80% extra dose are avoided by using the ASG.

III. Results

Figure 4 shows the SPR as a function of ASG height for a 32 cm water cylinder. It can be observed that already an ASG height of 5 mm leads to an SPR reduction of about 20% for the 16-row detector and about 30% for the 64-row detector. At 20 mm, the SPR is below 60% for 16 rows and below 40% for 64 rows. Figure 5 illustrates the $\mathrm{SNR}_{\mathrm{if}}$ for the same 32 cm water cylinder. The $\mathrm{SNR}_{\mathrm{if}}$ reaches its maximum at $\beta = 0°$, which corresponds to the detector pixel opposite to the active source. The maximum is approximately 1.2 and 1.05 for 16 and 64 rows, respectively. With increasing angle within the fan, the $\mathrm{SNR}_{\mathrm{if}}$ drops below 1. This occurs within the limits of the 32 cm cylinder.

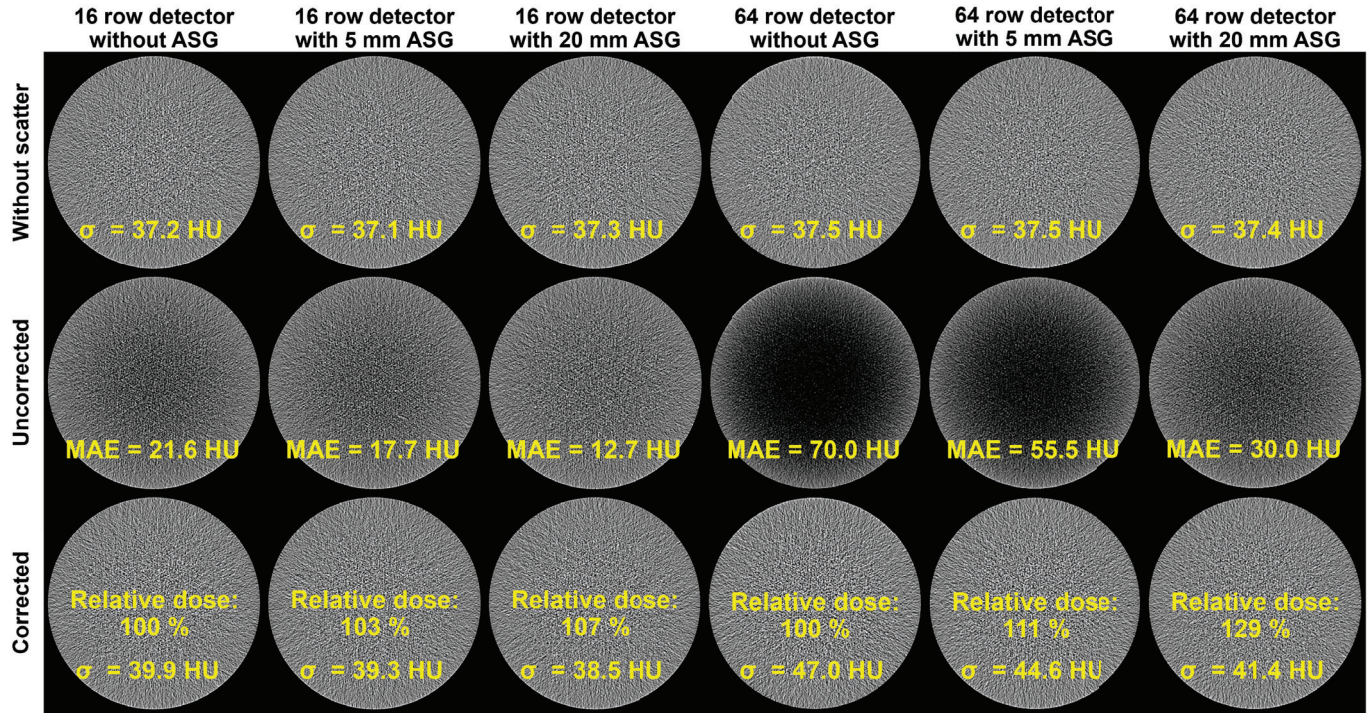


Fig. 3. Reconstructions of a 32 cm water cylinder with 16 rows (three rows on the left) and 64 rows (three rows on the right). Employing one-dimensional ASGs reduces the scatter artifacts (second row). Moreover, the noise level can be reduced compared to simulations without ASG (third row). Reconstructions: $C = 0$ HU, $W = 200$ HU

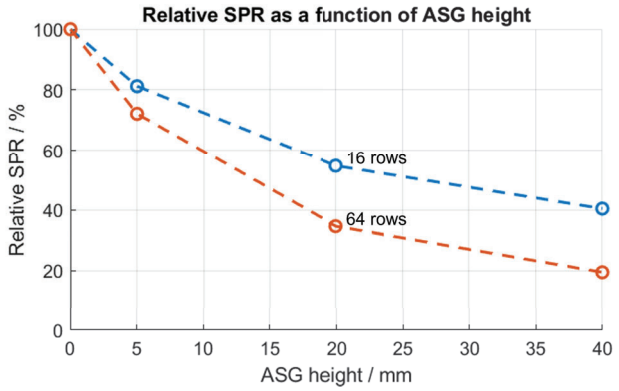


Fig. 4. Relative SPR as a function of the ASG height. 32 cm water phantom. Since the SPR correlates with the scatter artifacts in the image, it is a suitable measure to evaluate ASG performance. It can be observed that the ASG is more effective for detectors with higher z-coverage.

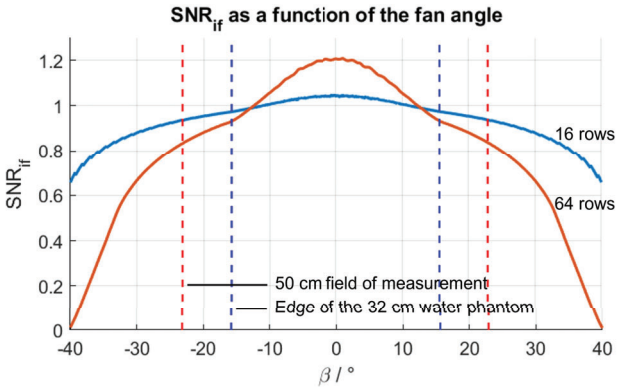


Fig. 5. $\mathrm{SNR_{if}}$ as a function of the angle β in the fan. It can be seen that the $\mathrm{SNR_{if}}$ has its maximum at $\beta = 0°$, where the ASG is exactly aligned to the opposite source. It drops with increasing angle within the fan until it eventually falls below 1. Moreover, it can be observed that the simulated ASG yields a higher $\mathrm{SNR_{if}}$ for the 64-row detector with a maximum of 1.2 compared to 1.05 for the 16-row detector.

Figure 3 shows the corresponding reconstructions. The scatter artifacts decrease consistently with the calculated relative SPRs for 16 rows and 64 rows. Slight deviations may stem from measuring the MAE over the entire phantom instead of only in the object center. After scatter correction with the simulated ground truth, reconstructions yield an increased noise level. This increase is more prominent for the 64-row scanner than for the 16-row scanner and can be reduced by employing the one-dimensional ASGs. For 16 rows, the 20 mm ASG yields a relative dose of 107%, for 64 rows, the relative dose at 20 mm ASG is 129%. Note that the noise level for relative dose calculation was averaged over the entire phantom and hence the saved dose may vary with respect to the position relative to the object center.

Figure 6 shows the impact of object size on ASG performance. Scatter artifacts are reduced for all three objects. For the 16 cm cylinder, the MAE is reduced to approximately 83%, for the 32 cm cylinder to approximately 59%, and for the 40×30 cm ellipse to approximately 58%. The noise level decreases when using the ASG. The last scenario we simulated was with multi-source operation, where three sources fire simultaneously. Figure 7 shows results. Numbers can be taken from the figure.

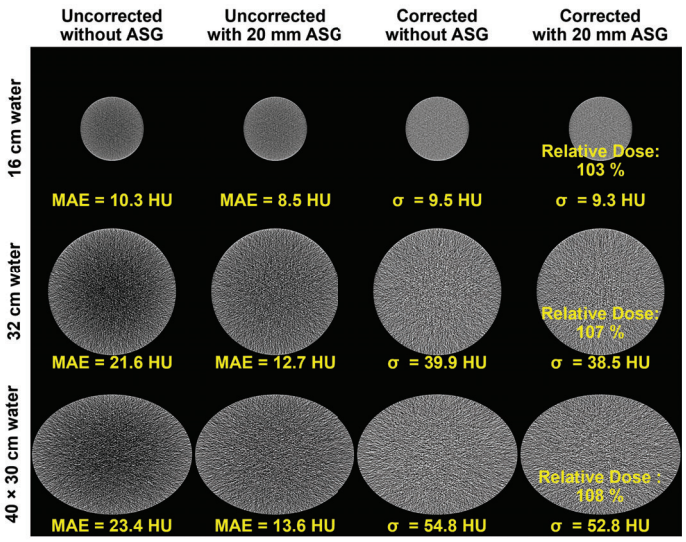


Fig. 6. Comparison of ASG performance in dependence of the object. The ASG reduces the MAE and image noise in all scenarios. However, more extra dose can be avoided for bigger objects. Reconstructions: C = 0 HU, W = 100 HU.

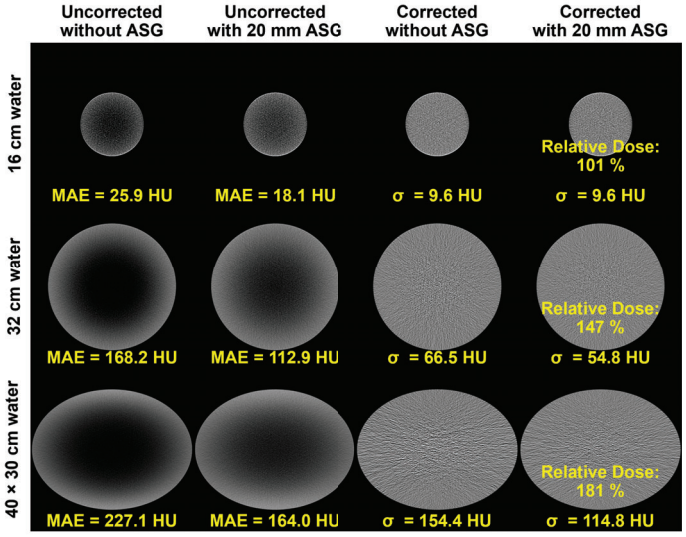


Fig. 7. Evaluation of ASG performance in multi-source operation. The ASG reduces the MAE, however, strong artifacts remain. The ASG can avoid extra dose especially for bigger objects (relative dose 181% for the 40 × 30 cm ellipse). Reconstructions: 16 cm cylinder: C = 0 HU, W = 100 HU, 32 cm cylinder: C = 0 HU, W = 600 HU, 40 × 30 cm ellipse: C = 0 HU, W = 800 HU.

IV. Discussion and Conclusion

In this work, we simulated one-dimensional ASGs with lamellae in z-direction. We compared ASG performance for detectors with 1.2 cm and 4.8 cm z-coverage, dose savings for single source operation. For multi-source simulations, dose savings depended strongly on phantom size, avoiding over 80% additional dose for the 40 × 30 cm ellipse. However, strong scatter artifacts remained visible for multi-source operation even with a 20 mm ASG. In all simulated scenarios, additional software-based scatter correction would always be necessary to mitigate residual scatter artifacts.

References

[1] T. McSkimming, A. Lopez-Montes, A. Skeats, C. Delnooz, B. Gonzales, K. J. Reynolds, W. Zbijewski, E. Perilli, and A. Sisniega, "Geometry adaptive projection-domain deep scatter estimation for multi-source semi-stationary cone-beam computed tomography," *Med. Phys.*, vol. 53, no. 1, Dec. 2025.

[2] T. McSkimming, A. Lopez-Montez, A. Skeats, C. Delnooz, B. Gonzales, E. Perilli, K. Reynolds, J. H. Siewerdsen, W. Zbijewski, and A. Sisniega, "Adaptive kernel-based scatter correction for multi-source stationary CT with non-circular geometry," *SPIE Medical Imaging Proc.*, p. 33, Apr. 2023.

[3] C. H. McCollough and P. S. Rajiah, "Milestones in CT: Past, present, and future," *Radiology*, vol. 309, no. 1, Oct. 2023.

[4] R. A. Schulz, J. A. Stein, and N. J. Pelc, "How CT happened: the early development of medical computed tomography," *Journal of Medical Imaging*, vol. 8, no. 05, Oct. 2021.

[5] E.-P. Rührnschopf and K. Klingenbeck, "A general framework and review of scatter correction methods in X–ray cone–beam computerized tomography. part 1: Scatter compensation approaches," *Med. Phys.*, vol. 38, no. 7, pp. 4296–4311, Jul. 2011.

[6] ——, "A general framework and review of scatter correction methods in X–ray cone beam CT. part 2: Scatter estimation approaches," *Med. Phys.*, vol. 38, no. 9, pp. 5186–5199, Sep. 2011.

[7] J. Erath, J. Magonov, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep scatter estimation for coarse anti-scatter grids as used in photon-counting CT," *SPIE Medical Imaging Proc.*, vol. 12304, Feb. 2025.

[8] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (DSE): accurate real–time scatter estimation for x–ray CT using a deep convolutional neural network," *Journal of Nondestructive Evaluation*, vol. 37, no. 3, p. 57 (9 pages), Jul. 2018.

[9] J. Maier, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. Kachelrieß, "Real–time scatter estimation for medical CT using the deep scatter estimation: method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation," *Med. Phys.*, vol. 46, no. 1, pp. 238–249, Nov. 2019.

[10] L. Hennemann, J. Erath, A. Heinkele, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep bowtie and patient scatter correction applied to clinical photon-counting CT," *SPIE Medical Imaging Proc.*, vol. 13405, Apr. 2025.

[11] A. Heinkele, J. Erath, L. Hennemann, J. Maier, E. Fournié, J. Sunnegaardh, C. Hofmann, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep scatter estimation for static CT using multiple projections," *SPIE Medical Imaging Proc.*, vol. 13405, Apr. 2025.

[12] L. Hennemann, J. Erath, A. Heinkele, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Spectral deep learning-based patient and bowtie scatter correction for clinical photon-counting CT," *Med. Phys.*, 2026, under revision.

[13] K. Stierstorfer, "Modeling the frequency-dependent detective quantum efficiency of photon-counting x-ray detectors," *Med. Phys.*, vol. 45, no. 1, pp. 156–166, 2018.

[14] K. Stierstorfer, M. Hupfer, and N. Köster, "Modeling the DQE (f) of photon-counting detectors: impact of the pixel sensitivity profile," *Physics in Medicine & Biology*, vol. 64, no. 10, p. 105008, 2019.

Foveated-Imaging Geometry CT Architecture and Prior-Guided Diffusion Model Enabling Global Super-Resolution Reconstruction

Wenxin Mo, Yingxian Xia, Hao zhou, Hewei Gao

Abstract—It is well-known that smaller CT detector elements usually lead to better spatial resolution of a CT scan but with inevitably significant increase in system costs. However, in many clinical applications, only regions of interest are particularly important, making universally small detector element size unnecessary. In this work, we propose the concept of Foveated-Imaging Geometry CT (FIGCT) architecture, which can balance the cost of a system and its high resolution imaging performance, and develop a global super-resolution image reconstruction algorithm by leveraging a prior-guided diffusion model for FIGCT. By arranging high-resolution detectors (HRD) in the central region and low-resolution detectors (LRD) on both sides, FIGCT can simultaneously acquire HR projections for the Region of Interest (ROI) and LR projections covering the entire field of view (FOV). Diffusion Probabilistic Foveated-imaging geometry CT Super Resolution Reconstruction (DPFSR) was proposed to generate global high-resolution image for FIGCT. AAPM low dose CT grand challenge dataset and a pig lung data from helical CT is conducted to validate super resolution performance of DPFSR combined with FIGCT. Experimental results demonstrate that the FIGCT architecture achieves high-precision reconstruction within the central ROI and the performance of DPFSR on global image super resolution.

Index Terms—FIGCT, super resolution, Diffusion model, deep learning

I. Introduction

HIGHER spatial-resolution has long been a pursuit in CT imaging, as it directly contributes to improved diagnostic accuracy and the detection of subtle pathological changes. To prevent increasing the radiation dose to patients, obtaining HR CT images often requires expensive CT hardware using detector with smaller pixel. However, they also introduce substantial increases in hardware cost and complexity. To obtain HR CT images within a limit system budget, two common approaches are hardware and algorithm based.

From a hardware perspective, the spatial resolution of CT is primarily determined by the aperture of detector cells, the X-ray focal spot size and geometrical magnification[1]. In general, the X-ray focal spot size is constrained by tube power and heat dissipation limitations, leaving other aspects as the primary technique to improve resolution. Early multi-slice CT systems adopte an adaptive array detector design, in which detector elements with different size are arranged along the z direction. This allowe high spatial resolution in the central slices while simultaneously providing a larger longitudinal field of view[2]. In clinical CT, only specific regions are of interest, such as cardiac imaging and image-guided radiotherapy. Such designs suggests that non-uniform sampling geometries may better align with clinical requirements and cost-control considerations.

Algorithmic methods can recover HR images from LR inputs, which doesn't need additional hardware cost. Traditional super-resolutioin (SR) algorithms encompass model-based iterative reconstruction methods and learning-based approaches including dictionary learning and sparse representation[3]. Recently, deep learning-based methods have been introduced to CT images super-resolution. By effectively extracting and utilizing the information embedded in CT images, neural network is capable of producing high-fidelity images with rich anatomical details. More recently, denoising diffusion probabilistic models(DDPMs) has been gaining increasing popularity[4], [5]. DDPMs learn to gradually remove various level of noise from output, enabling the transformation from a Gaussian noise to an arbitrary distribution. In SR task, cDDPM have emerged as a powerful alternative to traditional convolutional neural networks (CNNs) and generative adversarial networks (GANs). However, generative models can cause another type of error called hallucinations[6], which may lead the model to generate details that deviate from those in the original image.

In this paper, we presented a hardware and algorithmic combined approach to achieve high-resolution imaging at a reduced cost. This contribution consists of two parts:

Firstly, we presented the Foveated Imaging Geometry CT(FIGCT) architecture that combines HRD at the center with LRD on both sides, enabling higher density sampling in ROI. Multiplexing-Splitting Filtering Back-Projection(MS-FBP)[7] was proposed to reconstructed the multi-resolution data, which ensures high-resolution imaging of the interior region.

Secondly, we developed a novel DDPM-based super-resolution technique: Diffusion Probabilistic Foveated Imaging CT Super Resolution Reconstruction(DPFSR). DPFSR leverages HR segment of projection data during the inverse diffusion process to reconstruct high-resolution images that preserve fine anatomical details, which achieves super-resolution in the exterior region. Our method is able to recover image details that are more

This project was supported in part by the National Key R&D Program of China (No. 2024YFC2417500)

realistic and reliable guided by real high-resolution data. The proposed model was validated on AAPM low dose CT grand challenge dataset, demonstrate the robustness and generalization capability of our method.

II. Methods

A. FIGCT-type architecture

1) The Concept of FIGCT: FIGCT employs a hybrid detector architecture composed of HRD and LRD. The system uses only two pixel sizes, which reduces modeling complexity and facilitates image reconstruction. This architecture enables the simultaneous acquisition of both HR and LR projection P_H, P_{L_1}, P_{L_2} within single scan. For simplicity, a symmetric arrangement is used here, so the reconstruction naturally separates into an interior HR ROI and an exterior LR region. HRD is placed at the center, while LRD are positioned on both sides.

As shown in Fig.1, points within the ROI are scanned by the HRD over the full angular range, enabling high-resolution reconstruction. In contrast, points outside the ROI are only scanned by the HRD at limited angles. In medical diagnostics, critical anatomical information is often concentrated within specific ROI. Consequently, HR reconstruction can be selectively prioritized within these critical regions, while LR imaging suffices for the surrounding exterior areas to maintain contextual information. A example of FIGCT system was described in Fig.1(a).

To describe the extent of coverage provided by the HRD, we define the HR Coverage Ratio (HCR) r and Ratio of LRD and HRD element size (RLH) k as key parameters in FIGCT:

$$r = \frac{\beta_h}{\beta_f} \tag{1}$$

$$k = \frac{\Delta d_l}{\Delta d_h} \tag{2}$$

where β_h denotes the fan angle of the HRD, β_f represents the full fan angle of the entire detector, Δd_l denotes the pixel size of the LRD and Δd_h represents the pixel size of the HRD.

2) Sampling Characteristics and Reconstructed-Image Properties: We utilize equiangular fan-beam geometry as an example. The reconstructed image and sinogram of two points are shown in Fig.1(b). Point B in the interior region of the image are scanned by the HRD at full angles, while point A in the exterior region are scanned only over certain angles. This inherent sampling characteristic gives rise to a foveated-imaging characteristic, enabling HR detail recovery in the interior region while yielding reduced spatial resolution in the exterior region.

B. DPFSR: diffusion-based super resolution reconstruction algorithm for FSCT

HR prior is embedded in projection. Motivated by this, we propose DPFSR for FIGCT, which generates images with richer details and higher quality by using the HRD projection as a prior. An overview of DPFSR, which is based on cDDPM, is provided in Fig. 2.

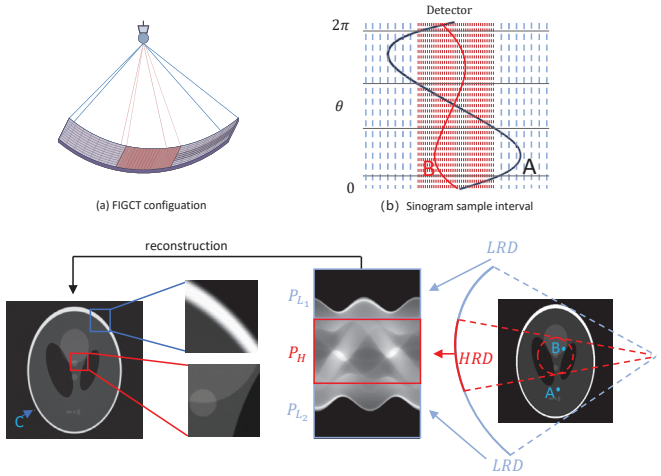


Fig. 1: FIGCT sampling characteristics and reconstructed image properties. FIGCT acquires 3 incomplete mismatched sinograms. Point B at the center region is fully sampled by high-resolution projections, while point A in the exterior region is only sampled at partial angles.

1) Conditional DDPM: Conditional DDPM(cDDPM) refers to DDPMs that samples with guided image, which consists of two parts: forward process and reverse process. The forward process gradually adds noise to an image, resulting in a series $x_0 \to x_1 \to \cdots \to x_T$, where the noise level progressively increases from the clean image x_0 to the pure noise image x_T, thereby transforming an arbitrary data distribution into a Gaussian distribution. For reverse process, cDDPM learns to generate HR images by modeling the reverse Markov process $p_\theta(x_{t-1}|x_t, f)$, conditioned on noisy inputs x_t and corresponding LR image f. In order to approximate $p_\theta(x_{t-1}|x_t, f)$, a neural network was train to learn to predict the noise $\tilde{\epsilon}_\theta(x_t, f, t)$ on x_t with the condition f and x_t. During sampling, we can infer x_{t-1} with the noise predicted by network.

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\sqrt{\alpha_t}}{\sqrt{1-\gamma_t}}\tilde{\epsilon}_\theta(x_t, f, t)) + \sigma_t z \tag{3}$$

where $z \sim \mathcal{N}(0, I)$.

Since the network f_θ is trained to estimate the noise ϵ , we can predict x_0 given x_t from definition. To distinguish the estimate of x_0 at each time step, we denote it by $x_{0|t}$.

$$x_{0|t} = \frac{1}{\sqrt{\gamma_t}}(x_t - \sqrt{1-\gamma_t}f_\theta(x_t, f, t)) \tag{4}$$

From [5], the posterior distribution of x_{t-1} given x_0, x_t can be derive as:

$$x_{t-1} = \frac{\sqrt{\gamma_{t-1}}\beta_t}{1-\gamma_t}x_0 + \frac{\sqrt{\alpha_t}(1-\gamma_{t-1})}{1-\gamma_t}x_t + \sigma_t z \tag{5}$$

where, $\sigma_t^2 = \frac{(1-\gamma_{t-1})(1-\alpha_t)}{1-\gamma_t}$.

2) HR seed guidance: Since the region outside the ROI was also scanned by HRD, the HR information is embedded in HR segment of projection. the reverse diffusion process is modeled as sampling from the conditional

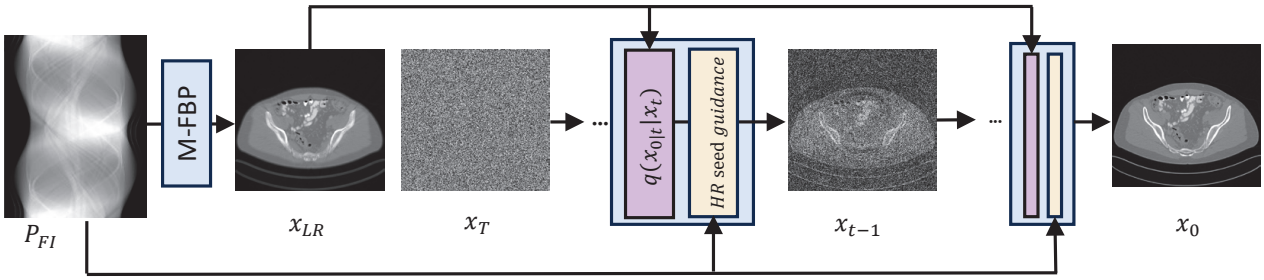


Fig. 2: Overall pipeline of DPFSR. MS-FBP first pre-reconstructs an image as the condition for cDDPM. A HR seed guided constraint step is performed at each reverse diffusion step.

Algorithm 1: DPFSR

Input: P_{L_1}, P_{L_2}: LR projections, P_H: HR segment, f: reconstructed FI image, $\tilde{\epsilon}_\theta$: Adjusted denoiser network

Sample $x_T \sim \mathcal{N}(0, I)$; // Run diffusion sampling

for $t = T, T-1, \ldots, 1$ do

$z \sim \mathcal{N}(0, I)$;

$x_{0|t} = \frac{1}{\sqrt{\gamma_t}} \left(x_t - \sqrt{1-\gamma_t}\, \tilde{\epsilon}_\theta(x_t, f, t)\right)$;

$\hat{x}_{0|t} = \mathrm{HRc}\left(x_{0|t}, P_H, f\right)$;

$x_{t-1} = \frac{\sqrt{\gamma_{t-1}}\,\beta_t}{1-\gamma_t}\, \hat{x}_{0|t} + \frac{\sqrt{\alpha_t}\,(1-\gamma_{t-1})}{1-\gamma_t}\, x_t + \sigma_t z$;

Output: x_0: SR image

distribution $p_\theta(x_{t-1}|x_t, P_H, f)$. As P_H is obtained from original data, the forward projection of the final x_0 is expected to be consistent with P_H.

Inspired by the recent successes of cDDPM[8], [9], DPFSR is proposed to leverage HR segment of projection as seed to guide generation of SR images with a well trained cDDPM during reverse diffusion step. HR seed guidance is achieved by replacing the corresponding HR segment region with the original HR segment, followed by reconstruction step using FBP. The overall step of HRc$(\cdot)$ is given below.

$$\hat{P}_{0|t} = (1 - m_p) \odot \mathcal{P}(x_{0|t}) + m_p \odot P_{H_{int}} \tag{6}$$

where m_p represents mask of the HR segment of projection.

$$\hat{x}_{0|t} = (1 - m_i) \odot \mathcal{F}(\hat{P}_{0|t}) + m_i \odot f \tag{7}$$

Where m_i represents mask for HR region in image domain, $\mathcal{C}(\cdot)$ denotes concatenation along the detector direction,the exterior operator Ext$(\cdot)$ denotes the extraction of the LR part from original FIGCT projection.

III. Experiments and Results

1) Reconstruction accuracy analysis: To evaluate the accuracy of reconstruction image, clinical image from AAPM Low Dose CT Grand Challenge was utilized to numerical simulation. For parameters of FIGCT, Source-Detector distance (SDD) was set to 1000 and Source-iso distance (SID) was set to 625. The element size of HRD was set to 0.5 mm, whereas RLH k was set to 4 as analyzed above. For each simulation, a total of 720 fan-beam projections were acquired over 360 degrees.

Reconstruction result is presented in Fig.4. It is observed that HRD could fully cover phantom while $r = 60\%$.The white circle indicates the region that is fully scanned and reconstructed at HR region. The PSNR and SSIM performance within white circle and overall image were presented at the corner of each image. It can observed that MS-FBP allows for high-quality reconstruction of the high-resolution regions. Specifically, in the central high-resolution area, the method achieves excellent quantitative metrics with a PSNR above 90 dB and an SSIM approaching 1.

2) SR performance study: To evaluate our proposed DPFSR, we used the AAPM Low Dose CT Grand Challenge (LDCT) dataset for simulation, which contains 10 normal-dose CT scans. Images were simulated at 70 keV and converted to sinograms using the parameters above. Poisson noise was added after photon attenuated by phantom. Foveated images were reconstructed with M-FBP, and HR images were generated in an equiangular fan-beam geometry with the parameters of $r = 100\%$ and reconstructed with FBP.

In this study, we compared DPFSR with state-of-art methods: U-Net, EDSR, GAN-CIRCLE, ESRGAN, cDDPM. The comparison methods only was trained $r = 10\%$ cases. Results shows that compared with other approaches, CNN-based methods generally obtain higher PSNR and SSIM scores. cDDPM achieves higher scores in LPIPS. As mentioned above, the spatial resolution of Foveated-image is shift-variant. From the visual results in Fig.3, GAN-based and PSNR-oriented method is capable of producing mages with higher resolution, since the input images themselves already contain some HR components. But the spatially variant characteristics are still preserved in the SR images. Compared with the above methods, cDDPM can generate SR images with visual quality comparable to the HR images. However, it fails to accurately

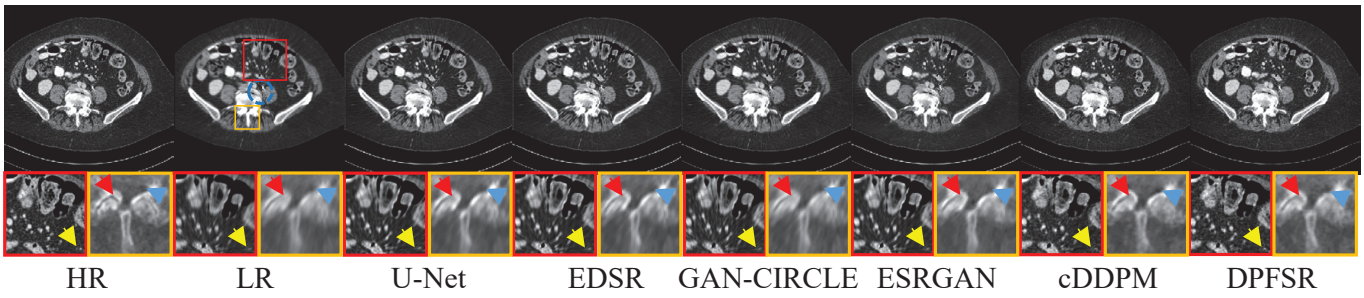


Fig. 3: Visual comparison of $r = 10\%$ case and $r = 0\%$ cases with different methods.The display windows is [-160,240] HU for whole image and ROI1, [-500,1000] HU for ROI2. Region in blue circle represents HR area, while anatomical features enclosed by the yellow and red boxes are zoomed in.

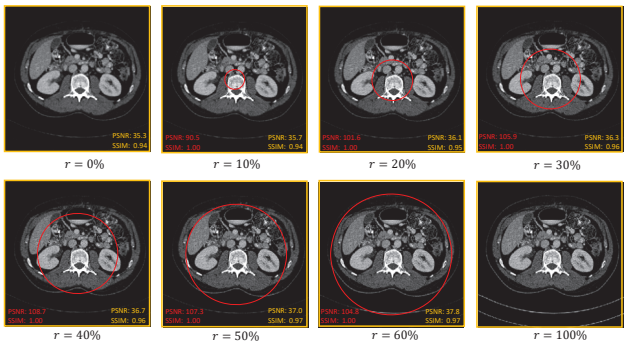


Fig. 4: Reconstruction result of different HCR r with k fixed to 4. Region in red circles were HR area of Foveated-image. The metric in the lower-left corner is computed between the red-circled region and the high-resolution (HR) reference, while the one in the lower-right corner is calculated over the entire image. The display windows is [-100,400] HU.

recover some fine structures, whereas DPFSR is able to reconstruct these details more clearly. As indicated by the yellow arrow, a small vessel is better preserved; the red arrow highlights the bone structure; and at the location marked by the blue arrow, cDDPM produces an undesired dark artificial structure that is effectively suppressed by DPFSR.

IV. Conclusion

In this paper, we introduced the concept of FIGCT, a novel imaging paradigm capable of simultaneously acquiring HR images of the ROI and LR images of the global background. FIGCT holds great promise for practical deployment in various CT system designs. In multi-row detector CT systems, detector elements with varying physical sizes can be employed within a single detector array. In flat-panel detector based cone-beam CT and photon-counting CT, FIGCT can be implemented by applying different binning modes to different detector regions to realize FIGCT. We further proposed the DPFSR framework to address the reconstruction challenge. By leveraging the physical consistency of HR segment of projections as constraints, our method successfully achieves global image super-resolution. Extensive experimental results have validated the effectiveness of the proposed method, demonstrating its potential for efficient and high-quality CT imaging.

References

[1] A. M. Hernandez, P. Wu, M. Mahesh, J. H. Siewerdsen, and J. M. Boone, "Location and direction dependence in the 3d mtf for a high-resolution ct system," Medical Physics 48(6), pp. 2760–2771, 2021.

[2] M. Prokop, "General principles of MDCT," European Journal of Radiology 45, pp. S4–S10, Mar. 2003.

[3] I. Steven Tilley, W. Zbijewski, and J. W. Stayman, "High-fidelity modeling of shift-variant focal-spot blur for high-resolution ct," Int'l Mtg. Fully 3D Image Recon. in Radiology and Nuc. Med .

[4] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in Advances in Neural Information Processing Systems, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds., 33, pp. 6840–6851, Curran Associates, Inc., 2020.

[5] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," IEEE Transactions on Pattern Analysis and Machine Intelligence 45(4), pp. 4713–4726, 2023.

[6] M. Tivnan, S. Yoon, Z. Chen, X. Li, D. Wu, and Q. Li, "Hallucination Index: An Image Quality Metric for Generative Reconstruction Models," in Medical Image Computing and Computer Assisted Intervention – MICCAI 2024, M. G. Linguraru, Q. Dou, A. Feragen, S. Giannarou, B. Glocker, K. Lekadir, and J. A. Schnabel, eds., pp. 449–458, Springer Nature Switzerland, (Cham), 2024.

[7] H. Gao and Y. Xia, "A novel coplanar multi-modality tomographic imaging for image guidance in radiotherapy using hybrid radiation detector," in MEDICAL PHYSICS, 47(6), pp. E570–E570, WILEY 111 RIVER ST, HOBOKEN 07030-5774, NJ USA, 2020.

[8] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "RePaint: Inpainting using Denoising Diffusion Probabilistic Models," in 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 11451–11461, IEEE, (New Orleans, LA, USA), June 2022.

[9] J. Liu, R. Anirudh, J. J. Thiagarajan, S. He, K. A. Mohan, U. S. Kamilov, and H. Kim, "DOLCE: A Model-Based Probabilistic Diffusion Framework for Limited-Angle CT Reconstruction," in 2023 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10464–10474, IEEE, (Paris, France), Oct. 2023.

PINN-Driven Scatter Correction for Multi-source Linear-array CT

Yankun Wang, Yule Li, Junhao Wu, Yuanxiang Gao, Haomin Hou, Deyang Kong, Limin Li
Jingyi Liao, Hao Wang, Jianhua Ma, Dong Zeng

***Abstract*—Multi-source linear static computed tomography (SLCT) achieves rapid scanning without mechanical rotation through a linearly arranged distributed ray source array, offering advantages of simple structure, compact size, and high temporal resolution. However, its non-centrosymmetric geometry causes significant edge source displacement, resulting in severe asymmetric distortion of the scatter point diffusion function. Traditional scattering correction methods based on symmetric kernels are prone to systematic errors, compromising projection data accuracy and reconstructed image quality. To address this challenge, this paper proposes a physics-informed neural network (PINN)-driven adaptive scattering correction framework. This framework introduces explicit geometry-aware correction terms to dynamically compensate for path anisotropy and forward scattering bias induced by source displacement. Concurrently, it constructs a PINN network to learn nonlinear relationships between hyperparameters and geometric variables, incorporating a physics-consistent loss to enhance the interpretability of scattering kernels for spatial scattering distributions. Simulation experiments on a cylindrical homogeneous water phantom validate that our method successfully suppresses asymmetric scattering artifacts under SLCT geometry while improving reconstructed CT value accuracy. This adaptive correction strategy, integrating physical modeling with deep learning, overcomes the poor adaptability of traditional methods to non-standard geometries, laying an algorithmic foundation for high-fidelity imaging in static CT.**

***Index Terms*—Multi-source Linear-array CT; Static CT; Scatter correction; Physical information neural network.**

This work was supported in part by the National Key R&D Program of China under Grant 2024YFA1012000 and 2024YFC2417800, the National Natural Science Foundation of China under Grant 62571228, Beijing Natural Science Foundation under Grant Z250002, and Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515140146. (Corresponding author: Hao Wang, Dong Zeng.)

Yankun Wang, Yule Li, Junhao Wu, Yuanxiang Gao, Haomin Hou, Deyang Kong, Jingyi Liao, Hao Wang and Dong Zeng are with the School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China. (e-mail: ykwang@smu.edu.cn, ylli@smu.edu.cn, wujunhao@smu.edu.cn, gyx2001@smu.edu.cn, houhm@smu.edu.cn, kongdy@smu.edu.cn, 1jy@smu.edu.cn, wang_hao@smu.edu.cn, zd1989@smu.edu.cn)

Limin Li is with the School of Mathematics and Statistics in Xi'an Jiaotong University, Xi'an 710049, China. (e-mail: liminli@xjtu.edu.cn)

Jianhua Ma is with the School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China. (e-mail: jhma@smu.edu.cn)

I. INTRODUCTION

COMPUTED tomography (CT) technology, as a core non-invasive detection method, plays an irreplaceable role in clinical medical diagnosis and high-precision industrial non-destructive testing. Traditional cone-beam CT (CBCT) systems primarily rely on a rotating gantry to drive the X-ray source and detector through a 360° circular scan to acquire multi-angle projections. However, this mechanical rotation mode is constrained by physical inertia, limiting further reduction in scan cycles. Additionally, geometric instability caused by mechanical wear and patient motion artifacts resulting from prolonged acquisition times remain persistent bottlenecks for high-fidelity imaging. In recent years, breakthroughs in carbon nanotube (CNT) field-emission source technology, with its ultra-fast electron switching capability, have enabled the development of SLCT. SLCT employs a fixed, linearly distributed beam source array for electronically controlled acquisition, completely eliminating the need for complex rotating gantries. This architecture significantly enhances scanning speed, simplifies equipment size, and greatly improves temporal resolution. It demonstrates remarkable application potential in time-sensitive scenarios such as emergency stroke examinations and rapid chest imaging.

However, the unique non-centrosymmetric geometric layout of SLCT poses unprecedented physical challenges for scatter correction [1]. Unlike the relatively stable symmetric scatter field in conventional CBCT, the SLCT radiation source exhibits linear distribution with significant lateral displacement at its peripheral regions [2]. This unique geometry causes primary rays to strike the detector at steep angles, altering the propagation paths of scattered photons. Consequently, the point spread function (PSF) of scattered photons exhibits extreme non-stationarity and pronounced anisotropy [3].

This geometry-induced distortion of the scatter field introduces severe low-frequency scatter artifacts, stripe artifacts, and extensive quantitative CT value shifts in reconstructed images, significantly compromising diagnostic accuracy [4]. Current correction strategies for such issues remain generally limited in SLCT scenarios: Traditional analytical models struggle to capture complex scattering patterns due to the lack of explicit physical descriptions of geometric distortions [5]. Purely data-

driven deep learning methods, while demonstrating strong fitting potential, face severe interpretability challenges due to their "black-box" mechanisms and often exhibit insufficient generalization when dealing with different phantom sizes or heterogeneous geometric configurations. Although prior studies attempted to compensate for geometric bias by introducing position correction terms [6], their core parameters typically rely heavily on cumbersome manual empirical fitting [7]. This makes achieving robust adaptive correction in dynamically changing imaging environments or complex lesion detection challenging.

To this end, this paper proposes a PINN-driven adaptive scattering correction framework. Building upon the geometric correction strategy, this framework employs a PINN architecture to perform end-to-end adaptive optimization of model parameters [8]. Through a carefully designed composite loss function, the framework simultaneously constrains data consistency and physical laws, achieving precise physical descriptions of non-stationary scattering fields. This dual-driven approach physically and data-driven compensates for the geometric limitations of traditional models while overcoming deep learning's reliance on large datasets. Monte Carlo water phantom simulations demonstrate that this method significantly suppresses asymmetric scattering artifacts, substantially improving CT value accuracy and contrast in reconstructed images [9]. This study not only validates the effectiveness of PINN in non-standard geometric CT systems but also provides critical technical support for the clinical quantitative application of SLCT [10].

II. METHODS

A. Geometric Analysis of SLCT Scatter Field

In traditional CBCT, the X-ray source and detector maintain a perfectly centered geometric alignment. This coaxial characteristic ensures that the mean free path of scattered photons remains highly consistent in the radial direction as they traverse the patient, resulting in a PSF that approximates an isotropic, radially symmetric distribution [11].

However, the linear array layout of SLCT systems disrupts this physical symmetry. When radiation sources at the array edges are activated, the emitted photon beams enter the patient with significant oblique angles, causing substantial lateral displacement on the detector plane. From a microscopic physics perspective, this non-centrosymmetric geometry causes the attenuation path length of scattered photons within the body to exhibit highly nonlinear variations depending on the incident position: on the side closer to the beam source offset direction, the exit path length after photon scattering is shorter, resulting in less energy attenuation; conversely, the exit path length for photons on the opposite side is significantly elongated. This leads to severe imbalance in the scattered energy spectrum distribution and intensity across the detector plane. Fig. 1 illustrates the differences in geometric structure and scattering distribution characteristics between the two designs.

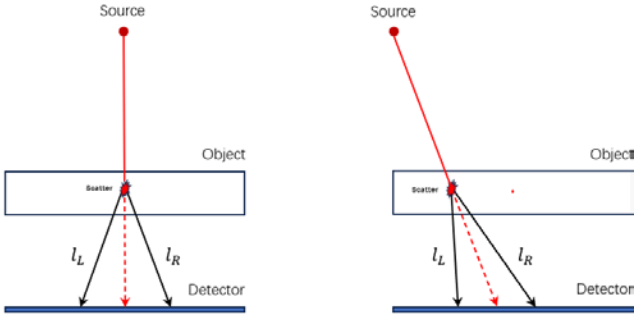


Fig. 1. (a) CBCT scatter distribution; (b) SLCT scatter distribution

B. Baseline Scatter Kernel Modeling

This paper constructs a scattering kernel model based on the fASKS algorithm: the method discretizes and groups projection data according to water equivalent thickness [12], modeling the total signal by superimposing the scattering contributions from each group:

$$I_s(x,y) = \sum_i \left(\iint_D I_p(x',y') R^i(x',y') c^i(x',y') g^i(x-x',y-y') \times \left(1 - \gamma \cdot (\tau(x,y) - \tau(x',y'))\right) dx'\,dy' \right), \quad (1)$$

where (x,y) and (x',y') are coordinates on the detector. I_s and I_p represent scattered and primary signals, respectively. The integration region D is the exposure area on the detector, known in Monte Carlo simulations and used to solve for scattered nuclear hyperparameters.

To address non-uniform object thickness, fASKS employs group-based modeling based on water equivalent thickness, configuring parameters independently for each group. This approach uses multiple group functions to cover substantial thickness variations, significantly enhancing fitting accuracy:

$$R^i(x',y') = \begin{cases} 1, & t_i \le \tau(x,y) < t_{i+1} \\ 0, & otherwise \end{cases}, \quad (2)$$

$$\tau(x,y) \approx \frac{1}{\mu} \cdot \ln\left(\frac{I_o(x,y)}{I_p(x,y)}\right), \quad (3)$$

where I_o is the empty sweep signal and I_p is the primary signal, μ is the water equivalent attenuation coefficient, i is the thickness group index, and N is the total number of thickness groups.

For each thickness group, the scattering kernel consists of an amplitude factor c and a shape function g. The amplitude factor c captures the nonlinear dependence between scattering intensity and the primary signal:

$$c(x',y') = \mathrm{A}\cdot\left(\frac{I_p(x',y')}{I_o(x',y')}\right)^{\alpha}\cdot\left(\ln\left(\frac{I_o(x',y')}{I_p(x',y')}\right)\right)^{\beta}, \quad (4)$$

where A, α, and β are fitting parameters, with the primary signal denoted as I_p and the empty scan signal as I_o.

The shape function g describes the spatial distribution of the PSF, represented by weighting two concentric, rotationally symmetric Gaussian functions:

$$g(x-x',y-y') = exp\left(-\frac{(x-x')^2+(y-y')^2}{2{\sigma_1}^2}\right) + Bexp\left(-\frac{(x-x')^2+(y-y')^2}{2{\sigma_2}^2}\right), \quad (5)$$

where B, σ_1, and σ_2 are fitting parameters, with σ_1 and σ_2 expressed in cm.

C. Asymmetric Geometry Correction Term

To accommodate local thickness variations in objects, fASKS incorporates an object-thickness-dependent disturbance term, yielding the adaptive kernel h_{af}:

$$h_{af} = c(x',y')g(x-x',y-y')\left(1-\gamma\cdot\left(\tau(x,y)-\tau(x',y')\right)\right), \quad (6)$$

where γ is the thickness adjustment coefficient, which controls the extent to which local thickness differences influence the scattering kernel's shape.

To overcome the symmetry constraints imposed by FASKS, this paper introduces a piecewise two-dimensional Gaussian correction term. By explicitly modeling the array orientation, it achieves an accurate representation of the SLCT's asymmetric scattering field:

$$G(x,y) = \begin{cases} k_3 * e^{\left(-\frac{((x-x')^2+(y-y')^2)}{2*{\sigma_3}^2}\right)}, & x < x' \\ k_4 * e^{\left(-\frac{((x-x')^2+(y-y')^2)}{2*{\sigma_4}^2}\right)}, & x > x' \end{cases}, \quad (7)$$

where k_3, k_4, σ_3, and σ_4 are fitting parameters. Revised Total Scattering Signal Estimation Model:

$$I_s(x,y) = \sum_i\left(\iint_D I_p(x',y')R^i(x',y')\times\left(h_{af}+G(x,y)\right)dx'\,dy'\right). \quad (8)$$

This revised model explicitly models the scatter flux imbalance and multiple scattering caused by path length differences, ensuring an accurate description of asymmetric scattering fields.

D. PINN Framework and Composite Loss Function

This paper constructs a physically embedded PINN framework, whose core lies in utilizing dual-stream coordinates (detector coordinates and ray source positions) as joint inputs, replacing the traditional single image mapping. Through a high-dimensional feature extraction layer, the network captures deep nonlinear mappings between geometric variables and scatter field distortions across projection spaces, enabling efficient adaptive solutions for highly complex hyperparameters of scattering kernels. As illustrated in Fig. 2, this architecture transforms the prior physical distribution obtained through Monte Carlo simulations into network weights. This approach ensures real-time correction efficiency while endowing the model with the ability to perceive different ray source angles.

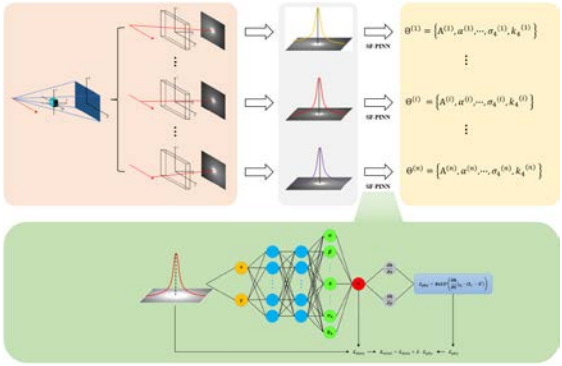


Fig. 2. PINN Data Flow Framework Diagram

To ensure that the parameter set for network learning strictly adheres to the geometry-driven scattering effects under non-circular geometry in SLCT, we constructed a composite loss function incorporating both data fidelity terms and physical geometric constraint terms:

$$\mathcal{L}_{total} = \mathcal{L}_{data} + \lambda\cdot\mathcal{L}_{phy}, \quad (9)$$

where $\mathcal{L}_{data}$ represents the data fidelity term, $\mathcal{L}_{phy}$ denotes the physical loss, and λ is the weighting factor for physical loss.

Data consistency loss employs $L1$ norm constraints to minimize discrepancies between network predictions and Monte Carlo simulation truths, ensuring global fitting accuracy:

$$L_{data} = \frac{1}{N}\sum_{x\in\Omega}\left|I_{pred}(x,y) - I_{gt}(x,y)\right|, \quad (10)$$

where I_{pred} is estimated scattered signal and I_{gt} is true scattered signal.

The physical consistency loss is based on the monomodal physical characteristics of the scattering distribution. By imposing a “geometric anticorrelation” constraint that forces the scattering gradient on the detector plane to remain geometrically aligned with the direction of the ray source offset, this loss term effectively prevents oscillatory or energy-conservation-violating prediction trends that may arise from deep learning during small-sample training. This physics-guided training mechanism ensures that the scattering kernel consistently adheres to the monomodal physical distribution near its extrema:

$$R_{offset} = (x' - x_c), \quad (11)$$

$$\mathcal{L}_{phy} = ReLU\left(\frac{\partial h}{\partial x}|_{x_c}\cdot R_{offset}\right), \quad (12)$$

where R_{offset} denotes the offset of the pencil beam's focal point x' relative to the detector center x_c.

III. EXPERIMENTS

A. Experimental Setup

In this work, we establishe a highly realistic SLCT imaging system based on the Geant4-v11.1.0 platform. In terms of hardware configuration, the ray source array consists of 31 equidistantly spaced carbon nanotube cold cathode sources, with an emission spectrum based on the publicly available 120kVp continuous multi-color spectrum from the XCIST simulation library.

A cylindrical uniform water phantom with a radius of 5 cm and height of 10 cm was employed as the detection target. To obtain high-precision physical benchmarks, pure scattered signals and primary signals in the projection space were extracted via Monte Carlo pencil beam simulations for subsequent scattering kernel fitting and training. Regarding algorithm implementation, the PINN framework was constructed using PyTorch. Training was conducted with 1000 iterations on an NVIDIA RTX 4090 GPU using the Adam optimizer to achieve adaptive solution of scattering physical parameters. Finally, all processed projections were reconstructed into 512×512 pixel images using an analytical reconstruction algorithm derived from FBP theory. The geometric parameters of the system are shown in Table I.

TABLE I
SYSTEM GEOMETRIC PARAMETERS

Parameter type	Parameter value
DSD	150 cm
DSO	100 cm
DDO	50 cm
Detector Pixel Array	1000 × 1250
Detector Pixel Size	0.04 cm × 0.04 cm
Number of radiation sources	31

B. *Evaluation Metrics*

To objectively evaluate algorithm performance, this paper uses Monte Carlo-generated reconstruction maps of the primary signal without scattering as ground truth. Root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) are employed as quantitative assessment metrics. By comparing the differences between the corrected images and the ground truth, the performance of this framework in eliminating non-stationary scattering, suppressing artifacts, and enhancing the quantitative reliability of the system is validated.

IV.RESULTS

Fig. 3 shows the reconstruction results and the corresponding error maps. As shown in Fig. 3, the uncorrected image exhibits severe low-frequency scattering artifacts due to the non-centrosymmetric geometry of the SLCT, with significant deviations in CT values in the central region. After correction using the proposed method, the visual quality of the image is substantially improved, and low-frequency artifacts are largely eliminated.

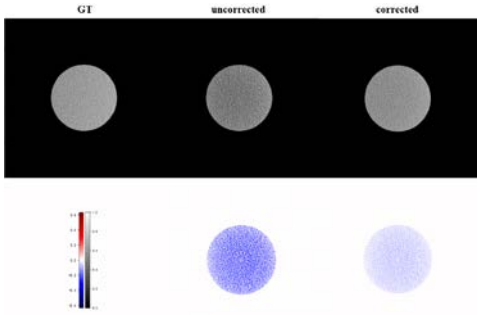


Fig. 3. Top Row: Cylindrical Water Mold Reconstructed Image. Bottom Row: Relative difference image between the reconstructed domain and the actual image of the cylindrical water mold

Analysis of the difference map reveals that the coupling between the physical correction term and the PINN accurately captures and compensates for the anisotropic scattering components induced by edge-source displacement. The significant reduction in residuals demonstrates that this framework, when handling non-stationary scattering fields, not only ensures the accuracy of the physical model but also leverages the neural network's adaptive adjustment capability for complex nonlinear parameters, effectively enhancing the system's quantitative accuracy.

Table II summarizes the reconstruction quality metrics across different methods. Quantitative results demonstrate that the proposed PINN-driven scattering correction framework achieves superior performance, with reconstructed images exhibiting an 8.6 dB increase in PSNR, a 52.2% reduction in RMSE, and a 0.5 improvement in SSIM. This validates the enhanced model generalization capability conferred by the physical constraint term within the composite loss function.

TABLE II
QUANTITATIVE INDICATOR COMPARISON

	Uncorrected	Corrected	Improvement
RMSE	73.4HU	35.1HU	52.2%
PSNR	24.5dB	33.4dB	8.6dB
SSIM	0.93	0.98	0.5

To provide a more intuitive comparison of calibration accuracy, we extracted the profile line from the center row of the reconstructed image. As shown in Fig. 4, the uncorrected profile line exhibits significant deviation from the ideal curve, whereas the curve calibrated by the algorithm proposed in this paper shows minimal discrepancy from the scatter-free reference value. These results collectively validate that the algorithm can accurately compensate for scatter interference under SLCT geometry, ensuring the quantitative accuracy of reconstructed images.

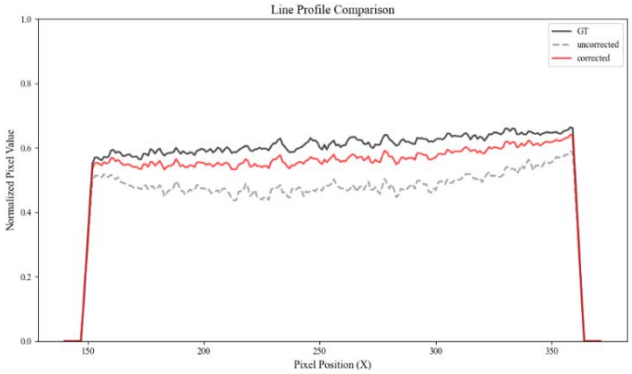


Fig. 4. Reconstructed Cross-Section of the Middle Row for Image Comparison

V.CONCLUSION

This study presents an adaptive scatter correction framework driven by PINN. Designed to address the non-centrosymmetric geometry of SLCT, the framework incorporates a geometry-aware correction term into the fASKS scatter model, thereby effectively compensating for anisotropic scattering and asymmetric point-spread functions caused by edge-source displacement. A key advantage of the approach is the use of a PINN to replace labor-intensive empirical parameter tuning; the network learns the hyperparameters of the scatter model in an end-to-end adaptive manner. By embedding a physics-consistent loss function during training, the method maintains high data fidelity while also preserving the physical interpretability of the scatter estimates. The experimental results demonstrate its efficiency.

REFERENCES

[1] W. Zhao, K. J. Little, and J. M. Starman, "A model-based scatter artifacts correction for cone beam CT," *Med. Phys.*, vol. 43, no. 4, pp. 1736–1753, Apr. 2016.

[2] J. Cai et al., "A practical cone-beam CT scatter correction method with optimized Monte Carlo simulations for image-guided radiation therapy," *Med. Phys.*, vol. 42, no. 5, pp. 2487–2498, May 2015.

[3] J. M. Starman and N. J. Pelc, "Efficient scatter correction using asymmetric kernels," *Proc. SPIE 7258, Medical Imaging 2009*, 2009, 72581Z.

[4] J. H. Siewerdsen and D. A. Jaffray, "Cone-beam computed tomography with a flat-panel imager: Magnitude and effects of x-ray scatter," *Med. Phys.*, vol. 28, no. 2, pp. 220–231, Feb. 2001.

[5] W. Zhao, K. J. Little, and J. M. Starman, "A model-based scatter artifacts correction for cone beam CT," *Med. Phys.*, vol. 43, no. 4, pp. 1736–1753, Apr. 2016.

[6] Y. Nomura et al., "Deep learning-based scatter correction of simulated x-ray images," *Phys. Med. Biol.*, vol. 65, no. 3, p. 035006, Jan. 2020.

[7] T. Würfl et al., "Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1454–1463, Jun. 2018.

[8] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.

[9] W. Huang et al., “A hybrid cone-beam CT scatter correction method combining fast Monte-Carlo simulation and deep neural network,” IEEE Trans. Instrum. Meas., vol. 74, pp. 1–13, Jan. 2025.

[10] Z. Piao, W. Deng, Y. Xu, and J. Ma, “Adaptive scatter kernel deconvolution modeling for cone-beam CT scatter correction via deep reinforcement learning,” Med. Phys., vol. 50, no. 11, pp. 6813–6826, Nov. 2023.

[11] H. Li, R. Mohan, and X. R. Zhu, “Scatter kernel estimation with an edge-spread function method for cone-beam computed tomography imaging,” Phys. Med. Biol., vol. 53, no. 23, pp. 6729–6748, Dec. 2008.

[12] Y. Sun and J. M. Starman, "Improved scatter correction using adaptive scatter kernel superposition," *Phys. Med. Biol.*, vol. 55, no. 22, pp. 6695–6720, Nov. 2010.

Upright Cone-Beam Breast CT: Towards Breast Cancer Screening using Short-Scan and Sparse-View Acquisitions

Hsin Wu Tseng PhD, Jing-Tzyh Alan Chiang, MD PhD, and Srinivasan Vedantham PhD

Department of Radiology and Imaging Sciences, The University of Arizona, Tucson, AZ 85724, USA

***Abstract*— Dedicated breast computed tomography (CT) offers three-dimensional visualization without tissue overlap or breast compression. For breast cancer screening, the mean glandular dose (MGD) needs to be reduced to approximately 4.5 mGy so that it is comparable to standard 2-view screening mammography. This study investigates sparse-view acquisition combined with Fast, TV-Regularized, Iterative Statistical reconstruction Technique (FRIST) for an upright cone-beam breast CT system. The system employs a 210-degree short-scan trajectory with 0.15 mm copper and 1 mm aluminum filtration, 60 kVp, 50 mA tube current, and 20 ms pulse width. For the default acquisition protocol of 210 views over 210 degrees (10.5 mGy MGD), traditional Feldkamp-Davis-Kress (FDK) reconstruction barely visualizes 0.24 mm microcalcifications—a critical limitation for screening applications. To investigate dose vs. microcalcification visibility, we evaluated sparse-view protocols with 120 views (6.0 mGy), 110 views (5.5 mGy), 100 views (5.0 mGy), and 90 views (4.5 mGy) reconstructed using FRIST algorithm. Results demonstrate that FRIST reconstruction with 90 views successfully enables visualization of 0.24 mm calcifications at MGD of 4.5 mGy—a 57% dose reduction from the reference protocol while maintaining diagnostic image quality. The upright geometry eliminates prone positioning discomfort. These findings establish the technical feasibility of sparse-view upright breast CT for population-based breast cancer screening, addressing the primary barrier of radiation dose while preserving microcalcification visibility that is essential for early cancer detection.**

***Index Terms*— breast computed tomography, sparse-view reconstruction, iterative reconstruction, screening, radiation dose reduction, FRIST algorithm, upright geometry, short-scan.**

I. Introduction

Breast cancer screening with mammography has significantly reduced mortality, yet inherent limitations including tissue superposition, breast compression discomfort, and reduced sensitivity in dense breasts motivate the development of alternative three-dimensional imaging modalities [1]. Dedicated breast computed tomography (CT) addresses these challenges by providing true volumetric visualization without compression and tissue overlap [2-3].

Current clinical breast CT systems employ prone patient positioning with full 360-degree gantry rotation [2-3]. While these systems have demonstrated improved diagnostic performance over mammography [4], they present several practical challenges. Prone positioning requires patients to climb onto a table and turn their heads sideways during the scan, causing neck and shoulder discomfort, and these have been reported in up to 39% of study participants [3]. This positioning is particularly challenging for elderly patients and those with mobility limitations or musculoskeletal conditions—populations that constitute a significant proportion of the screening demographic [5].

The upright cone-beam breast CT system investigated in this study addresses these limitations through an innovative geometry that enables patient-friendly positioning and a compact system design (Fig. 1). The upright configuration allows patients to simply stand during image acquisition, eliminating the need to climb onto a table. Furthermore, the compact footprint enables installation in standard mammography rooms without extensive facility modifications.

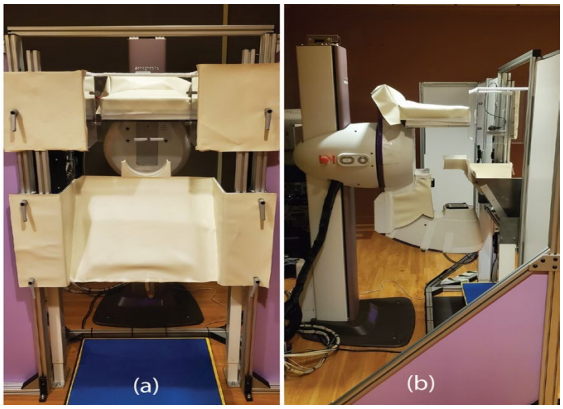


Fig. 1 Upright cone-beam breast CT system. (a) Front view. (b) Side view.

Beyond positioning considerations, radiation dose remains the primary barrier to translating breast CT from diagnostic imaging to screening applications. The U.S. Food and Drug Administration (FDA) limit the mean glandular dose (MGD) to 3 mGy for a phantom representing the standard compressed breast [6]. For screening, the goal of breast CT is to replace the two standard views with a single scan. Hence, 6 mGy is considered the upper limit for screening using breast CT. A standard uncompressed breast has not been defined. Hence, we

use a phantom representative of an average uncompressed breast for dosimetry and target a MGD of 4.5 mGy.

The upright breast CT system employs a 210-degree short-scan trajectory, which satisfies the data sufficiency conditions for the central plane [7]. Preliminary evaluation using Feldkamp-Davis-Kress (FDK) analytical reconstruction [8] at MGD of approximately 4.5 mGy indicated insufficient image quality for reliable detection of microcalcifications and served as a motivation for this study.

Advanced iterative reconstruction algorithms based on compressed sensing theory offer potential solutions for maintaining image quality with reduced projection data [9-11]. The Fast, TV-Regularized, Iterative Statistical reconstruction Technique (FRIST) [12] represents an accelerated implementation of total variation minimization specifically designed for breast CT, providing substantial noise reduction while preserving spatial resolution. Prior investigations have demonstrated FRIST's capability to suppress artifacts and improve image quality in sparse-view cone-beam breast CT acquisitions [13].

When optimizing breast CT systems for screening applications where the lesion type is unknown a priori, a recent study using cascaded systems analysis demonstrated that microcalcification detection represents the more demanding imaging task compared to soft-tissue lesion detection [14]. Consequently, this study focuses on microcalcification visualization as the primary metric for evaluating sparse-view reconstruction performance.

This work investigates the combination of upright patient positioning, 210-degree short-scan geometry, and FRIST iterative reconstruction at a MGD of approximately 4.5 mGy, while maintaining visualization of microcalcification clusters. We hypothesize that sparse-view acquisition protocols reconstructed with FRIST can reduce radiation dose by more than 50% compared to the reference protocol, while preserving visualization of microcalcifications.

II. MATERIALS AND METHODS

A. Upright Breast CT System

The upright cone-beam breast CT system employs a compact geometry optimized for patient comfort and clinical workflow integration. The system utilizes a W-anode x-ray tube (60 kV, 50 mA, 20 ms pulse width) and a CsI:Tl scintillator-coupled CMOS detector with source-to-axis distance (SAD) of 503 mm and source-to-detector distance (SDD) of 703 mm. The detector has a pixel pitch of 151.8 μm and an active area of 30 cm × 30 cm. A 0.15 mm copper filter followed by a 1 mm Al filter is placed at the x-ray tube output. The gantry traverses a 210-degree arc inferior to the breast.

B. Phantom and Acquisition Protocols

Image quality evaluation was performed using a breast quality control (QC) phantom containing microcalcification clusters (calcium carbonate) of known sizes (Fig. 2). The phantom includes a pentagon-shaped calcification cluster with six individual calcifications—one at the center and five at the vertices—with calcification sizes ranging from 0.16 mm to 0.32 mm in diameter. The 0.24 mm calcifications were selected as the primary evaluation target. Visualization was defined as the ability to identify individual calcification specks, with successful detection requiring identification of at least 4 of 6 calcifications per cluster (≥67% detection rate).

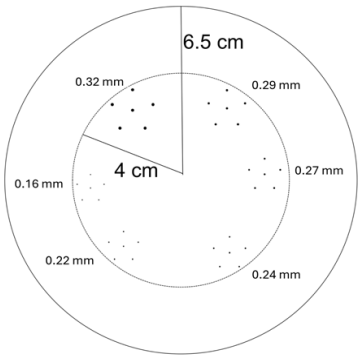


Fig. 2 Breast QC phantom has six clusters of pentagon shape. Each cluster has six calcifications (calcium carbonate).

Five acquisition protocols were investigated: (1) Reference protocol: 210 views over 210 degrees (1.0° angular sampling; 10.5 mGy MGD); (2) Protocol A: 120 views over 210 degrees (1.75° angular sampling; 6.0 mGy MGD); (3) Protocol B: 110 views over 210 degrees (1.91° angular sampling; 5.5 mGy MGD); (4) Protocol C: 100 views over 210 degrees (2.10° angular sampling; 5.0 mGy MGD); (5) Protocol D: 90 views over 210 degrees (2.33° angular sampling; 4.5 mGy MGD). All protocols maintained identical X-ray beam quality (HVL = 4.23 mm of Al) and technique factors (60 kV, 50 mA, 20 ms).

C. Image Reconstruction Algorithms

Two reconstruction approaches were compared: conventional filtered back-projection using the Feldkamp-Davis-Kress (FDK) algorithm [9] with Parker weights [15] for short-scan data, and the Fast Iterative TV-Regularized Statistical reconstruction Technique (FRIST) [12, 13]. The reference 210-view dataset was reconstructed using FDK algorithm with ramp filter and Parker weights. Sparse-view protocols were reconstructed using FRIST, which minimizes the total variation (TV) of the image subject to data consistency constraints.

Both FDK and FRIST reconstructions initially used 0.1 mm × 0.1 mm × 0.1 mm isotropic voxels, followed by 2×2×2 rebinning to 0.2 mm × 0.2 mm × 0.2 mm to suppress salt-and-pepper artifacts. FRIST hyperparameters were: $\alpha = 2\times10^{-3}$ (TV reduction parameter), $\beta = 1$ (data fidelity weight), and 130 iterations. The reconstruction matrix was 1024 × 1024 pixels in-plane with 500 slices, yielding an in-plane field-of-view of 204.8 mm × 204.8 mm and chest-wall to nipple coverage of 100 mm.

D. Radiation Dose Estimation

Mean glandular dose (MGD) was estimated using a two-step methodology [16]. First, air kerma was measured at the axis-of-rotation using a calibrated ion chamber under the reference acquisition protocol. Second, Monte Carlo simulations were performed using the GEANT4 toolkit to determine the normalized glandular dose coefficient (DgN) specific to the

system geometry and beam quality. The phantom represented a semi-ellipsoidal breast with 13 cm diameter at the chest wall and 10 cm length from chest wall to nipple, composed of 20% glandular tissue. For sparse-view protocols, MGD was estimated by linear scaling from the reference protocol based on the number of projection views.

E. Image Quality Assessment

Image quality was assessed based on visualization of the 0.24 mm calcification cluster. Successful visualization was defined as the ability to identify at least four of the six calcifications (≥67% detection rate) within the cluster.

III. RESULTS

Table I summarizes the results from microcalcification visualization for all acquisition protocols and reconstruction methods. The number of visible calcifications within each cluster (out of 6 total) is reported for each size.

TABLE I

CALCIFICATION VISUALIZATION RESULTS FOR DIFFERENT ACQUISITION PROTOCOLS

Protocol	Views	MGD (mGy)	Recon Method	0.32 mm	0.29 mm	0.27 mm	0.24 mm
Reference	210	10.5	FDK	6	6	4	2
A	120	6.0	FDK	6	4.5	2	1
B	110	5.5	FDK	6	2.5	1	0
C	100	5.0	FDK	6	0.5	0.5	0
D	90	4.5	FDK	4.5	0	0	0
D	90	4.5	FRIST	6	5.5	6	4.5

Values represent number of visible calcifications out of 6 total per cluster size.

The reference protocol with 210 views and FDK reconstruction (10.5 mGy) failed to detect the 0.24 mm cluster, as only 3 of 6 calcifications were visible (50% detection rate, below the 67% threshold). As the number of views decreased, with FDK reconstruction, image quality progressively degraded. At 120 views (6.0 mGy), none of the 0.24 mm calcifications were visible.

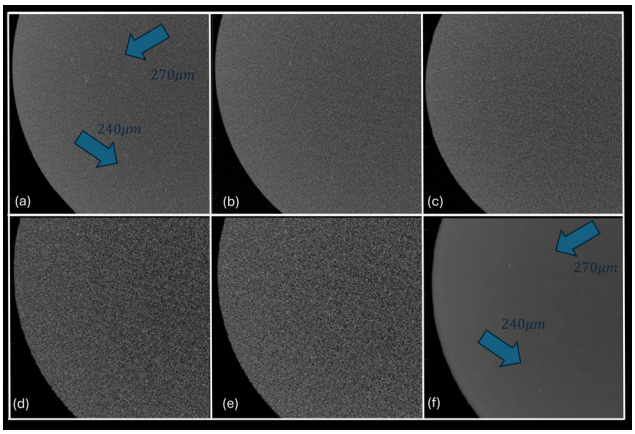


Fig. 3 FDK and FRIST images of 10-slices maximum intensity projection (MIP). (a) Reference: FDK using 210 views (10.5 mGy). (b) FDK using 120 views (6 mGy). (c) FDK using 110 views (5.5 mGy). (d) FDK using 100 views (5 mGy). (e) FDK using 90 views (4.5 mGy). (f) FRIST using 90 views (4.5 mGy).

In contrast, Protocol D with FRIST reconstruction (90 views, 4.5 mGy) successfully achieved the 0.24 mm visualization criterion with 4 of 6 calcifications clearly visible (67% detection rate). FRIST also demonstrated superior performance for larger calcifications. Notably, FRIST at 90 views and 4.5 mGy achieved better 0.24 mm visualization than FDK at 210 views and 10.5 mGy (57% dose reduction while improving detection capability). Figure 3 shows representative maximum intensity projection images.

IV. DISCUSSION

This study demonstrates that upright cone-beam breast CT with 210-degree short-scan geometry, sparse-view acquisition, and FRIST reconstruction achieves the ability to visualize microcalcifications at MGD suitable for screening. The key finding is that Protocol D (90 views, 4.5 mGy) with FRIST reconstruction successfully visualized 0.24 mm calcifications (4 of 6 visible, 67% detection rate), while the reference protocol (210 views, 10.5 mGy) with FDK reconstruction failed this criterion despite 2.3× higher dose.

The superior performance of FRIST can be attributed to several factors. First, total variation minimization exploits the sparse-gradient property of medical images, effectively suppressing streak artifacts that severely degrade FDK reconstructions from sparse-view data. Second, the statistical iterative framework models Poisson noise characteristics of quantum-limited acquisitions, optimally weighting projection data based on variance. Third, the two-stage reconstruction strategy eliminates salt-and-pepper artifacts while maintaining sharpness needed for calcification detection.

The potential clinical implications of achieving microcalcification visualization at MGD of 4.5 mGy are substantial. This can pave the way for future clinical trials to investigate the potential of breast CT that enables compression-free imaging, three-dimensional visualization, and eliminates tissue superposition. Furthermore, the upright geometry offers significant patient comfort advantages over prone systems, potentially improving compliance in screening populations.

Several limitations should be acknowledged. First, this study evaluated a single breast QC phantom rather than clinical cases, and calcification detectability in real breast tissue requires validation. Second, our binary visualization criterion provides an objective threshold but does not capture subtle differences in calcification conspicuity.

Future work will focus on clinical validation with reader studies using actual patient cases containing pathology-confirmed microcalcifications. Investigation of dual-energy acquisition protocols combined with FRIST reconstruction may enable material decomposition for calcification composition analysis, potentially improving specificity for malignancy prediction.

V. CONCLUSIONS

This study demonstrates the feasibility of upright cone-beam breast CT for screening applications using the combination of 210-degree short-scan geometry and FRIST iterative reconstruction. The optimal protocol (90 views, 4.5 mGy MGD) achieves 0.24 mm calcification detection capability

while operating at radiation doses comparable to two-view screening mammography. Notably, FRIST reconstruction at 4.5 mGy outperforms conventional FDK reconstruction at 10.5 mGy, representing a 57% dose reduction with simultaneous image quality improvement. These findings support investigating the clinical translation of upright breast CT as a screening modality that combines the three-dimensional diagnostic advantages of CT with patient comfort, accessibility for women with mobility limitations, and radiation safety suitable for population-based screening programs.

ACKNOWLEDGMENT

This work was supported in part by Institutional Research Grant number IRG-18-161-40 from the American Cancer Society (ACS) and the NIH R01 CA241709. The contents are solely the responsibility of the authors and do not represent the official views of the ACS or the NIH.

REFERENCES

[1] E. D. Pisano et al., "Diagnostic performance of digital versus film mammography for breast-cancer screening," N. Engl. J. Med., vol. 353, no. 17, pp. 1773–1783, Oct. 2005.

[2] K. K. Lindfors et al., "Dedicated breast CT: Initial clinical experience," Radiology, vol. 246, no. 3, pp. 725–733, Mar. 2008.

[3] A. O'Connell et al., "Cone-beam CT for breast imaging: Radiation dose, breast coverage, and image quality," AJR Am. J. Roentgenol., vol. 195, no. 2, pp. 496–509, Aug. 2010.

[5] E. B. Cole et al., "Clinical performance of dedicated breast computed tomography in comparison to diagnostic digital mammography," presented at the Radiological Society of North America Annual Meeting, Chicago, IL, USA, 2015.

[6] N. C. Wright et al., "The recent prevalence of osteoporosis and low bone mass in the United States based on bone mineral density at the femoral neck or lumbar spine," J. Bone Miner. Res., vol. 29, no. 11, pp. 2520–2526, Nov. 2014.

[7] H. K. Tuy, "An inversion formula for cone-beam reconstruction," SIAM J. Appl. Math., vol. 43, no. 3, pp. 546–552, Jun. 1983.

[8] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," J. Opt. Soc. Am. A, vol. 1, no. 6, pp. 612–619, Jun. 1984.

[9] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," IEEE Trans. Inf. Theory, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[10] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," Phys. Med. Biol., vol. 53, no. 17, pp. 4777–4807, Sep. 2008.

[11] J. Bian et al., "Investigation of iterative image reconstruction in low-dose breast CT," Phys. Med. Biol., vol. 59, no. 11, pp. 2659–2685, Jun. 2014.

[12] H. W. Tseng, S. Vedantham, and A. Karellas, "Cone-beam breast computed tomography using ultra-fast image reconstruction with constrained, total-variation minimization for suppression of artifacts," Phys. Med., vol. 73, pp. 117–124, May 2020.

[13] H. W. Tseng, A. Karellas, and S. Vedantham, "Sparse-view, short-scan, dedicated cone-beam breast computed tomography: Image quality assessment," Biomed. Phys. Eng. Express, vol. 6, no. 6, p. 065015, Sep. 2020.

[14] T. Larsen, H. W. Tseng, J.-T. A. Chiang, and S. Vedantham, "Soft-tissue lesion and microcalcification detectability in cone-beam breast CT: cascaded system analysis," J. Med. Imaging, vol. 12, no. 6, p. 063501, Nov. 2025, doi: 10.1117/1.JMI.12.6.063501.

[15] S. Wesarg, M. Ebert, and T. Bortfeld, "Parker weights revisited," Med. Phys., vol. 29, no. 3, pp. 372–378, Mar. 2002, doi: 10.1118/1.1450132.

[16] H. W. Tseng, S. Vedantham, and A. Karellas, "Radiation dosimetry of a clinical prototype dedicated cone-beam breast CT system with offset detector," Med. Phys., vol. 48, no. 4, pp. 1955–1970, Apr. 2021.

Analytical Weighted Reconstruction for Longitudinal Multiple X-ray Source Array CT (MXA-CT)

Jiaming Liu, Gengxin Shi, Andrew M. Hernandez, Alejandro Sisniega, Richard E. Colbeth, Vance Robinson, Jeffrey H. Siewerdsen, Craig K. Abbey, Paul Schwoebel, John M. Boone, and Wojciech Zbijewski

***Abstract*—Longitudinal Multiple X-ray Source Array Computed Tomography (MXA-CT) extends the axial coverage of circular CT by distributing multiple x-ray focal spots along the rotation axis. By alternating simultaneous exposures of odd and even sources, MXA-CT can increase volumetric coverage per rotation while keeping the cone angle of each exposure moderate. However, the resulting volumetric sampling becomes voxel-dependent: the angular range of rays illuminating a voxel and redundancy from conjugate views is inhomogeneous. Direct application of conventional cone-beam filtered backprojection (FBP) can therefore lead to intensity bias and shading in the interface regions between the sources.**

This work presents an analytical reconstruction algorithm for longitudinal MXA-CT that avoids explicit cone-to-parallel rebinning. Building on a closed-form characterization of single-source illumination and conjugate-ray redundancy, we introduce a voxel-wise weighting strategy composed of a partial-scan weight that allocates complementary contributions across redundant views and a normalization weight that compensates for nonuniform angular sampling of each voxel. We apply this framework to MXA-CT setting by deriving geometry-dependent region partition and combination rules for reconstructions of neighboring sources. The proposed approach is evaluated in numerical experiments in idealized and anthropomorphic phantoms assuming a 6-source MXA CT configuration.

***Index Terms*—MXA-CT, image reconstruction, image quality.**

I. Introduction

A Multiple X-ray Source Array CT (MXA-CT) system based on a longitudinal configuration of rotating tungsten anodes within a common vacuum enclosure is under development, as shown in Fig. 1. The proposed architecture builds on earlier distributed x-ray source prototypes for breast tomosynthesis with a stationary gantry [1]. MXA CT can potentially provide wide volumetric coverage in a single gantry rotation with reduced cone-beam and scatter artifacts compared to traditional single-source Cone-Beam CT (CBCT) and possibly with a faster imaging time than with a helical scan and a narrow detector.

Image reconstruction in the MXA CT configuration shown in Fig. 1 has to contend with a highly irregular voxel sampling pattern [2]. Specifically, depending on system configuration (source spacing δ and magnification) and location in the volume, voxels may receive: (i) full 360° coverage from only one source (central region of that source), (ii) partial scan coverage from only one source (the so-called extended region of that source towards the edge of its beam), or (iii) some combination of partial or full coverage of two neighboring sources. In analytical reconstruction, the resulting voxel-dependent sampling density and data redundancy need to be handled by appropriate weighting to avoid shading and streak artifacts [3]–[6].

Here, we adapt a voxel-wise partial scan weighting scheme for the extended region of single-source CBCT of [3] to handle the proposed longitudinal multi-source configuration. We develop a modification to the weights to avoid explicit cone-to-parallel rebinning [7] and investigate optimal combinations of reconstructions of neighboring sources to achieve artifact-free imaging in the regions of overlapped sampling.

II. Methodology

This section reviews the sampling properties of single-source CBCT in the central and extended regions, proposes a redundancy weighting reconstruction that avoids cone-to-parallel rebinning in the single source case, and extends it to MXA-CT.

A. Single-source CBCT volumetric sampling properties

We adopt a standard circular trajectory with source radius R_F as shown in Fig. 1(c). According to [3], a voxel $\mathbf{p} = (r, \varphi, z)$ is illuminated at a certain gantry angle α if the ray connecting the voxel and source falls within the maximum

J. Liu, G. Shi, and W. Zbijewski are with Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA.

A. Sisniega is with Sedecal Corporation, Madrid, Spain.

J. H. Siewerdsen is with Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, TX, USA.

R. E. Colbeth and V. Robinson are with Varex Imaging Corporation, Salt Lake City, UT, USA.

C. K. Abbey is with the Department of Psychological & Brain Sciences, University of California, Santa Barbara, Santa Barbara, CA, USA.

P. Schwoebel is with the Department of Physics and Astronomy, University of New Mexico, Albuquerque, NM, USA.

A. M. Hernandez and J. M. Boone are with Department of Radiology, University of California Davis, Sacramento, CA, USA.

Work supported by NIH U01-EB034205-02

cone angle $\gamma_{\max}$. Consequently, the range of gantry angles illuminating the voxel is:

$$\alpha \in \left[\varphi - \frac{\pi}{2} - \Delta\alpha(r,z),\ \varphi + \frac{\pi}{2} + \Delta\alpha(r,z)\right], \quad (1)$$

where $\Delta\alpha_1 = \arcsin[(R_F - cz_1)/r]$, $\Delta\alpha_2 = \arcsin[(R_F - cz_2)/r]$, $c = \cot(\gamma_{\max})$. The corresponding parallel-beam angular sampling range θ at voxel $\mathbf{p}$ is

$$\theta \in \left[\varphi - \frac{\pi}{2} - \Delta\theta(r,z),\ \varphi + \frac{\pi}{2} + \Delta\theta(r,z)\right]. \quad (2)$$

When $\Delta\theta(r,z) = \frac{\pi}{2}$, the voxel receives full 360° parallel beam coverage (full scan), where each ray has a redundant (conjugate) counterpart 180° apart. When $0 < \Delta\theta(r,z) < \frac{\pi}{2}$, the voxel receives more than 180° but less than 360° parallel beam coverage (partial scan) and only some rays passing through the voxel are redundant, resulting in shading if treated with unweighted FDK reconstruction. For an in-plane field-of-view radius R, these angular ranges can be converted to two boundary surfaces in the volume (see Fig. 1):

$$Z_{\text{extended}} = \frac{1}{c}\left(R_F - \frac{R^2}{R_F}\right), Z_{\text{central}} = \frac{1}{c}\left(R_F - R\right). \quad (3)$$

When $|z| < Z_{\text{central}}$ (central region), all voxels in the slice receive full coverage. When $Z_{\text{FDK}} \le |z| < Z_{\max}$ (extended region), some voxels receive only partial scan coverage.

B. Single-source CBCT weighted reconstruction algorithm

Following [3], we introduce voxel-wise weighting during backprojection to address spatially-variant ray redundancy patterns in single-source FDK reconstruction:

$$f_{\text{wFDK}} = \frac{1}{2}\int_0^{2\pi} \frac{w_{\text{PS}}}{w_{\text{Norm}}\, w_Z} \times \left[\int_{-\infty}^{\infty} \cos(\xi)\, q(s,v,\alpha)\, h(s'-s)\, ds\right] d\alpha, \quad (4)$$

where $q(s,v,\alpha)$ is the log-transformed projection data, $h(\cdot)$ is 1D ramp-type filter, ξ is the local tilt angle. The weight consists of three components: w_{PS} is a voxel-dependent partial-scan weight, w_{Norm} is a normalization weight that compensates for nonuniform sampling in the parallel domain without explicit rebinning, and w_Z is a z-axis calibration weight to ensure consistent attenuation scaling across z.

1) w_{PS}: we adopt w_{PS} [3] as a smoothly fading function in θ domain. In particular, the weights of two redundant rays add up to 2 and the ray with smaller cone angle is given a larger weight. The weight of unpaired rays is also 2.

2) w_{Norm}: While the gantry angles are uniformly spaced, the parallel beam ray samples around each voxel are not. For a fixed voxel, $\theta(\alpha) = \alpha + \beta(\alpha)$ with

$$\beta(\alpha) = -\arctan\left(\frac{r\sin(\alpha-\varphi)}{R_F + r\cos(\alpha-\varphi)}\right). \quad (5)$$

Differentiating (5) yields

$$\frac{\mathrm{d}\theta}{\mathrm{d}\alpha} = \frac{R_F\left(R_F + r\cos(\alpha-\varphi)\right)}{R_F^2 + 2R_F r\cos(\alpha-\varphi) + r^2}. \quad (6)$$

Thus, a uniform step $\Delta\alpha$ corresponds to a nonuniform step $\Delta\theta(\alpha) \approx \frac{d\theta}{d\alpha}\Delta\alpha$ depending on voxel location. For example, when $\alpha = \varphi$, $\frac{d\theta}{d\alpha} = \frac{R_F}{R_F+r} < 1$ and when $\alpha = \varphi + \pi$, $\frac{d\theta}{d\alpha} = \frac{R_F}{R_F-r} > 1$. Therefore, θ samples become denser when the source is near the voxel and sparser when the source is far from the voxel. If the redundancy weights w_{PS} are applied in the non-uniformly sampled θ domain, the sum of contribution across all views can become inconsistent across voxels, leading to shading artifacts. A direct cone-to-parallel rebinning solves this by interpolating the rays onto an equispaced grid in θ, but we wish to avoid this step for implementation efficiency. Therefore, we propose a normalization weight that enforces a consistent average contribution per voxel:

$$w_{\text{Norm}}(r,\varphi,z) = \frac{1}{N_\alpha}\sum_{k=1}^{N_\alpha} w_{\text{PS}}(r,\varphi,z,\theta_k). \quad (7)$$

3) w_{Z}: To minimize residual discontinuities between the slices, we also apply a z-dependent normalization based on a reconstruction of a water calibration phantom (different from the object being reconstructed). Let $f_{\text{FDK}'}$ denote the weighted reconstruction of the calibration without the w_{Z} term (but including w_{PS} and w_{Norm}) and μ_{water} be the attenuation of water:

$$w_{\text{Z}}(z) = \frac{\dfrac{1}{|\Omega_{\text{FOV}}|}\displaystyle\sum_{(x,y)\in\Omega_{\text{FOV}}} f^{\text{water}}_{\text{wFDK}'}(x,y,z)}{\mu_{\text{water}}}. \quad (8)$$

C. Volumetric sampling properties of MXA-CT

In MXA-CT geometry, all sources S_i share the same R_F and maximum cone angle. The system alternates between simultaneous exposures of odd and even sources during acquisition. We consider a longitudinal two-source configuration, which serves as the basic building block of the MXA-CT configuration. Let S_1 and S_2 denote the sources located at longitudinal offsets $\pm\delta/2$ relative to the x–y plane. For voxel p, its effective distances to each source plane are $z_1 = |z + \frac{\delta}{2}|$, $z_2 = |z - \frac{\delta}{2}|$. The view-dependent illumination condition for each source follows the single-source derivation by replacing $|z|$ with z_i. Consequently, the gantry angle ranges for which $\mathbf{p}$ is illuminated can be written as the union

$$\alpha \in \left[\varphi - \frac{\pi}{2} - \Delta\alpha_1,\ \varphi + \frac{\pi}{2} + \Delta\alpha_1\right] \cup \left[\varphi - \frac{\pi}{2} - \Delta\alpha_2,\ \varphi + \frac{\pi}{2} + \Delta\alpha_2\right]. \quad (9)$$

where $\Delta\alpha_1 = \arcsin[(R_F - cz_1)/r]$, $\Delta\alpha_2 = \arcsin[(R_F - cz_2)/r]$,$c = \cot(\gamma_{\max})$. The range of θ could be derived in the same way. The reconstruction for voxels where both

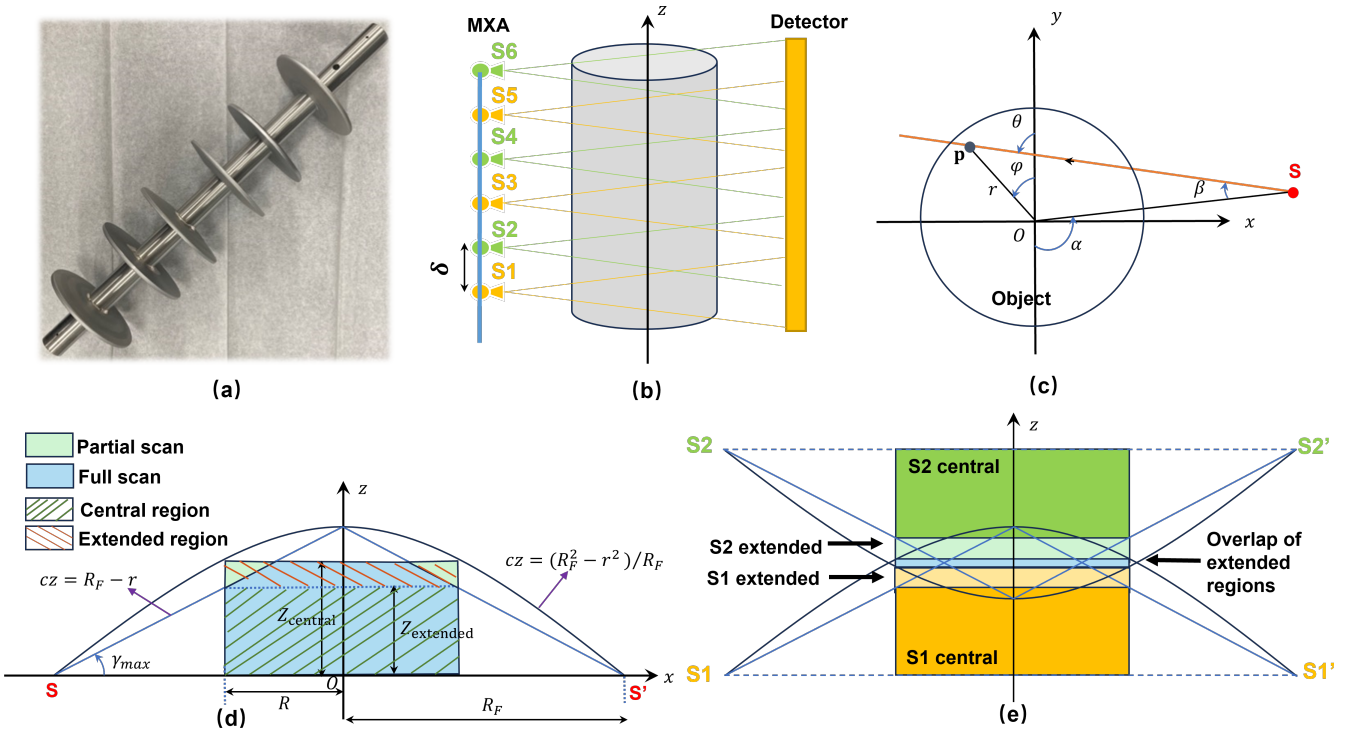


Fig. 1. (a) Prototype multiple X-ray source anode array, (b) geometry of MXA-CT (coronal plane), (c) reconstruction geometry in an axial slice, (d) sampling properties in the coronal plane of single-source CBCT, (e) sampling properties in the coronal plane of MXA-CT

sources contribute angular sampling (i.e., where both sets in Eq. 9 are non-empty) is performed by weighted summation of individual reconstructions of each source.

D. MXA-CT weighted reconstruction algorithm

For a given field of view R, source-axis distance R_F, and cone angle $\gamma_{\max}$, different values of source spacing δ correspond to different amounts of sampling region overlap of two neighboring sources. When $\delta = \delta_{\max} = 2(R_F^2 - R^2)/cR_F$, the extended regions of the two sources intersect exactly at their boundaries; when $\delta = \delta_D = (2R_F^2 - RR_F - R^2)/cR_F$, the two extended regions completely overlap; when $\delta = \delta_{\min} = 2(R_F - R)/c$, the two central regions intersect exactly at their boundaries. Together, these values define the following situations: (i) $\delta > \delta_{\max}$ yields a data-incomplete band between neighboring sources, (ii) $\delta_D < \delta \leq \delta_{\max}$ yields overlap of extended regions, (iii) $\delta_{\min} < \delta \leq \delta_D$ yields overlap of the extended region one source with the central region of the other source, (iv) $\delta \leq \delta_{\min}$ yields overlapping central regions where weighting is unnecessary. Using (ii) as an example (illustrated in Fig. 1(e)), the reconstructed volume f can be obtained as the following weighted combination:

$$f = \begin{cases} f_{\mathrm{FDK}}^{S_1}, & -\frac{\delta}{2} \leq z < -\frac{\delta}{2} + Z_{central}, \\ f_{\mathrm{wFDK}}^{S_1}, & -\frac{\delta}{2} + Z_{central} \leq z < \frac{\delta}{2} - Z_{extended}, \\ \frac{f_{\mathrm{wFDK}}^{S_1}}{2} + \frac{f_{\mathrm{wFDK}}^{S_2}}{2}, & \frac{\delta}{2} - Z_{extended} \leq z < Z_{extended} - \frac{\delta}{2}, \\ f_{\mathrm{wFDK}}^{S_2}, & Z_{extended} - \frac{\delta}{2} \leq z < \frac{\delta}{2} - Z_{central}, \\ f_{\mathrm{FDK}}^{S_2}, & \frac{\delta}{2} - Z_{central} \leq z \leq \frac{\delta}{2}. \end{cases} \tag{10}$$

III. Results and Discussion

We evaluated the proposed normalization weight w_{Norm} and the z-axis calibration weight w_{Z} in a two-source MXA-CT simulation study of a digital model of the Corgi phantom (voxel size: 0.5 mm). The geometry is summarized in Table I. Fig. 2 shows reconstructions across the interface between two sources, including axial and sagittal slices under different weighting strategies: a simple unweighted sum of conventional FDK reconstructions of the two sources (1st row), a sum of two FDK reconstructions naively weighted by the total number of projection rays contributing to each voxel (2nd row), and the proposed combination of two weighted reconstructions with only the redundancy weights w_{PS} (3rd

TABLE I
GEOMETRY SETTINGS [DISTANCES IN MM].

views	$\gamma_{\max}$	R_F	R_D	R	δ	$\delta_{\max}$	δ_D	$\delta_{\min}$
720	4.5°	580	420	114	43	44.6	40.9	37.3

row), with redundancy and normalization weights (4th row), and with redundancy, normalization, and z-weights (bottom row). We also show maps of inter-slice intensity variation obtained for the selected axial slice by subtracting the average of its two direct neighbors above and below .

Compared to a sum of conventional FDKs, the counting weight mitigates the prominent streaks in the overlap region but remains affected by shading because it does not account for ray redundancy. Introducing the partial-scan weight largely removes the interslice discontinuity. However, a doming pattern is visible in the axial slices. The proposed normalization weight w_{Norm} reduces the doming, with the remaining slight interslice variations handled by the z-axis calibration weight w_{Z}.

To quantitatively assess image uniformity, we used the mean attenuation coefficient of spherical inserts in the central region of the top source (indicated as ROI 2 in Fig. 2) as a reference and computed a difference to the inserts in the adjacent overlap region (ROI 1). As shown in Fig. 3, our method substantially reduces the mean difference to nearly zero with a small standard deviation across voxels, indicating that the reconstruction is free of bias and uniform.

Fig. 4 presents a two source MXA-CT simulation using of digital head phantom (source: CQ500 dataset, voxel size 0.5 mm). The display window is relatively narrow to emphasize soft-tissue contrast. In addition to the MXA-CT acquisition, a single cone-beam source simulation with the same longitudinal coverage was performed, which required a large cone angle of 9.4°. As expected, MXA-CT exhibits substantially reduced cone-beam artifacts compared to single-source CBCT (yellow arrow).

Our method achieved a highly uniform reconstruction with only minor discontinuities in the longitudinal direction that we attribute to the calibration-based w_{Z} correction. Theoretically, redundancy weights should be applied after cone-to-parallel rebinning. In the proposed implementation, the redundancy weights are instead applied directly in cone-beam geometry; as a result, the voxel-to-source-distance-dependent FDK backprojection weights may be incorrectly accumulated across redundant views. The calibration-based w_{Z} accounts for most of this mismatch in practice, but residual artifacts may remain for highly eccentric objects. We are therefore investigating alternative correction approaches that more directly address this error.

IV. CONCLUSION

We presented an analytical reconstruction framework for longitudinal MXA-CT that targets the complex voxel-dependent sampling characteristics at interfaces between neighboring sources. Using a voxel-wise parallel beam sampling analysis, we proposed a redundancy-weighting approach that avoids explicit cone-to-parallel rebinning. Strategies for combining weighted reconstructions of neighboring sources were developed. Simulation studies using a digital Corgi phantom and a head phantom verified efficient removal of shading artifacts and smooth inter-slice continuity in images obtained using a 6-source MXA-CT configuration.

Future work will involve investigation of various MXA-CT geometric configurations differing in magnification and source spacing and validation under poly-energetic beams and noise.

REFERENCES

[1] A. Sisniega, A. M. Hernandez, S. A. Shakeri, E. A. Morris, J. M. Boone, J. H. Siewerdsen, and P. R. Schwoebel, "A multiple x-ray-source array (mxa) system with a planar two-dimensional source distribution for digital breast tomosynthesis," *Medical physics*, vol. 51, no. 12, pp. 8709–8724, 2024.

[2] H. Zhou, A. M. Hernandez, A. Sisniega, R. E. Colbeth, V. Robinson, J. H. Siewerdsen, C. K. Abbey, P. Schwoebel, J. M. Boone, and W. B. Zbijewski, "Best in physics imaging: Volumetric sampling properties of longitudinal multiple x-ray-source array computed tomography (mxa-ct)," in *AAPM 67th Annual Meeting & Exhibition*, AAPM, 2025.

[3] R. Grimmer, M. Oelhafen, U. Elstrøm, and M. Kachelrieß, "Cone-beam ct image reconstruction with extended range," *Medical physics*, vol. 36, no. 7, pp. 3363–3370, 2009.

[4] S. Mori, M. Endo, S. Komatsu, S. Kandatsu, T. Yashiro, and M. Baba, "A combination-weighted feldkamp-based reconstruction algorithm for cone-beam ct," *Physics in Medicine & Biology*, vol. 51, no. 16, p. 3953, 2006.

[5] K. Zha, Q. Zhao, S. Luo, and Y. Li, "Bilateral extended fdk: An improved weighting method for static ct imaging," *Medical Physics*, vol. 51, no. 1, pp. 251–266, 2024.

[6] X. Tang, J. Hsieh, R. A. Nilsen, and S. M. McOlash, "Extending three-dimensional weighted cone beam filtered backprojection (cb-fbp) algorithm for image reconstruction in volumetric ct at low helical pitches," *International Journal of Biomedical Imaging*, vol. 2006, no. 1, p. 045942, 2006.

[7] M. Kachelrieß, M. Knaup, and W. A. Kalender, "Extended parallel backprojection for standard three-dimensional and phase-correlated four-dimensional axial and spiral cone-beam ct with arbitrary pitch, arbitrary cone-angle, and 100% dose usage," *Medical Physics*, vol. 31, no. 6, pp. 1623–1641, 2004.

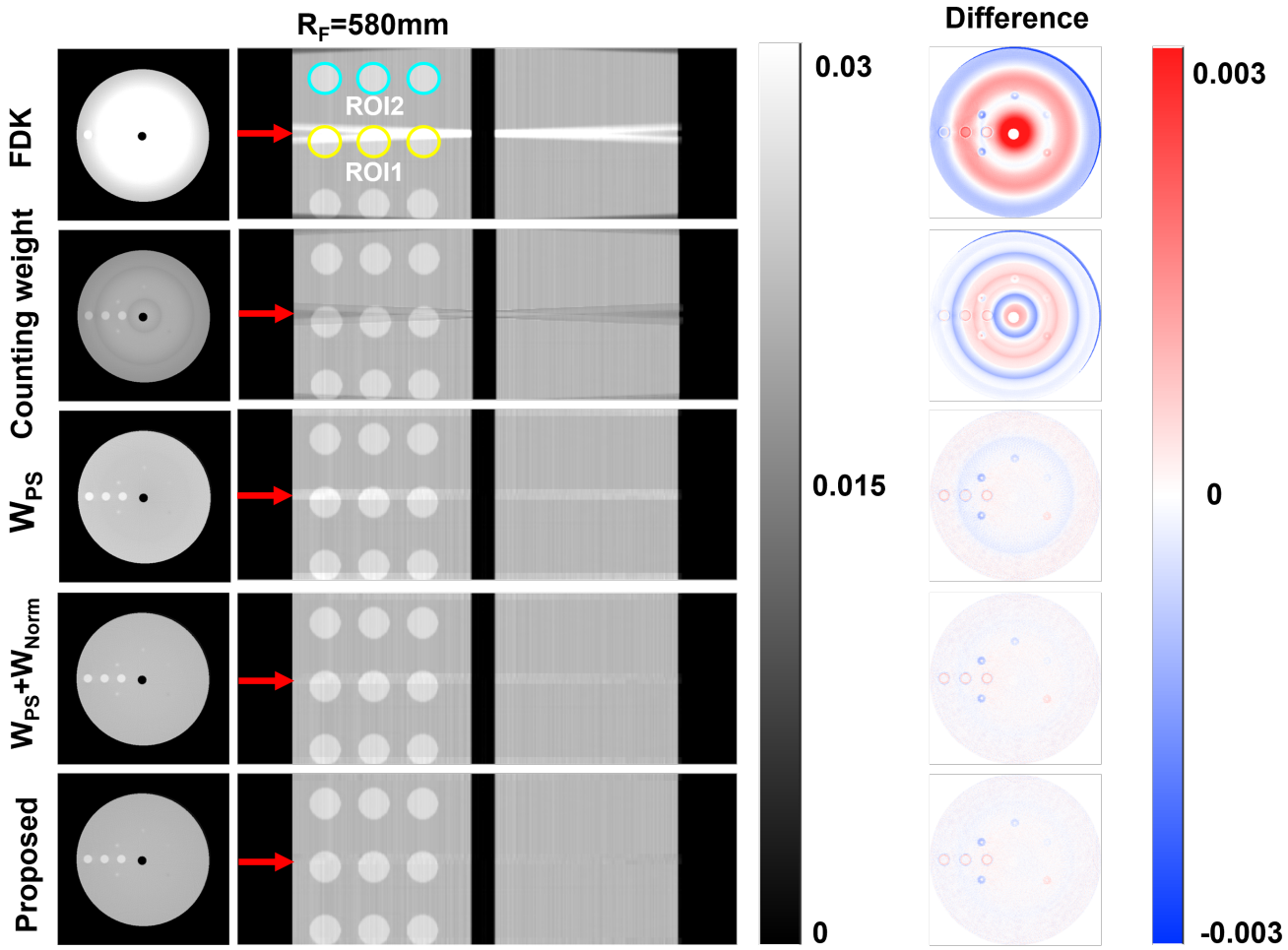


Fig. 2. MXA-CT reconstructions using different weighting schemes.

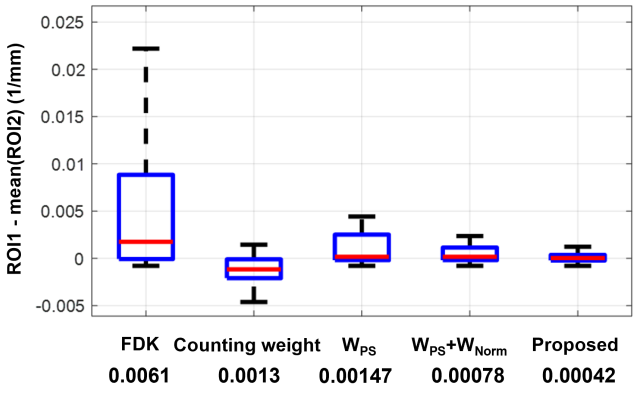


Fig. 3. Distributions of voxel value differences between ROI1 and the mean of ROI2 (see Fig. 2) for different weighting schemes. Standard deviations of the distributions are indicated at the bottom.

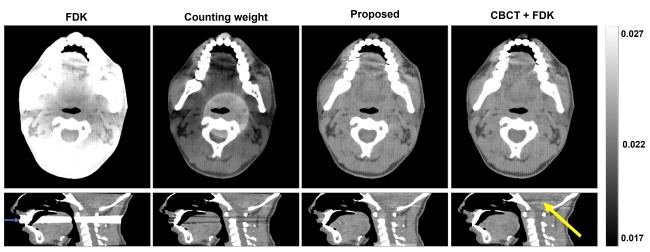


Fig. 4. MXA-CT and CBCT reconstructions of a head phantom using different weighting schemes.

Bayesian Regulators to Promote Sharp Image Edges in Limited-Angle Tomography

Gengsheng L. Zeng

***Abstract*—When the image reconstruction problem is severely ill-conditioned, for example, when the scanning angle is small, the analytic reconstruction algorithms produce image with too many artifacts to be useful in practice, while iterative Bayesian reconstruction algorithms can produce better images. The main purpose of Bayesian constraints is to regulate and stabilize the algorithm. Bayesian penalty minimization is a popular and effective method in regulating, denoising, and edge preserving. The main players for the Bayesian regulators are the total-variation (TV), Huber, L_0, L_1, L_2, and L_q norms. This paper suggests some other methods to regulate an iterative algorithm and to encourage sharp image edges. Some computer simulations are provided. The L_q-like regulators seem to be more effective than Huber regulators in terms of edge-preserving. In this paper, we propose the combination of the user-friendly L_1 and L_2 norms to approximate the user-unfriendly L_q norm.**

***Index Terms*—Bayesian objective function, total variation, edge preservation, L_0, L_1, L_2 and L_q optimization.**

I. Introduction

IN incomplete-data tomography, for example, in limited angle tomography, the image reconstruction problem is severely ill-conditioned [1, 2]. Without regularization, the reconstructed images contain too many artifacts to have any practical applications. An effective way to incorporate regularization is through an iterative algorithm that has Bayesian terms. The total variation (TV) regulator is effective in stabilizing the reconstruction procedure, reducing noise, preserving sharp edges, and promoting piecewise-constant images.

In section II, the two versions of the famous TV functionals for two-dimensional images are discussed. These two versions are isotropic TV and anisotropic TV [3]. Huber functional is another popular functional used to denoise and preserve the sharp image edges [4]. Some other *ad hoc* Huber-lookalike functionals are suggested in this paper. The suggested new Huber-lookalike functionals are shown to be more effective in sharp-edge preserving than the original Huber functional.

In section III, some limited-angle computer simulations are carried out. The equivalence of the two TV versions is illustrated. The Huber-lookalike functions are compared when they are used in an iterative algorithm.

We believe that deep-learning-based image reconstruction is another form of Bayesian image reconstruction, which is not in the scope of this paper.

This work was supported in part by the National Institutes of Health under Grant 2R15EB024283.

G. L. Zeng is with the Department of Computer Science, Utah Valley University, Orem, UT 84058 USA (e-mail: larry.zeng@uvu.edu) and the Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, UT 84108 USA (larry.zeng@hsc.utah.edu).

II. Methods

A. Isotropic TV and Anisotropic TV Regulators

Anisotropic TV of a two-dimensional (2D) image $F(x,y)$ is defined as

$$TV_{aniso}(F) = \iint \|\nabla F\|_1 \, dxdy \tag{1}$$

with the local L_1 norm of the image gradient magnitude as

$$\|\nabla F\|_1 = |\nabla_x F| + |\nabla_y F| . \tag{2}$$

At every point (x,y), the gradient of a 2D function $F(x,y)$ is a vector $\left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}\right)$. The L_1 norm this gradient vector is defined as $\left|\frac{\partial F}{\partial x}\right| + \left|\frac{\partial F}{\partial y}\right|$. Therefore, the anisotropic TV of image $F(x,y)$ is the summation of the L_1 norms the gradient vectors $\left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}\right)$ over all points (x,y).

If we replace the L_1 norm with the *local* L_2 norm at each point (x,y), $\left|\frac{\partial F}{\partial x}\right| + \left|\frac{\partial F}{\partial y}\right|$ becomes $\sqrt{\left(\frac{\partial F}{\partial x}\right)^2 + \left(\frac{\partial F}{\partial y}\right)^2}$. As a result, the isotropic TV of a 2D image $F(x,y)$ is defined as the summation of the *local* L_2 norms the gradient vectors $\left(\frac{\partial F}{\partial x}, \frac{\partial F}{\partial y}\right)$ over all points (x,y):

$$TV_{iso}(F) = \iint \|\nabla F\|_2 dxdy. \tag{3}$$

with the local L_2 norm of the image gradient magnitude as

$$\|\nabla F\|_2 = \sqrt{(\nabla_x F)^2 + \left(\nabla_y F\right)^2} . \tag{4}$$

We point out that the integral in (3) is the sum of the square-roots, which are the "local" L_2 norms of the partial gradients at a point. On the other hand, the L_2 norm of the gradient images is the square-root of the sum of the partial gradient images with respect to x and with respect to y, as expressed in (5) below:

$$\|\nabla F\|_2 = \sqrt{\iint \left[(\nabla_x F)^2 + \left(\nabla_y F\right)^2\right] dxdy}. \tag{5}$$

The name 'isotropic TV' implies that the value of $TV_{iso}(F)$ is rotational invariant. Let us consider a rotated coordinate system

$$\begin{pmatrix} s \\ t \end{pmatrix} = \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{6}$$

Then we have

$$\left(\frac{\partial F}{\partial s}\right)^2+\left(\frac{\partial F}{\partial t}\right)^2$$
$$=\left(\frac{\partial F}{\partial x}cos\theta+\frac{\partial F}{\partial y}sin\theta\right)^2+\left(-\frac{\partial F}{\partial x}sin\theta+\frac{\partial F}{\partial y}cos\theta\right)^2$$
$$=\left(\frac{\partial F}{\partial x}\right)^2cos^2\theta+2\frac{\partial F}{\partial x}\frac{\partial F}{\partial y}cos\theta sin\theta+\left(\frac{\partial F}{\partial y}\right)^2sin^2\theta$$
$$+\left(\frac{\partial F}{\partial x}\right)^2sin^2\theta-2\frac{\partial F}{\partial x}\frac{\partial F}{\partial y}cos\theta sin\theta+\left(\frac{\partial F}{\partial y}\right)^2cos^2\theta$$
$$=\left(\frac{\partial F}{\partial x}\right)^2+\left(\frac{\partial F}{\partial y}\right)^2. \quad (7)$$

The anisotropic TV is sensitive to horizontal/vertical edges, while the isotropic TV is rotation invariant.

B. Huber Regulators

A Huber regulator is expressed as a combination of a quadratic penalty for small values and a linear penalty for large values as follows.

$$H(g)=\begin{cases}\frac{1}{2}g^2, & |g|\le c\\ c|g|-\frac{1}{2}c^2, & |g|>c.\end{cases} \quad (8)$$

In (8), g is image gradient $\sqrt{(\nabla_x F)^2+\left(\nabla_y F\right)^2}$ given in (4).

C. Huber-Lookalike Regulators

We now define some more regulators. Similar to the Huber regulator, we will define some regulators that have different expressions in different regions. We set a threshold c. When the magnitude of the gradient is greater than c, $|g|>c$, the gradient of the regulator is assigned as a zero-mean random variable. When $|g|<c$, the regulators are defined as

$$I(g)=\frac{1}{2}g^2; \quad (9)$$

$$J(g)=\|\nabla F\|_2-\frac{\alpha}{2}g^2; \quad (10)$$

$$K(g)=\|\nabla F\|_1-\frac{\alpha}{2}g^2. \quad (11)$$

In the region of $|g|<c$, regulator I in (9) is the square of the local L_2 norm of the image gradient, that is, $(\nabla_x F)^2+\left(\nabla_y F\right)^2$.

In the region of $|g|<c$, regulator J in (10) is a combination of $TV_{iso}(F)$ and the L_2 norm. The parameter α can be positive or negative.

In the region of $|g|<c$, regulator J in (11) is a combination of $TV_{aniso}(F)$ and the L_2 norm. To make sure J and K are nondecreasing, we require that $1-\alpha c\ge 0$, that is, $\alpha\le 1/c$.

Both local L_1 norm and local L_2 norm behave like an L_1 norm. The motivation for these combinations is to approximate an L_q norm, with $0<q<1$. Some curves are shown on the interval [0, 0.5] in Fig. 1. If the second order derivative of the function is positive (negative), the curve is convex (concave). The L_q norm is defined by the function of $y=x^q$[5]. If $0<q<1$, $y"=q(q-1)x^{q-2}<0$ for $x>0$. Thus, $y=x^q$ is concave.

The function $y=x-\alpha x^2$ for $x>0$ and $\alpha>0$ is also concave. An advantage of using $y=x-\alpha x^2$ to replace $y=x^q$ is that $x=0$ is a singular point for the first order derivative of $y=x^q$. On the other hand, there is no singularity for any order derivative of $y=x-\alpha x^2$.

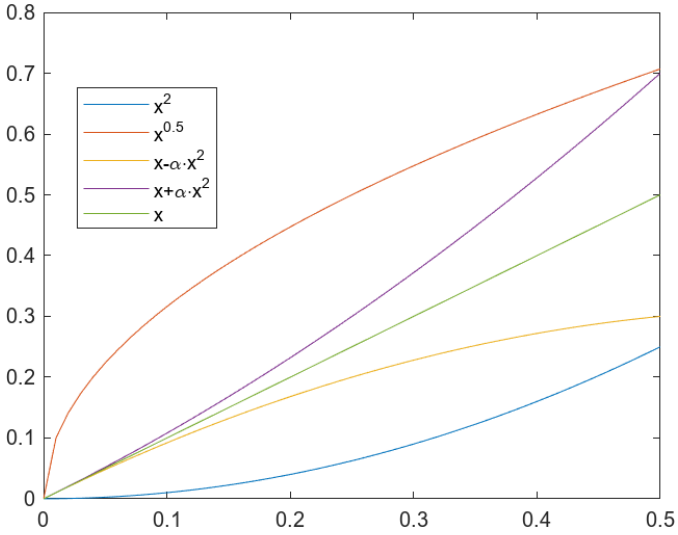


Fig. 1. Some monotonic functions for $0\le x\le 0.5$.

Some one-dimensional profiles of derivative cases are depicted in Figs. 2-5, respectively. The purpose of 'zero mean random' is to give the algorithm a chance to get out of a local minimum during an optimization procedure.

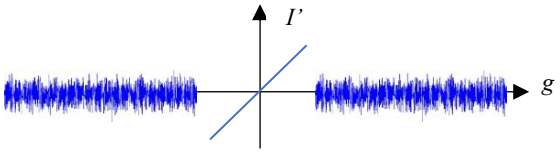


Fig. 2. The gradient of penalty *I*.

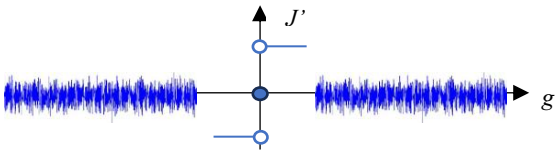


Fig. 3. The gradient of penalty *J* with $\alpha=0$.

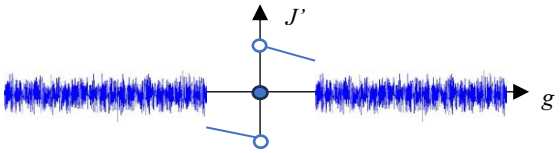


Fig. 4. The gradient of penalty *J* with $\alpha>0$.

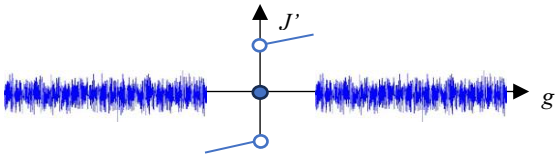


Fig. 5. The gradient of penalty *J* with $\alpha<0$.

D. Computer Simulations

A computer simulated noiseless phantom, and its rotated version were used in the computer simulations. All images were 256×256. The imaging geometry was parallel-beam, and the scan-angle was 40° with 400 views.

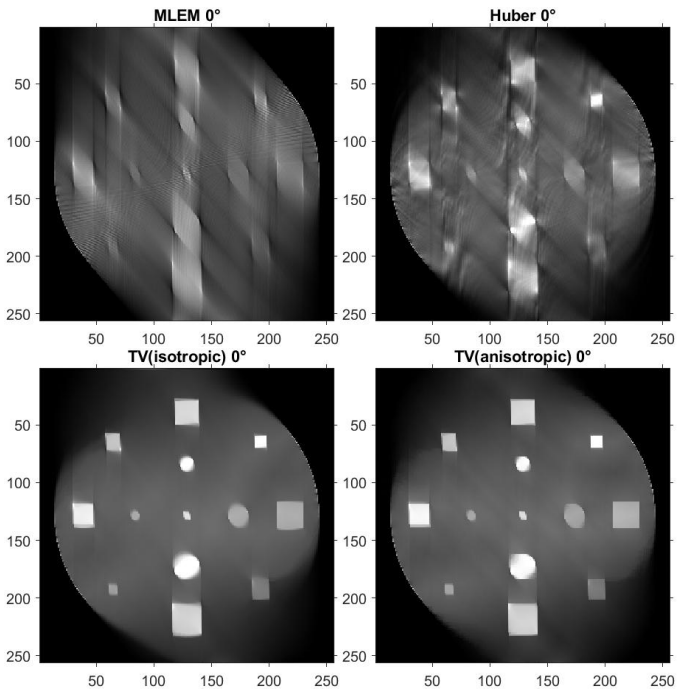
There are many image reconstruction algorithms. In this paper, we adopted a "Projections onto Convex Sets" (POCS) algorithm [6], in which the images were reconstructed with 10000 iterations, each iteration containing 10 iterations of the "Maximum-Likelihood Expectation-Maximization" (MLEM) algorithm [7] and one iteration of the gradient descent algorithm [8]. The MLEM algorithm enforces data fidelity, and the gradient descent algorithm minimizes the regulation function of the image gradient.

When the magnitude of the gradient is less than parameter c, the Regulator I (9) and Huber Regulator (8) are the L_2 norm of the gradient. The parameter c was set as 0.5 in our simulations. The phantom took 4 values: 0, 0.5, 1, and 1.5. The parameter α was set as 0, 0.001, and -0.001, respectively.

III. Results

Two sets of reconstructed images are shown in this section. The first set of the reconstructed images are shown in Fig. 6. After rotating the phantom by 45°, the second set of the reconstructed images are shown in Fig. 7. The structural similarity index measure (SSIM) metrics [9] are calculated and listed in Table 1 for comparisons.

The scanning angle was 40°, which is too small to provide enough data for any useful reconstructions. The purpose of the comparison studies here is to illustrate the different effectiveness of the regulators.



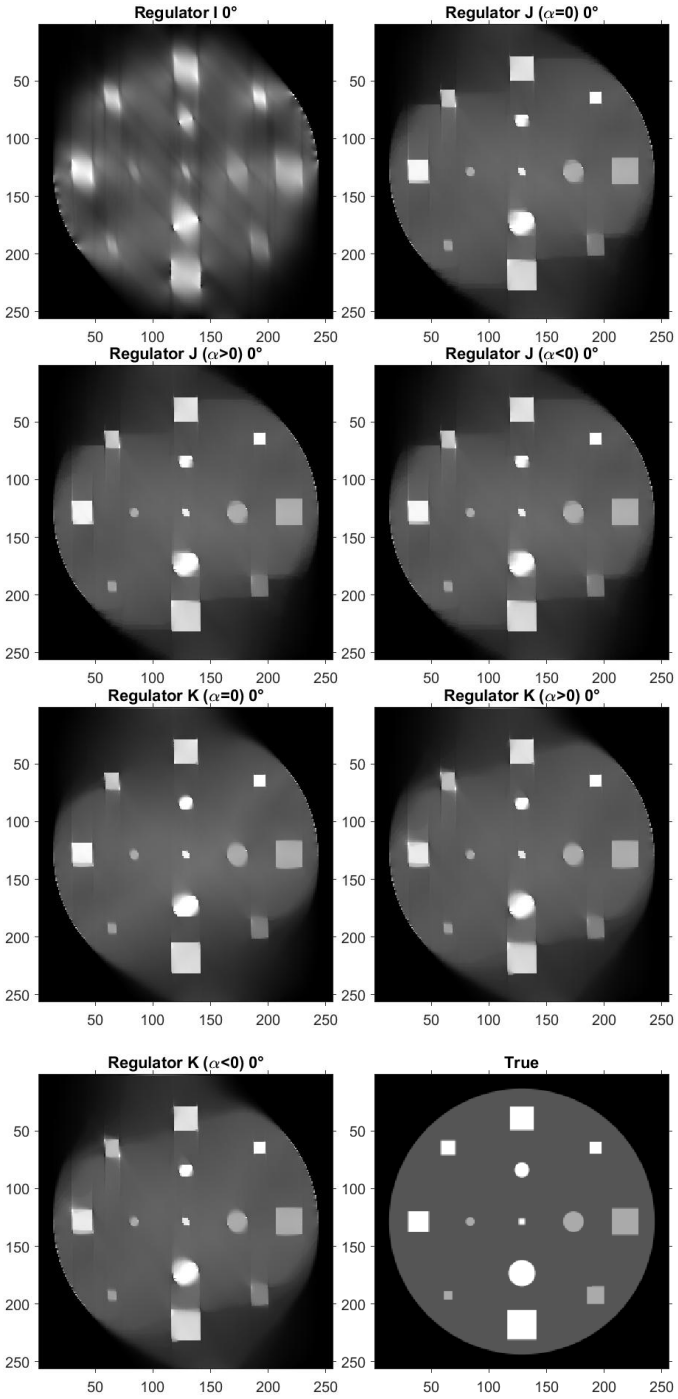
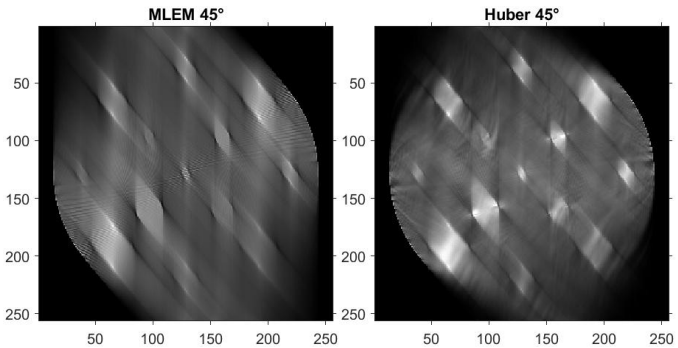


Fig. 6. Reconstructed images without phantom rotation.



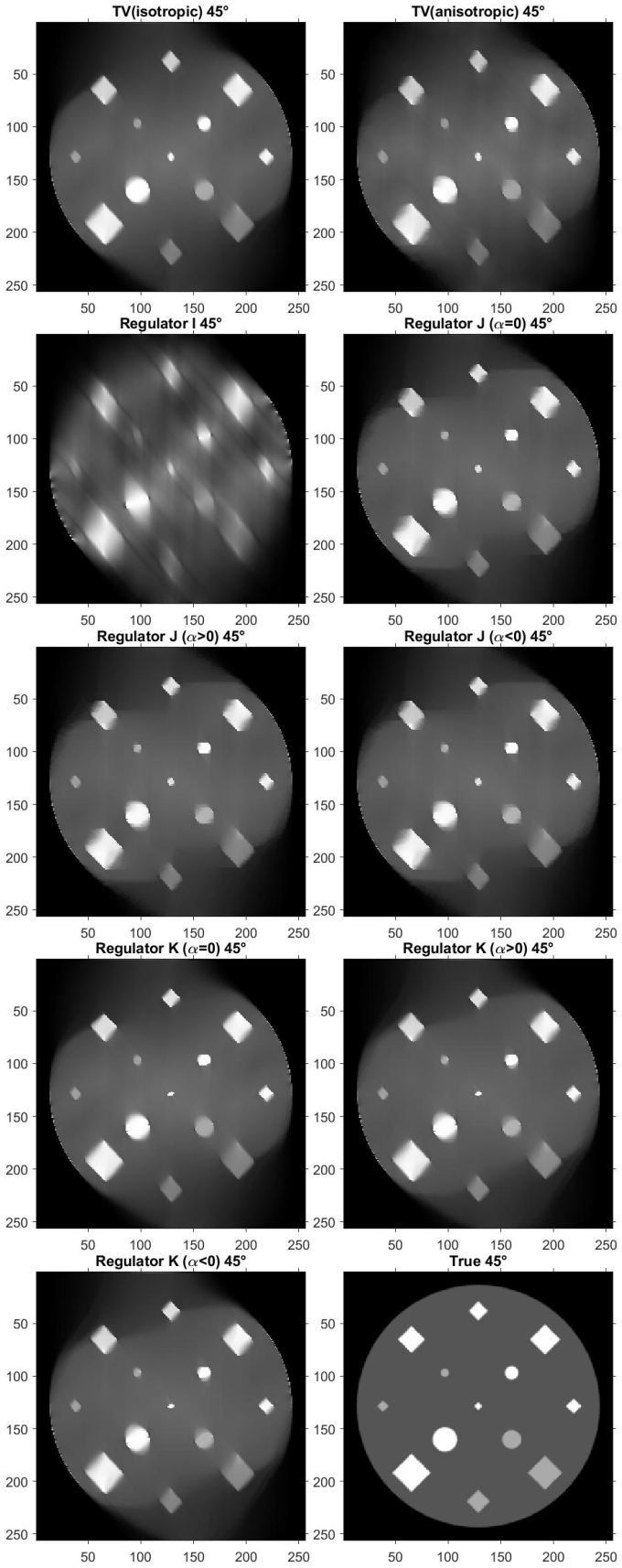


Fig. 7. Reconstructed images with phantom rotation.by 45°.

Table 1. SSIM for reconstructions (A bigger value is a better vaue)

Regulator	0° (Fig. 6)	45° (Fig. 7)
MLEM	0.5417	0.5420
Huber	0.5540	0.5828
TV(isotropic)	0.7451	0.7447
TV(anisotropic)	0.7417	0.7297
I	0.6285	0.6191
J, α = 0	0.7455	0.7383
J, α = 0.01	0.7450	0.7379
J, α = -0.01	0.7457	0.7378
K, α = 0	0.7461	0.7443
K, α = 0.01	**0.7498**	**0.7517**
K, α = -0.01	**0.7498**	0.7515

IV. Discussion and Conclusions

Bayesian regulators are effective in image reconstruction especially when projection data is incomplete, for example, in limited angle tomography. Popular Bayesian regulators include the TV norms, L_1 norm, L_0 norm, and L_q norm. Taking the derivative of an L_0 norm or an L_q norm can be problematic, due to the singularity at the origin when $0 \le q < 1$.

This paper proposes the use of regulator $y = |x| - \frac{\alpha}{2}|x|^2$ for a small positive α to replace the L_q regulator $y = |x|^q$ for $0 < q < 1$. When $x > 0$, both functions are positive (if $0 < \alpha \le 2/c$); the first-order derivative functions are positive (if $0 < \alpha \le 1/c$); the second-order derivative functions are negative (if $0 < \alpha$). When $x \to 0$, the first-order derivative function of $y = |x| - \frac{\alpha}{2}|x|^2$ has finite left and right limits; however, the first-order derivative function of $y = |x|^q$ does not have finite left and right limits.

Our computer simulation examples demonstrate better performance of our proposed regulators than the popular TV regulators and the Huber regulator.

References

[1] Davison, M.E., 1983. The ill-conditioned nature of the limited angle tomography problem. *SIAM Journal on Applied Mathematics*, *43*(2), pp.428-448.

[2] Barutcu, S., Aslan, S., Katsaggelos, A.K. and Gürsoy, D., 2021. Limited-angle computed tomography with deep image and physics priors. *Scientific reports*, *11*(1), p.17740.

[3] Shi, Y. and Chang, Q., 2013. Efficient algorithm for isotropic and anisotropic total variation deblurring and denoising. *Journal of Applied Mathematics*, *2013*(1), p.797239.

[4] Bai, F., Franchois, A. and Pizurica, A., 2016. 3D microwave tomography with Huber regularization applied to realistic numerical breast phantoms. *Progress In Electromagnetics Research*, *155*, pp.75-91.

[5] Livadiotis, G., 2020. General fitting methods based on Lq norms and their optimization. *Stats*, 3(1), pp.16-31.

[6] Cetin, A.E., Bozkurt, A., Gunay, O., Habiboglu, Y.H., Kose, K., Onaran, I., Tofighi, M. and Sevimli, R.A., 2013, December. Projections onto convex sets (POCS) based optimization by lifting. In *2013 IEEE Global Conference on Signal and Information Processing* (pp. 623-623). IEEE.

[7] Shepp, L.A. and Vardi, Y., 2007. Maximum likelihood reconstruction for emission tomography. *IEEE transactions on medical imaging*, *1*(2), pp.113-122.

[8] Meza, J.C., 2010. Steepest descent. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*(6), pp.719-722.

[9] Hore, A. and Ziou, D., 2010, August. Image quality metrics: PSNR vs. SSIM. In *2010 20th international conference on pattern recognition* (pp. 2366-2369). IEEE.

Breaking the G-Force Barrier in CT: A Sixth-Generation Architecture Using a Stationary Magnetically Steered X-ray Source

Hui Hu, Vitaliy Ziskin, Andrew Cross, Hong Chen
Nautilus X-Ray, Inc., Henderson, NV, USA Corresponding author: h.hu@nautilusxray.com

Abstract **Current third-generation CT scanners are fundamentally limited by the centrifugal forces (G-forces) experienced by the rotating X-ray tube, constraining temporal resolution limit to ~140 ms. We introduce a sixth-generation CT architecture that combines a novel stationary magnetically steered (SMS) X-ray source with a conventional rotating detector-collimator-filter assembly. The SMS source uses a compact toroidal vacuum envelope and a two-stage magnetic steering system to sweep the focal spot across a 230° C-shaped anode ring in 30 ms, ~ 5× faster than state-of-the-art rotating tubes. By keeping the heavy X-ray source stationary, rotating mass is reduced by ~75% (~250 kg), enabling ~2x faster gantry rotation. Prototype results show focal spot size ~1.2 × 1.0 mm, tube output of 550 mA at 90 kV, and sufficient angular coverage for short-scan reconstruction. The design preserves third-generation advantages in scatter rejection and bowtie filtration while remaining compatible with photon-counting detectors, reconstructions and AI developments. Projected temporal resolution reaches ~75 ms with optimized 480+ RPM rotation, offering a pathway to motion-robust cardiac, pediatric, and trauma imaging without clinical compromises.**

Index Terms: Temporal resolution, stationary X-ray source, magnetically steered electrons, sixth-generation CT, cardiac CT

I. Introduction

Third-generation CT has been the clinical standard for over 50 years, but mechanical rotation of a ~1000 kg gantry imposes fundamental G-force limit on speed. Previous stationary-source approaches (e.g., electron-beam CT, fifth-generation) failed clinically due to poor scatter rejection, limited z-coverage and other reasons. We present a hybrid architecture that retains a rotating detector assembly while using a stationary magnetically steered (SMS) source to overcome this long-standing barrier.

II. Prototype Design & Methods

We designed a stationary magnetically steered (SMS) X-ray source, featuring an electron gun and a toroidal vacuum envelope enclosing a 230° C-shaped anode ring. (Figs 1-3) The SMS source guides electrons onto targeted focal spots along the stationary C-shaped anode ring. This mechanism generates X-ray projections across the requisite angular range of 230° surrounding the region of interest without rotating the tube.

The SMS source is further designed with a two-stage magnetic steering system (transport + deflection) (Fig. 2-3). Electron transport along the ring tube envelope is achieved with Helmholtz coils (122A/B) with its B-field parallel to ring axis. The electron deflection from the transport path to strike specific focal spots on the anode ring to generate X-ray is controlled by distributed deflection coils (such as Helmholtz coils on non-magnetic yokes, 124A/B).

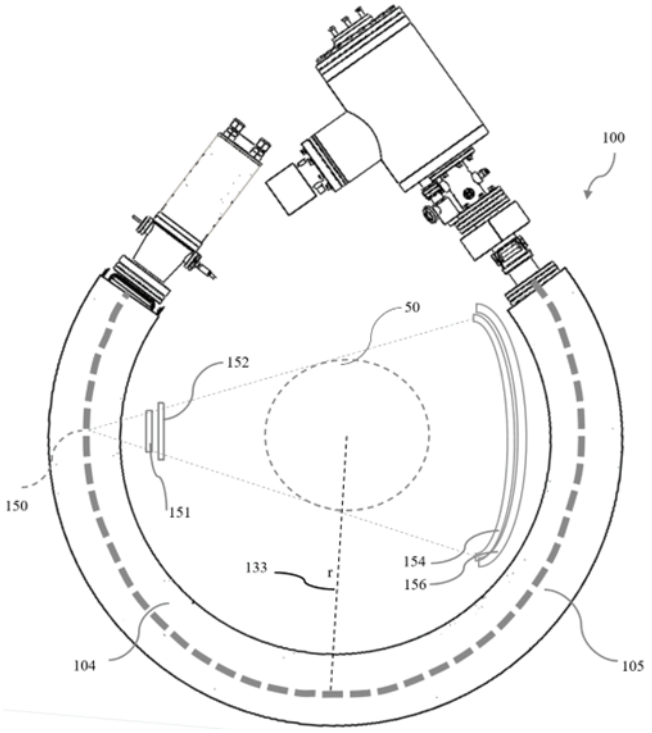


Fig.1 Schematic top view of the novel Stationary Magnetically Steered (SMS) X-ray source and a sixth-generation CT architecture. The SMS source (100) includes an electron gun, a scan tube (104), a C-shaped anode ring (105) and one focal spot (150). The sixth-generation CT architecture includes the SMS source (100), a rotating X-ray detector array (156), a post-object anti-scatter collimator (ASG,154), a X-ray filter (such as bowtie) (151), a pre-object collimator (152), and the region of interest (50).

We further designed a hybrid architecture (Fig 1, 4), termed sixth-generation CT architecture, that uses rotating detectors, collimator, and filter similar to the third-generation CT while the SMS source remains stationary. Focal-spot sweep is synchronized with detector rotation (Fig. 1, Fig. 4).

This architecture retains the critical geometric relations of a third-generation scanner – preserving critical strengths of third-generation CT such as scatter rejection via anti-scatter grids (154) and dose efficiency via rotating bowtie filters (151) – while breaking the G-force limit associated with rotating the massive X-ray tubes.

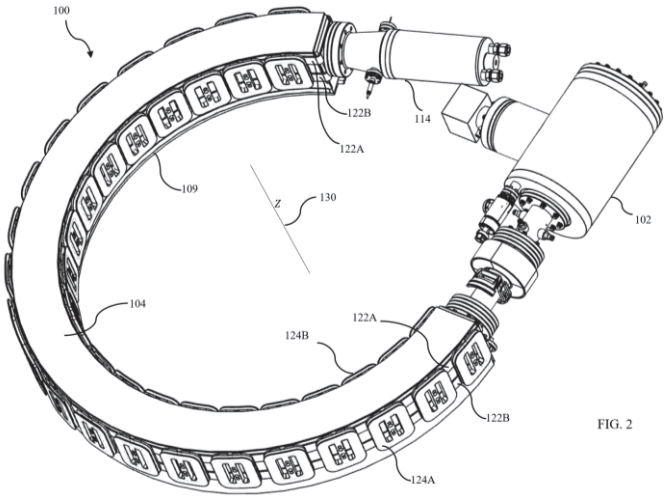


Fig.2 Perspective expanded view the novel SMS X-ray source of Fig. 1, showing a toroidal enclosure and an open bore.

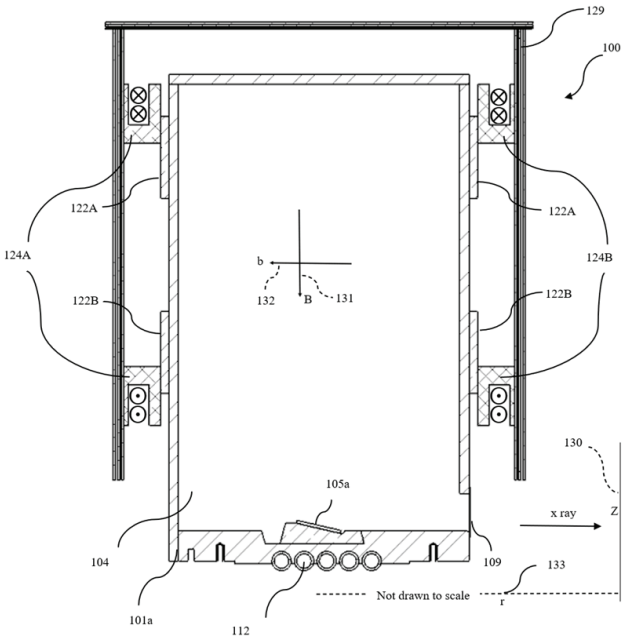


Fig. 3. Cross-sectional view perpendicular to the C-shaped anode ring, illustrating sidewall (101a), anode segment (105a), ring axis (130), radius (133), transporting magnetic field (131), and deflecting magnetic field (132)

Fig. 4 shows a photograph of a recent prototype we developed while Figs. 1-3 show the design drawings. Performance of our prototype was evaluated via:

- Wire-array current-pulse timing for sweep speed
- Rotating pinhole camera for focal spot size and shape
- High-voltage power supply readout for flux
- Structural analysis for gantry mass and speed margins

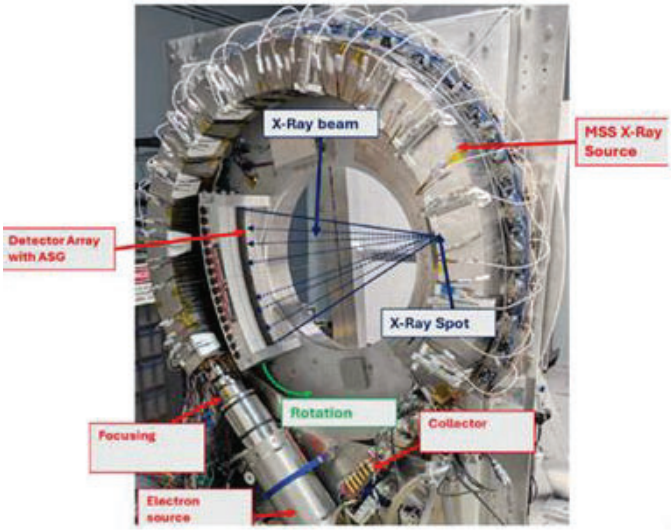


Fig. 4: Photograph of a recent prototype

III. Results

- *Focal spot sweep time*: 30 ms over 230° arc (≈5× faster than modern rotating tubes)



Fig. 5. Oscilloscope traces showing current pulses induced in a series of electrically isolated wires placed along the C-shaped anode ring. As the magnetically steered electron beam sweeps across the anode, it sequentially crosses each wire, generating distinct current pulses. The total duration from the first to the last pulse, marked by the red arrow, measures 30 ms - corresponding to the complete 230° X-ray focal spot sweep time.

- *Tube performance:* 550 mA at 90 kV

- *Angular coverage:* 230° (sufficient for complete short-scan reconstruction)

- *Rotating mass:* ~250 kg (~75% reduction of conventional third-generation CT)

- *Current gantry speed:* 270 RPM; design supports >480 RPM (projected 125 ms/rotation)

- *Focal spot size*: ~1.2 × 1.0 mm² (measured via rotating pinhole, at Full Width 15% of Maximum)

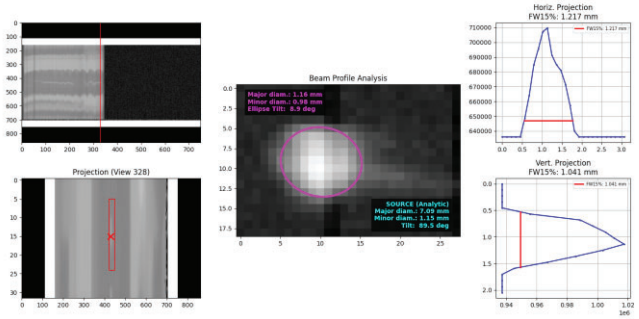


Fig. 6 Rotating pinhole camera measurement of the SMS X-ray focal spot. Central panel: 2D focal spot image with profile outline (purple). Horizontal and vertical projections (right) yield FW15% = 1.2 x 1.0 mm, respectively.

Table 1 summarizes the performance of our prototype, its near-term achievable (with italic font) and compares them with the third-generation architecture.

IV. Discussion

This sixth-generation architecture decouples the heavy X-ray tube from gantry rotation, overcoming the primary mechanical G-force barrier that has limited CT temporal resolution for nearly 50 years. By keeping the source stationary while retaining a rotating detector-collimator-filter assembly, the design preserves the core strengths of third-generation CT -including excellent scatter rejection via rotating anti-scatter grids and dose optimization via a rotating bowtie filter- while eliminating the dominant inertial constraint.

The 30 ms focal-spot sweep achieved by the SMS source is independent of mechanical gantry rotation and is not the fundamental speed limit of the source. With the ~75% reduction in rotating mass (~250 kg) enabled by the sixth-generation architecture, substantially higher (~2×) gantry speeds is achievable while remaining within existing structural safety margins. The next prototype, with hardware optimized for doubled rotational speed, targets >480 RPM (125 ms per rotation), yielding a projected system temporal resolution of ~75 ms - approximately double the performance of current state-of-the-art scanners.

Although the current focal spot size (~1.2 × 1.0 mm) is larger than modern clinical standards, it is sufficient for many general-purpose applications, and targeted optimizations (beam focusing and deflection coil refinement) are underway.

The platform is designed as a general-purpose CT scanner, with the ~2× speed gain being particularly valuable for motion-sensitive applications such as cardiac, pediatric, and trauma imaging.

Future work will focus on system development, integration, optimization, radiation-dose characterization, regulatory (FDA 510(k)) clearance, and in vivo clinical validation.

The architecture is directly compatible with existing and emerging technologies, including photon-counting detectors, dual-source configurations, spectral imaging, advanced dose-reduction techniques, reconstructions algorithms, and AI-based algorithms.

V. Conclusion

We have demonstrated a functional prototype of a stationary magnetically steered X-ray source integrated into a sixth-generation CT architecture. By dramatically reducing rotating mass and achieving a 30 ms source sweep, this platform provides a realistic and manufacturable pathway to sub-100 ms temporal resolution in a general-purpose scanner. This innovation has the potential to transform motion-robust, real-time CT imaging across multiple clinical domains.

References

1. Behling R. Modern Diagnostic X-Ray Sources: Technology, Manufacturing, Reliability. 2nd ed. CRC Press; 2021.
2. Kalisz K, Buethe J, Saboo SS, Abbara S, Halliburton S. Rajiah P Artifacts at Cardiac CT: Physics and Solutions. RadioGraphics. 2016;36(7):2064-2086. doi:10.1148/rg.2016160079
3. Jokerst, C. E. & Rajiah, P. S. Practical tips and tricks for salvaging ECG-gated cardiac CT from motion artifacts. RadioGraphics 45, e240247 (2025). doi:10.1148/rg.240247
4. Ren, P., He, Y., Zhu, Y., Zhang, T., Cao, J., Wang, Z. & Yang, Z. Motion artefact reduction in coronary CT angiography images with a deep learning method. BMC Medical Imaging22, 184 (2022). doi:10.1186/s12880-022-00914-2
5. Yao, J. et al. Deep learning-based motion correction algorithm for coronary CT angiography: Lowering the phase requirement for morphological and functional evaluation. Journal of Applied Clinical Medical Physics24, e14104 (2023).
6. Alkhodari, M., Alefishat, E., Jelinek, H. F. et al. Enhancing CCTA image quality: a review of deep learning approaches for advanced artifact correction and denoising. Artificial Intelligence Review58, 331 (2025). doi:10.1007/s10462-025-11311-w
7. Boyd, D. P. & Lipton, M. J. Cardiac computed tomography. Proc. IEEE 71, 298–307 (1983)
8. Parker DL. Optimal short scan convolution reconstruction for fanbeam CT. Med Phys. 1982;9(2):254-257.
9. Silver MD. A method for including redundant data in computed tomography. Med Phys. 2000;27(4):773-774.
10. Hsieh, J. *Computed Tomography: Principles, Design, Artifacts, and Recent Advances* (SPIE, 2009).
11. Tang, X. *Spectral Multi-Detector Computed Tomography (sMDCT): Data Acquisition, Image Formation, Quality Assessment and Contrast Enhancement* (Series in Medical Physics and Biomedical Engineering)
12. Flohr, T. G. et al. Cardiac CT imaging with a dual-source CT scanner. European Radiology 16, 256–268 (2006).
13. Achenbach, S. et al. Coronary computed tomography angiography with a consistent dose below 1 mSv using prospective electrocardiogram-triggered high-pitch spiral acquisition. *European Heart Journal* **31**, 340–346 (2010).
14. Flohr T, Petersilka M, Henning A, et al. Photon-counting CT review. Physica Medica. 2020;79:126-136. doi:10.1016/j.ejmp.2020.10.030
15. Willemink, M. J. et al. Photon-counting CT: Technical Principles and Clinical Prospects. *Radiology* **289**, 293–312 (2018).
16. McCollough, C. H. et al. Dual-Energy CT: A Phantom Comparison of Different Platforms for Abdominal Imaging. *Academic Radiology* **16**, 1398–1410 (2009).

Table 1 Generations/Architectures Comparison		
Feature / Parameter	3rd Gen Gold-Standard Benchmark	6th Gen Achieved / *Projected*
Source Motion	Mechanical Rotation (Gantry)	Stationary (SMS Sweep)
Detector Motion	Mechanical Rotation (Gantry)	Mechanical Rotation (Synchronized)
Geometry	Compact (Gantry)	Compact (Toroidal/Short-Throw)
Patient Access	Open Bore (Front/Rear)	Open Bore (Front/Rear)
X-Ray source speed (ms)	N/A	30
Gantry Speed per Rotation (ms)	>230	250 /*125*
Rotating Mass (kg)	~1000	*~250*
Temporal Resolution (ms)	146 Tube Rotation limited	<30 ms Source 160/*75* ms Gantry Ultra-Fast
Detector sample rate (μs)	200	200/*50*
Max Flux /Emission Power (kW)		550 mA/ 90 kV
Focal Spot Size (mm)		~1.2 × 1.0 mm^2
Z-coverage in one heartbeat (mm)	40-160	20/*160*
Compatibility with Photon-Counting detectors Dual Source Spectral CT Dose Reduction More detector rows Reconstruction AI technologies	✓	Directly Compatible
Scatter Rejection using ASG (Anti-Scatter Grids)	High / Focused ASG 1D,2D	Same
Bowtie Filter	Yes	Same
Both General and dedicated CT	✓	✓
Bore / FOV (mm)	700,780+/500	Same

Table 1 summarizes the performance of our prototype, its near-term achievable (with italic font) and compares them with the third-generation architecture.

A Static Dual-Ring CT System with Integrated Cold-Cathode X-Ray Source and Detector Arrays: A Simulation Study

Bohong Ye, Yule Li, Yankun Wang, Hao Wang, Jingyi Liao, Limin Li, Zhaoying Bian, Zhipeng Zhang, Jianhua Ma, and Dong Zeng

***Abstract*—Conventional CT imaging relies on thermionic X-ray tubes and mechanical rotation via slip rings, which limits temporal resolution and poses challenges for real-time imaging of dynamic organs such as the heart. Mechanical systems also introduce constraints related to size, maintenance, and vibration. In contrast, zinc oxide (ZnO) cold-cathode X-ray sources offer rapid electron emission, compact structure, low power consumption, and programmable multiplexed exposure, making them suitable for stationary CT architectures. By arranging sources and detectors in a fixed coaxial dual-ring layout, mechanical rotation can be eliminated entirely. This work presents a static dual-ring CT (SDR-CT) system design in which a ring of cold-cathode X-ray sources is closely integrated with an outer ring of detectors, enabling full 360° data acquisition through electronic scanning without any moving parts. Using the Geant4 Monte Carlo platform, we built a full simulation model that accurately represents X-ray spectra, photon interaction with phantoms, and detector response. The simulation confirms the correct operation of the dual-ring scanning sequence and yields realistic projection data. This design not only supports high-temporal-resolution imaging but also favors miniaturization and integration, offering a promising direction for portable and efficient CT systems.**

***Index Terms*—Computed tomography (CT), Cold cathode X-ray source, Stationary CT, Dual-ring architecture, Monte Carlo simulation**

I. Introduction

SINCE the first CT scanner was introduced in the 1970s, computed tomography has grown into a fundamental imaging tool essential to both medical diagnosis and industrial nondestructive testing [1]. Driven by ongoing advances in detectors, X-ray sources, and reconstruction algorithms, computed tomography (CT) has experienced rapid development [2], [3]. Over the following decades, advances such as slip-ring technology allowed third-generation CT systems to perform continuous helical scanning, greatly enhancing imaging efficiency. Yet despite these improvements, rising clinical needs have increasingly highlighted the limited temporal resolution of conventional CT, particularly in imaging dynamic structures like the coronary arteries [4]. Current mechanically rotating systems face inherent constraints, i.e., the high centrifugal forces during rapid gantry rotation, and push rotational speeds toward a practical ceiling, leaving little room for further gains. Beyond speed, mechanical rotation also introduces persistent challenges. It not only restricts imaging velocity and tends to produce motion artifacts when scanning moving subjects, but also leads to considerable system bulk, thermal management difficulties, and elevated maintenance costs [5].

In recent years, advances in static CT have the potential to address the inherent limitations of conventional CT systems. Static CT technology replaces mechanical rotation with fixed, distributed arrays of X-ray sources and detectors. By electronically controlling the timing of each source, multi-angle scanning is achieved without moving parts [6]. This approach removes the mechanical stress, wear, and complexity associated with traditional slip rings and rotating gantries. As a result, static CT systems can achieve faster imaging, reduced operational complexity, extended service life, and greater flexibility in deployment. They also offer distinct benefits for high-temporal-resolution imaging of dynamic organs such as the heart.

The advancement of cold cathode X-ray sources has provided fundamental support for the realization of static CT. As a novel type of nanomaterial-based X-ray source, the zinc oxide cold cathode emitter utilizes field emission to deliver key advantages including instant room-temperature startup, compact size, low power consumption, and extended operational life [7], [8]. These properties allow for dense circular array arrangements, thereby offering essential hardware foundations for the compact design of static CT systems. Building on this technology, this work introduces a static dual-ring CT (SDR-CT) imaging system, featuring an architecture that integrates cold cathode X-ray sources and detectors. The system adopts a coaxial and parallel dual-ring

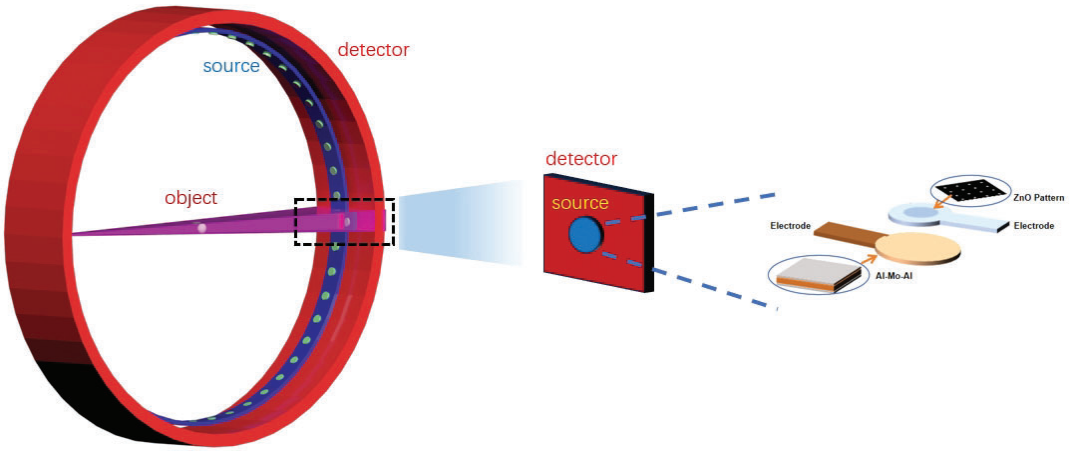
This work was supported in part by the National Key R&D Program of China under Grant 2024YFA1012000 and 2024YFC2417800, the National Natural Science Foundation of China under Grant 62571228, Beijing Natural Science Foundation under Grant Z250002, and Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515140146. (Corresponding author: Hao Wang, Dong Zeng.)

Bohong Ye, Yule Li, Yankun Wang, Hao Wang, Jingyi Liao, Zhaoying Bian, and Dong Zeng are with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China (e-mail: ybh22520356@smu.edu.cn; ylli@smu.edu.cn; ykwang@smu.edu.cn; wang_hao@smu.edu.cn; 1jy@smu.edu.cn; zybian@smu.edu.cn; zd1989@smu.edu.cn).

Limin Li is with School of Mathematics and Statistics in Xi'an Jiaotong University, Xi'an 710049, China (e-mail: liminli@xjtu.edu.cn).

Zhipeng Zhang is with School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510275, China (e-mail: zhangzhp25@mail.sysu.edu.cn).

Jiahua Ma is with School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: jhma@smu.deu.cn).



(a) (b)

Fig. 1. (a) Schematic of a dual-ring static CT system with a fixed inner X-ray source array and an outer curved detector array operating in a cone-beam geometry. (b) Schematic of a ZnO nanowire-based cold cathode X-ray source unit enabling electrically controlled multi-source X-ray emission.

layout that enables symmetrically sequenced imaging [9]. In this design, the source ring is positioned immediately adjacent to the detector ring, departing from the conventional separated and approach reduces redundant space, optimizes the internal arrangement, and contributes to a lighter and more portable system.

Using the Geant4 Monte Carlo simulation toolkit [10], we developed a full physical model of the SDR-CT system (Fig. 1). The simulation covers the complete imaging chain from electron beam excitation to energy deposition in the detector, allowing a feasibility assessment of the proposed architecture. Results demonstrate that the system retains imaging performance comparable to conventional CT while offering improved geometric symmetry and higher spatial efficiency. This design provides a promising research direction and technical pathway for developing the next generation of portable, integrated, and miniaturized CT devices.

II. Methods

This paper presents a SDR-CT system, which is built around a zinc oxide cold-cathode X-ray source array, as shown in Fig. 1(a). The source consists of a two-dimensional array of cold-cathode electron sources that produce large-area X-ray emission by bombarding an anode target. The anode target material is fabricated in a three-layer structure of aluminum, molybdenum, and aluminum, as shown in Fig. 1(b). To maintain simulation accuracy, the energy spectrum file of the cold-cathode X-ray source was incorporated into the model, accurately representing both electron beam excitation and photon energy distribution, as shown in Fig. 2(b).

Unlike the traditional rotational CT geometry, the SDR-CT system employs a compact, low-power cold cathode X-ray source array. Arranged in a full 360° ring, 100 sources are positioned at 3.6° intervals to cover the entire scanning range. As illustrated in Fig. 1(a), the source ring (blue) is mounted directly inside the detector ring (red), creating a symmetrical integrated structure. Between each source and its adjacent detector, a dedicated thermal and electromagnetic shielding layer is incorporated to dissipate heat and block interference during operation. When the system is active, the X-ray sources are exposed sequentially, each illuminating its corresponding high-pixel detector to collect projection data.

To mitigate dark- stripe artifacts in the SDR-CT system, we developed a projection-domain polynomial regression correction algorithm. The method first fits a mapping relationship using projection data extracted from air regions in both the object scan and the blank (empty) scan. This mapping is then applied to adjust the affected regions in the blank-scan projections, targeting the locations of the dark stripes.

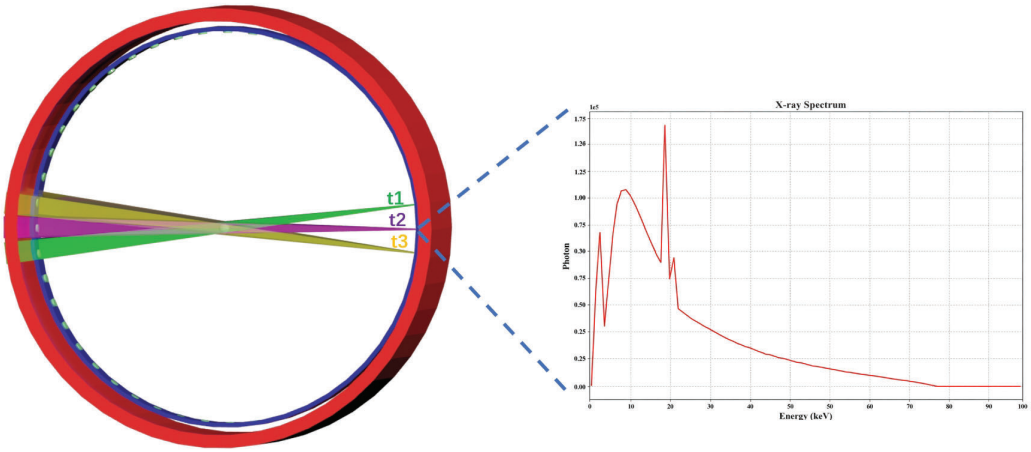
In this study, the imaging system was simulated using the Geant4 platform, where a physical model of the proposed system was implemented. The relevant simulation parameters are provided in TABLE I. Here, the source-to-isocenter distance (SID) and source-to-detector distance (SDD) define the system geometry and cone-beam magnification. CsI was used as the detector material. During operation, the radiation sources activate in sequence, capturing one projection every 3.6° to obtain a total of 100 projection frames.

TABLE I
Imaging System Parameters In The Simulation

Simulation	Value
Pixel size	0.2 mm × 0.2 mm
Detector size	100 mm × 157 mm
SID	286 mm
SDD	586 mm

III. Results

Fig. 3 shows a comparison of projection images of the water-sphere phantom obtained from the simulated system before and after correction, including the ground truth image as well as the corresponding absolute and signed residual maps used for quantitative error evaluation. In the experimental evaluation of the SDR-CT system, analysis of a single-angle



(a) (b)

Fig. 2. (a) Schematic of time-multiplexed multi-source illumination in the dual-ring static CT system, where different source groups (t1 - t3) are sequentially activated to acquire multi-view projection data. (b) Representative X-ray energy spectrum used in the simulation for polychromatic cone-beam imaging.

projection image revealed an axial dark-stripe artifact at the center of the logarithmically transformed data (Fig. 3, Uncorrected). This artifact arises because the Beer–Lambert law presumes a perfectly linear imaging system, whereas practical CT systems exhibit nonlinear responses and source-dependent fluctuations. Under polychromatic illumination, lower-energy photons are preferentially absorbed by the scanned object (a water-sphere phantom), which increases the mean energy of the transmitted spectrum, i.e., beam hardening. In the proposed dual-ring geometry, the physical presence of the source ring partially obstructs the X-ray beam. This obstruction not only amplifies the beam-hardening effect but also introduces additional scatter, collectively contributing to the persistent dark-stripe artifacts observed in the log-domain projections. As demonstrated in Fig. 3, the corrected logarithmic image shows a substantial reduction in the central dark-stripe artifact, though minor residual artifacts remain.

Residual analysis between the uncorrected, corrected, and ground truth images confirms that the deviation caused by the central dark stripe is notably decreased after correction. Quantitative evaluation further supports this improvement: as listed in TABLE II, the root-mean-square error (RMSE) between the corrected and ground truth images decreased by 30.79 % relative to the uncorrected, while the peak signal-to-noise ratio (PSNR) increased by 3.20 dB. These results verify the feasibility of the proposed correction approach in suppressing nonlinear artifacts induced by the structure.

TABLE II
QUANTITATIVE ASSESSMENT OF CORRECTION PERFORMANCE

	Ori vs. Ref	Corr vs. Ref	Improvement
RMSE	0.017163	0.011879	30.79%
PSNR	35.51 dB	38.70 dB	3.20 dB

Fig. 4 shows a comparison of projection images of the digital human phantom obtained from the simulated system before and after correction, including the ground truth image as well as the corresponding absolute and signed residual maps used for quantitative error evaluation. Table III lists the quantitative evaluation results of RMSE and PSNR in comparison with the ground truth image, indicating a 42.75% reduction in RMSE and a 4.84 dB improvement in PSNR, which further validate the effectiveness of the proposed correction algorithm in mitigating system nonlinear artifacts.

TABLE III
QUANTITATIVE ASSESSMENT OF CORRECTION PERFORMANCE

	Ori vs. Ref	Corr vs. Ref	Improvement
RMSE	0.020284	0.011613	42.75%
PSNR	40.40 dB	45.25 dB	4.84 dB

IV. CONCLUSION

This study introduces a stationary CT design based on a dual-ring coaxial configuration that integrates zinc oxide cold-cathode X-ray sources with an outer detector array. This approach addresses several limitations inherent in conventional CT systems, such as restricted temporal resolution, large physical footprint, and high maintenance demands. Using Geant4 Monte Carlo simulations, we modeled the system, acquired projection data, and evaluated its performance. The proposed architecture employs 100 cold-cathode sources densely arranged in an inner ring, paired with an outer detector ring, to achieve full 360° electronic scanning without mechanical rotation [11]. The closely integrated source–detector layout, together with optimized thermal management and shielding, provides a feasible route toward compact and portable CT systems. The experimental results demonstrate the efficiency of the proposed SDR-CT system. In the future, more comprehensive experiments will be conducted.

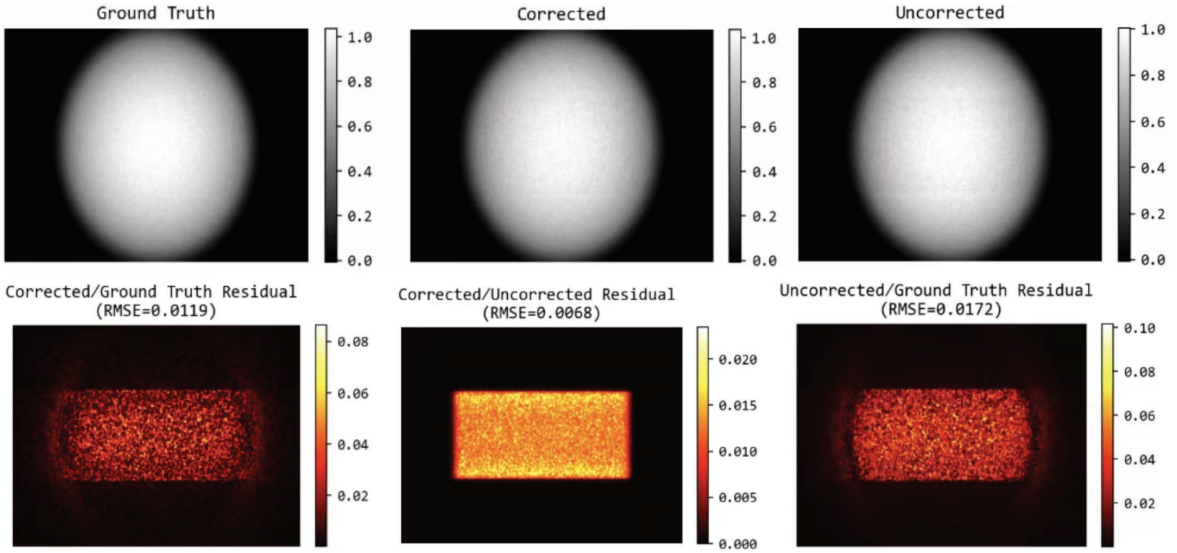


Fig. 3. Comparison of projection images before and after correction, including ground truth images and the corresponding absolute residual maps for quantitative error evaluation.

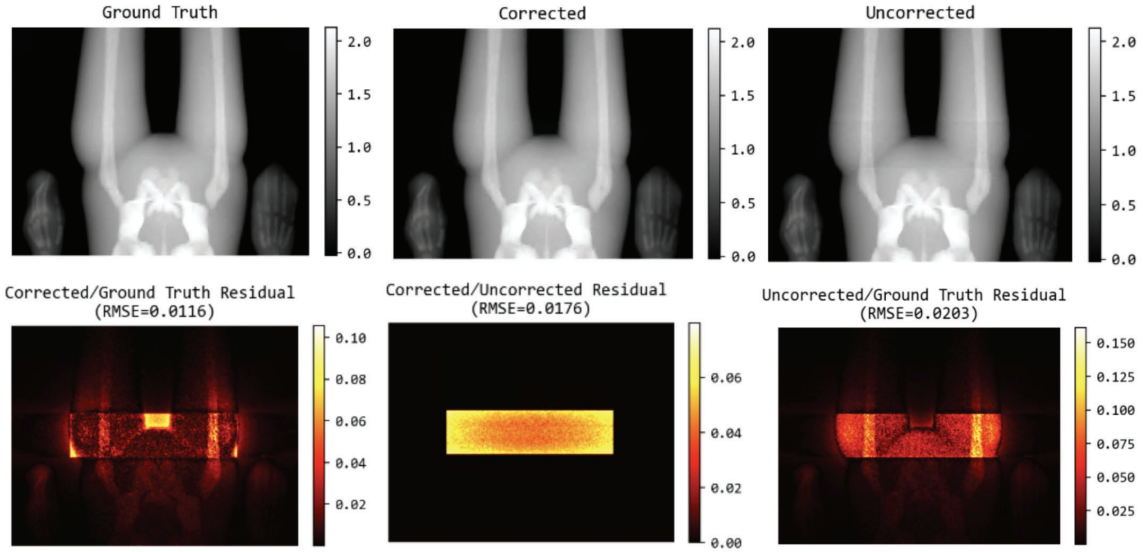


Fig. 4. Comparison of projection images of a digital human phantom before and after correction, with ground truth image and the corresponding absolute residual maps.

REFERENCES

[1] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. Description of system," *Br. J. Radiol.*, vol. 46, no. 552, pp. 1016–1022, 1973.

[2] D. Zeng et al., "Low-dose medical CT imaging technology: Past, present and future," *Chin. J. Stereol. Image Anal.*, vol. 27, no. 4, pp. 355–378, 2022, doi: 10.13505/j.1007-1482.2022.27.04.005.

[3] H. Wang et al., "Collaborative and privacy-preserving cross-vendor united diagnostic imaging via server-rotating federated machine learning," *Commun. Eng.*, vol. 4, no. 1, pp. 148, 2025, doi: 10.1038/s44172-025-00485-4.

[4] J. Hsieh and T. Flohr, "Computed tomography recent history and future perspectives," *J. Med. Imaging.*, vol. 8, no. 5, Art. no. 052109, Sep. 2021, doi: 10.1117/1.JMI.8.5.052109.

[5] N. J. Pelc, "Recent and future directions in CT imaging," *Ann. Biomed. Eng.*, vol. 42, no. 2, pp. 260–268, Feb. 2014, doi: 10.1007/s10439-014-0974-z.

[6] A. Cramer et al., "Stationary computed tomography for space and other resource-constrained environments," *Sci. Rep.*, vol. 8, no. 1, Art. no. 14195, Sep. 2018, doi: 10.1038/s41598-018-32505-z.

[7] C. Wang et al., "Fabrication of ZnO nanowire cold cathode flat-panel X-ray source with a reflective anode," *Nanomaterials*, vol. 14, no. 1, Art. no. 1504, 2024.

[8] M. Qi et al., "Dosimetric characterization for the flat-panel X-ray source: An electronic brachytherapy unit," *Med. Phys.*, vol. 52, no. 1, Art. no. e17988, 2025, doi: 10.1002/mp.17988.

[9] H. Ding et al., "Development and clinical application of a high-performance medical static computed tomography system," *J. Appl. Clin. Med. Phys.*, vol. 27, no. 1, Art. no. e70456, Jan. 2026, doi: 10.1002/acm2.70456.

[10] S. Agostinelli et al., "Geant4-A simulation toolkit," *Nucl. Instrum. Methods Phys. Res. A*, vol. 506, no. 3, pp. 250–303, 2003.

[11] D. Spronk et al., "Evaluation of carbon nanotube X-ray source array for stationary head computed tomography," *Med. Phys.*, vol. 48, no. 3, pp. 1089–1099, Mar. 2021, doi: 10.1002/mp.14696.

New directions for old detectors

Scott S. Hsieh

***Abstract*— The energy-integrating detector (EID) has powered diagnostic CT for more than 50 years. The prevailing view today is that EIDs are mature and that no major new changes are anticipated. We disagree with this assessment and examine three different components of the energy-integrating detector – the scintillator, reflector, and ASIC – and suggest how changes to these components could add new functionality. First, numerical modeling of gadolinium oxysulfide scintillator suggests that depth of interaction may be inferred through the color of the scintillated light. Second, we show through Monte Carlo simulations that a tilted geometry can hide the fill factor penalty of the reflector that ordinarily limits high resolution EIDs. Third, we show that a highly oversampling energy integrating ASIC might provide some spectral capability. EIDs designed according to these principles might provide a practical alternative to photon counting CT in cost-conscious markets.**

***Index Terms*—energy integrating detectors, spatial resolution, spectral imaging**

I. Introduction

ENERGY integrating detectors (EIDs) have served as the foundation for diagnostic CT since the debut of the EMI Mark I in 1971. While the basic principle of energy integration has been maintained over the past 50+ years, the components of the EID have undergone several major changes. The earliest EIDs used xenon gas to sense X-rays; this was soon replaced with cesium iodide (CsI), providing better stopping power in a more compact package. Finer pixelation of these detectors made multi-slice CT possible [1]. Improved application-specific integrated circuits (ASICs) reduced noise and increased frame rate, allowing fast gantry rotation times and techniques such as flying focal spot. Over the past two decades, however, major new innovations in the EID have ceased, and academic research has been minimal. The last major innovation was arguably dual-layer, proposed in 2005 [2]. Reduced noise ASICs have continued to appear since then but fundamentally new concepts have been scarce. Since 2010, the field has looked towards photon counting detectors (PCDs) as the future [3].

While PCDs do offer substantial advantages, we argue that the death knell for EIDs has been sounded too soon. The EID still offers several routes for improvement that have not been fully explored. Fig. 1 shows the different components of the EID. The redesign of each of these components might add new capabilities to the EID. First, X-rays incident on the EID are converted by the scintillator into pulses of scintillated light. This scintillated light is channeled via reflectors towards the pixel's photodiode, where it is converted to photoelectrons. The photocurrent is digitized by an application-specific integrated circuit (ASIC). Each of these components, the scintillator (Sec. II), the reflector (Sec. III), and the ASIC (Sec. IV), could be changed. Because of space limitations, we will describe the methods and results of these new innovations only briefly.

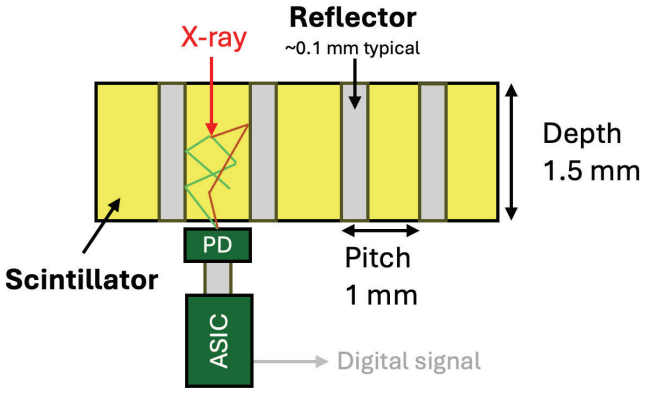


Fig. 1. Major components of the energy-integrating detector (EID), with typical dimensions. The scintillator converts X-ray into visible light. The reflector keeps the light contained with the pixel. The photodiode (PD) converts light into electrons, and the ASIC digitizes the photocurrent. A common scintillator, GOS, is depicted that produces both green and red light. One green light photon is depicted as self-absorbed by GOS. Drawing not to scale.

II. The Scintillator: Spectral Capability

The CsI scintillator used in early solid state EIDs was replaced by $CdWO_4$ (among other options) and eventually by gadolinium oxysulfide (GOS) [4]. Aside from the practical considerations of cost and stability, the key engineering figures of merit for a scintillator are light yield, transparency, and speed. GOS became dominant because of its outstanding light yield of ~60 photons/keV, about double that of $CdWO_4$ [5]. Its decay time of ~3 microseconds is faster than older scintillators, such as GE HiLight [6], but not as fast as garnets available today. The optical transparency of GOS, however, is mediocre. The optimal thickness of GOS was estimated around 2000 to be 1.4 mm [4]. Thicker GOS would increase X-ray stopping power but decrease detected scintillated light because of self-absorption.

While self-absorption is normally regarded as a non-ideality, we submit that it is actually an opportunity: a way to achieve spectral imaging with only a single layer of GOS. Our key observations are (1) a common formulation of GOS (Siemens UFC) has two emission peaks, green light at 669 nm

This work was supported in part by NIH R01EB035553. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

S. S. Hsieh is with the Mayo Clinic, Rochester MN 55902 (e-mail: Hsieh.Scott@mayo.edu).

(~75%) and red light at 512 nm (~25%), but (2) the self-absorption coefficient at 669 nm is orders of magnitude higher than that of 512 nm. These observations lead us to predict, on the basis of its datasheet, that the ratio of red to green light can be used to estimate depth of interaction. We will study the feasibility of this concept below, noting that it requires a "dual color" detector: segmenting the photodiode into red-green color-sensitive subpixels, analogous to the Bayer color filter used in consumer cameras.

A. Methods

We simulated the interaction of X-rays with GOS, assuming that the entire X-ray was converted into visible light at its first interaction (ignoring characteristic X-rays). The number of visible light photons followed a Poisson distribution (Fano factor = 1). Interaction depth was governed solely by linear attenuation coefficient. Optical transport characteristics were estimated from the Siemens UFC datasheet. We approximated the optical scattering coefficient as 400 cm^{-1} across red and green wavelengths. The absorption coefficient for red light is nearly zero (10^{-4} cm^{-1}), while for green light, it is 0.19 cm^{-1}.

We used a simple, 1D model of optical transport. Visible light photons live on a line segment, discretized at steps one scattering mean free path (25 microns) apart. After traveling one step, their travel direction is re-randomized and they have a chance for self-absorption. At one end of the line segment lies a reflector with 100% reflectivity. At the other end is the photodiode with 70% quantum efficiency across all energies. Our 1D model probably overestimates backscatter (in our model, photons are as likely to scatter backwards as forwards) and does not model losses from the reflector between pixels.

The GOS was irradiated by a 140 kV spectra filtered by 25 cm of water. The number of detected red and green quanta was used, without additional electronic noise, as the dual energy signal. This was compared to three reference systems: (1) an ideal dual layer EID with 1 mm YAG (a low-Z scintillator) as the front layer and a perfect absorber as the back layer, (2) an ideal GOS/GOS dual layer detector with 0.25 mm thickness as the front layer and 1.25 mm thickness as the back layer (no color signal but traditional depth segmentation), and (3) an ideal 2-bin PCD. By changing the object filtration slightly, we calculated the Cramér–Rao bound of the variance for material decomposition tasks.

B. Results

Fig. 2 shows the fraction of red/green quanta arriving on the photodiode as a function of interaction depth. The red quanta, which comprise 25% of the scintillated light, is rarely absorbed. The total amount of absorbed light varies from about 50% to 100% depending on depth of interaction. For comparison, Hell showed experimentally that increasing GOS thickness from 1 mm to 2 mm causes a loss of 30% of detected light [4]. This is a slightly different experiment but roughly consistent with the data in Fig. 2.

The CRLB of the material decomposition task shows that the spectral performance of this detector is expected to be about 75% worse (in standard deviation) than the YAG dual-layer detector. The GOS dual layer detector is intermediate. The difference between YAG dual layer and GOS dual layer can be attributed to the K-edge of GOS. The difference between GOS dual layer and the proposed dual color EID can be attributed to the color transmission function on the top panel of Fig 2. That is, if the optical transport was such that only red light were received between [0, 0.25] mm, and only green light were received between [0.25, 1.5] mm, the dual color EID should match GOS dual layer exactly. The optical transport is comparatively less efficient.

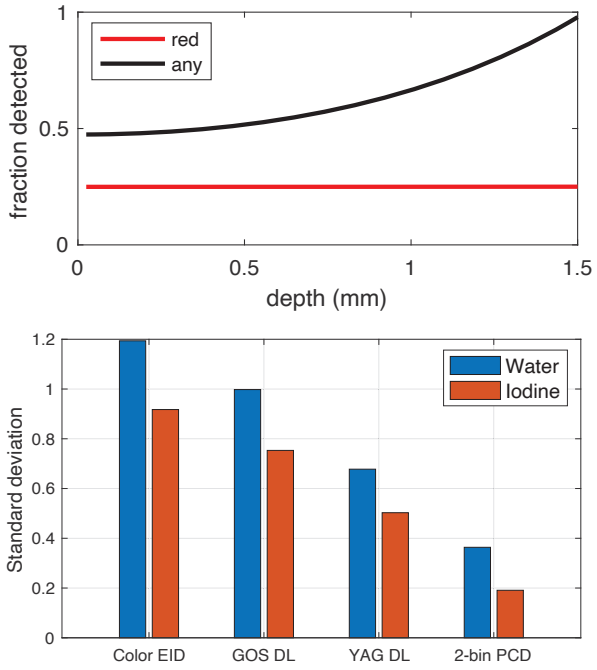


Fig. 2. (Top) Probability of light detection, for red or any light, as a function of interaction depth. Light signal becomes relatively greener as interaction depth increases. (Bottom) CRLB of material decomposition. DL is dual layer. YAG is the low-Z scintillator.

C. Discussion

There are two possible interpretations for Fig. 2. The pessimist might say: dual layer CT provides only middling spectral separation, so why pursue dual color scintillator CT when its spectral separation is worse? The optimist might say: the alternative to dual color EIDs is not really dual layer at all, but conventional single-energy CT.

The advantage of dual color EIDs is not their improved performance but their convenient packaging. The only changes that are needed are in the silicon circuitry: a color-segmented photodiode and associated changes to the ASIC. While dual color doubles the number of pixels that must be digitized, that does not necessarily mean that costs will increase proportionally. Cell phone cameras today digitize tens of megapixels (MP), but 64-slice CT has been digitizing ~50,000 pixels for the past two decades. There may be room for improvement. In the long run, the cost of producing dual color CT might be comparable to existing EIDs, in the same way that the 48 MP chipsets used in the iPhone 17 today probably have comparable production costs to the 2 MP chipsets for the first iPhone in 2007.

Besides this, it is fair to question if the results of Fig. 2 are

realistic. Our simulations of optical transport are simplistic and should be interpreted cautiously. Better yet, they should be corroborated experimentally.

III. The Reflector: Spatial Resolution

Many authors (including our own past work [7]) have stated that the spatial resolution of the EID is limited by the reflectors [8, 9]. If the reflector width is assumed fixed at ~0.1 mm, then smaller pixel pitch will cost a greater fraction of surface area to be consumed by the reflectors [10]. This decreasing fill factor causes more incident X-rays to be undetected. PCDs have no such restriction and have 100% fill factor (unless counting the external anti-scatter grid), and thus PCDs seem like the obvious choice for higher resolution CT.

We argue that if the reflector is tilted relative to the direction of the primary X-rays, then the incident X-rays can penetrate through the reflector and arrive at the scintillator underneath. This could enable practical construction of detectors at double the spatial resolution.

A. Methods

We used GEANT4 to model the transmission of X-rays through 3 mm of scintillator. This is thicker than most GOS scintillators but is compatible, for example, with HiLight [6] which was transparent but had long decay time (~1 ms). A thicker scintillator is helpful with the tilted geometry because the tilted geometry converts rays that were previously incident onto 100% reflector and 0% scintillator into rays that are incident onto N% reflector and (100-N)% scintillator, where N varies between 0 and 100 depending on the point of interaction. Therefore, there are rays that encounter very little scintillator, and it is helpful for the scintillator to be thicker. A modern option for the scintillator is single-crystal GAGG(Ce), which is nearly as bright as GOS, transparent, and with decay time of ~100 ns [11]. Our figure of merit was the zero-frequency detective quantum efficiency, or DQE(0).

B. Results

Fig. 3 shows the aperture and spatial resolution of the simulated pixel. The tilted geometry "fills out" more of the nominal pixel pitch. Its MTF is slightly worse than the conventional geometry at the same pixel pitch, but substantially better than the conventional geometry at a normal 1.0 mm pixel pitch.

The DQE(0) was calculated at 120 kV without an anti-scatter grid to be 72%, 58%, and 76% for the conventional 1.0 mm pitch, conventional 0.5 mm pitch, and tilted 0.5 mm pitch geometries, respectively. The conventional geometry pixels have DQE(0) worse than their fill factor because of the Swank factor. An anti-scatter grid reduces the advantage of tilting, depending significantly on the grid geometry. For example, a 1D grid (lamellae width, 0.15 mm; pitch, 1 mm) causes the DQE(0) to decrease to 68%, 53%, and 67%, respectively.

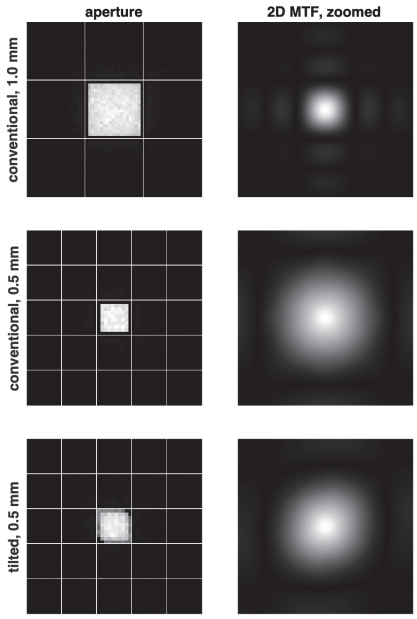


Fig. 3. Aperture from Monte Carlo simulations using 120 keV monoenergetic photons. The black border at the boundary of the pixel for the conventional geometries corresponds to poor fill factor, and this is reduced with the tilted geometry.

IV. The ASIC: Spectral Imaging

The ASIC digitizes the photocurrent from the photodiode and must operate at high speed, high precision, and low noise. Historically, ASICs for EIDs have developed with reduced noise as a major figure of merit.

Early in the multi-slice CT era, it was known that electronic noise was a major problem in highly attenuated rays. In 2008, a prototype EID ASIC demonstrated electronic noise equivalent to Poisson variance of 50 x-ray quanta. The substitution of $CdWO_4$ with GOS, which was twice as bright, reduced this crossing point by a factor of 4, and continuing improvements in electrical engineering have brought this figure of merit down further. In 2025, an EID ASIC (AS5952M, ams OSRAM) is available for purchase that specifies a maximum electronic noise of 2250 electrons, or about two x-ray quanta, despite being designed with affordability as a major goal.

In this context, there is little to gain from further electronic noise reduction. This is one reason why we think that the color photodiode concept (the AS5952M integrates both photodiodes and digitization electronics on a single chip) is an attractive path towards further development. Other paths exist that do not require color photodiodes, such as oversampling.

Fast oversampling is possible with modern analog-digital converters (ADCs). CT applications require 20-bit precision at a sampling frequency of ~5 kHz. Shen et al. describe a 16-bit SAR ADC that samples at 16 MHz in a silicon area of 0.55 mm^2 while consuming 16 mW [12]. By scaling the sampling rate down, the power is reduced. If GOS is used, the fastest useful digitization rate is ~320 kHz (given its decay time), meaning that the photodiode can be resampled 64 times during a 0.2 ms projection readout time. Adding together 64

subsamples allows us to populate 6 more bits, achieving 22 total bits of precision. With GAGG(Ce), sampling at several MHz could be useful. Variation within these samples can be used for dual energy imaging, as we will show below.

A. Methods

We modeled a 140 kV spectrum that is filtered by 25 cm of water, with a flux of 250 X-rays per view. This corresponds roughly to 0.5 mm pixel pitch, 0.3 ms view time, and 300 mA. For this feasibility study, we assume perfect integration, no discretization or electronic noise, and zero decay time. Each incident X-ray is randomly assigned to one of the subsamples, and we considered choices of 16, 64, or 256 subsamples. In addition to the usual signal of the total energy of subsamples (i.e., standard energy integrating), we considered a second dual energy signal that is the variance among the subsamples. From this, we calculated the CRLB.

B. Results

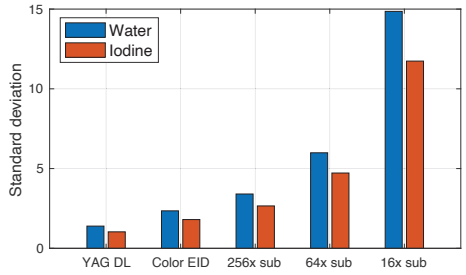


Fig. 4. Subsampling is a path towards material decomposition.

Fig. 4 shows that subsampling itself provides dual energy signal, although the quality of this is worse than the color photodiode approach unless a large number of subsamples can be used. To our knowledge, Roessl was the first to propose this basic mechanism for spectral imaging [13].

Another possibility for oversampling is single photon detection. If electronic noise is further reduced, the signal within each subsample can sometimes be thresholded to zero or one photon. This could provide Swank noise improvements in the low signal limit.

V. Conclusions

We have described three new ways in which future EIDs could be qualitatively different from past EIDs. Many of our proposals would use the same basic manufacturing process: fabricating a GOS ceramic scintillator wafer, structuring it into pixels using a fine wire saw, and bonding it to silicon electronics. While there are further improvements from using a transparent garnet scintillator, it is not strictly necessary for many of our concepts to see some improvement. The major manufacturing changes that we require are (1) the saw must make finer cuts, (2) the silicon circuitry should be redesigned, and (3) the module should be mounted at an angle. While no change to a CT scanner is "easy," these modifications could be achieved without major retooling of the existing production line.

The changes in the silicon circuitry, in particular, are easy to scale up. New ASICs are routinely introduced with new EID scanner models [14]. While design costs are high, once the ASIC is in mass production, it can use the economics of scale provided by the consumer electronics industry.

It is therefore possible to imagine future EIDs that capture some of the advantages of today's PCDs – high resolution, retrospective spectral capability, or higher dose efficiency – thus providing these clinical advantages to a larger patient population.

References

[1] W. A. Kalender, "X-ray computed tomography," *Physics in medicine & Biology,* vol. 51, no. 13, p. R29, 2006.

[2] R. Carmi, G. Naveh, and A. Altman, "Material separation with dual-layer CT," in *IEEE Nuclear Science Symposium Conference Record, 2005*, 2005, vol. 4: IEEE, pp. 3 pp.-1878.

[3] S. Kappler, F. Glasser, S. Janssen, E. Kraft, and M. Reinwand, "A research prototype system for quantum-counting clinical CT," in *Medical Imaging 2010: Physics of Medical Imaging*, 2010, vol. 7622: SPIE, pp. 664-669.

[4] E. Hell, W. Knüpfer, and D. Mattern, "The evolution of scintillating medical detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 454, no. 1, pp. 40-48, 2000.

[5] B. Heismann, L. Bätz, K. Pham-Gia, W. Metzger, D. Niederlöhner, and S. Wirth, "Signal transport in Computed Tomography detectors," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 591, no. 1, pp. 28-33, 2008.

[6] S. J. Duclos *et al.*, "Development of the HiLightTM scintillator for computed tomography medical imaging," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 505, no. 1-2, pp. 68-71, 2003.

[7] S. S. Hsieh, "Possible improvements in effective fill factor using X-ray fluorescent interpixel reflectors," *Medical physics,* vol. 51, no. 3, pp. 1617-1625, 2024.

[8] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: technical principles and clinical prospects," *Radiology,* vol. 289, no. 2, pp. 293-312, 2018.

[9] F. R. Schwartz, A. D. Sodickson, P. J. Pickhardt, D. V. Sahani, M. H. Lev, and R. Gupta, "Photon-counting CT: technology, current and potential future clinical applications, and overview of approved systems and those in Various stages of research and development," *Radiology,* vol. 314, no. 3, p. e240662, 2025.

[10] J. Hsieh and T. Flohr, "Computed tomography recent history and future perspectives," *Journal of Medical Imaging,* vol. 8, no. 5, pp. 052109-052109, 2021.

[11] K. Kamada *et al.*, "Crystal Growth and Scintillation Properties of Ce Doped ${\rm Gd} _ {3}({\rm Ga},{\rm Al}) _ {5}{\rm O} _ {12} $ Single Crystals," *IEEE Transactions on Nuclear Science,* vol. 59, no. 5, pp. 2112-2115, 2012.

[12] J. Shen *et al.*, "A 16-bit 16-MS/s SAR ADC with on-chip calibration in 55-nm CMOS," *IEEE Journal of Solid-State Circuits,* vol. 53, no. 4, pp. 1149-1160, 2018.

[13] E. Roessl, A. Thran, G. Martens, T. Istel, R. Proksa, and J.-P. Schlomka, "Dual-energy x-ray imaging by simultaneous integration and campbelling readout," in *IEEE Nuclear Science Symposium & Medical Imaging Conference*, 2010: IEEE, pp. 2112-2115.

[14] Y. Liu *et al.*, "Reducing image noise in computed tomography (CT) colonography: effect of an integrated circuit CT detector," *Journal of computer assisted tomography,* vol. 38, no. 3, pp. 398-403, 2014.

Practical Geometric Decoupling for Projection Unmixing in Multi-Source Linear-Array CT

Qiule Chen, Yule Li, Junhao Wu, Yankun Wang, Jingyi Liao, Yuanxiang Gao, Hao Wang, Jianhua Ma, and Dong Zeng

***Abstract*—Eliminating mechanical rotation through stationary multi-source computed tomography (CT) is critical for achieving high-speed scanning and reducing system complexity. However, the simultaneous acquisition required by such architectures inevitably leads to severe projection aliasing, where overlapping signals on the detector compromise reconstruction fidelity. This paper presents a Practical Geometric Decoupling (PGD) framework to address this challenge in multi-source linear-array CT. The proposed method begins with precise geometric calibration to map divergent source-detector geometries onto a unified angular coordinate system, thereby reformulating the projection aliasing problem into a tractable linear signal separation task. Subsequently, an efficient optimization model incorporating geometric consistency and smoothness constraints is introduced to robustly recover the de-aliased projections. Validation on a dual-source linear-array configuration—covering both dual-energy and variable-offset scenarios—demonstrates that PGD effectively eliminates overlapping artifacts and restores clear structural details from mixed measurements. Quantitative assessments reveal that the proposed method improves the Peak Signal-to-Noise Ratio (PSNR) by approximately 10 dB compared to direct reconstruction baselines and achieves a Structural Similarity Index (SSIM) exceeding 0.86. These results suggest that the proposed decoupling strategy offers a robust and scalable solution for projection unmixing in future multi-source and fully stationary CT systems.**

***Index Terms*—Stationary CT; Multi-source Imaging; Projection Aliasing; Geometric Decoupling;**

I. Introduction

Computed Tomography (CT) has been widely used in modern medical diagnosis, yet its reliance on mechanical gantry rotation fundamentally restricts temporal resolution, system stability, and device portability [1]. To overcome these limitations, stationary CT systems with fixed source-detector arrays have emerged as a promising solution, enabling real-time imaging through single-exposure acquisition using distributed X-ray sources, such as carbon nanotube (CNT) emitters [2],[3].These architectures, often employing inverse-geometry or linear-array designs, aim to eliminate motion blur and facilitate high-speed volumetric coverage [4].

However, in stationary CT, the shift from single-source to multi-source exposure introduces complex projection aliasing absent in conventional systems. In a shared detector configuration, overlapping signals from spatially distinct sources create ambiguous measurements that directly lead to severe structural artifacts and scatter contamination in the final reconstructed CT images [5]. Recent evaluations of multi-source configurations have highlighted the severity of these image quality degradations, specifically regarding scatter and dose efficiency [6],[7]. While some recent approaches have attempted to mitigate artifacts using deep learning or sparse sampling in distributed source designs [8], the aliasing problem remains particularly pronounced in systems where simultaneous multi-source exposure is essential for high-speed imaging. Consequently, effective signal separation becomes a non-negotiable requirement for diagnostic quality.

To address these limitations, we propose a Practical Geometric Decoupling (PGD) framework anchored in the geometric consistency of stationary CT architectures. While the framework is applicable to general multi-source systems, this study operationalizes PGD within a dual-source linear-array configuration. By mapping the divergent projection geometries onto a common angular domain, the framework transforms the complex aliasing challenge into a tractable linear separation problem solvable via regularized convex optimization [9]. Simulation results demonstrate that PGD effectively suppresses artifacts and enhances reconstruction fidelity, establishing a generalized foundation that offers clear potential for extension to higher-order multi-source systems.

II. Methods

2.1 Projection Aliasing Model Construction

As shown in Fig. 1, in dual-source linear-array CT, a global Cartesian coordinate system in the two-dimensional Euclidean space is built to describe the projection geometry. For source geometry, all X-ray sources are distributed along the line $y=D$. The spatial coordinate of the i-th source, S_i, is denoted as (x_i,D), where x_i represents the lateral offset from the central axis. For detector geometry, The detector is modeled as a linear array

This work was supported in part by the National Key R&D Program of China under Grant 2024YFA1012000 and 2024YFC2417800, the National Natural Science Foundation of China under Grant 62571228, Beijing Natural Science Foundation under Grant Z250002, and Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515140146. (Corresponding author:Dong Zeng.)

Qiule Chen, Yule Li, Junhao Wu, Yankun Wang, Jingyi Liao, Hao Wang, and Dong Zeng are with School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China (e-mail: cql@smu.edu.cn; ylli@smu.edu.cn; wujunhao@smu.edu.cn; ykwang@smu.edu.cn; 1jy@smu.edu.cn; gyx2001@smu.edu.cn; wang_hao@smu.edu.cn; zd1989@smu.edu.cn).

Limin Li is with School of Mathematics and Statistics in Xi'an Jiaotong University, Xi'an 710049, China (e-mail: liminli@xjtu.edu.cn).

Jiahua Ma is with School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: jhma@smu.deu.cn).

located at $y=-D^{'}$. The physical position of an arbitrary channel index k is given by $(k,-D^{'})$. the central ray angle is $\theta_i=\arctan(-x_i/D)$, and the half-fan-angle covering the ROI is $\alpha_i=\arcsin\left(R/\sqrt{x_i^2+D^2}\right)$. The effective projection interval K_i^{ROI} on the detector is defined by the intersection of the bounding rays with the detector plane:

$$K_i^{ROI}=\left[x_i+(D+D^{'})tan(\theta_i-\alpha_i),x_i+(D+D^{'})tan(\theta_i+\alpha_i)\right] \quad (1)$$

When multiple sources are active simultaneously, geometric aliasing occurs in the overlap region $K_{\cap}^{ROI}=K_1^{ROI}\cap K_2^{ROI}\cap\cdots$. Within this region, the detector signal is a superposition of varying ray paths, necessitating a decoupling strategy.

Without loss of generality, we derive the mathematical formulation based on a dual-source configuration, which serves as the fundamental unit of the multi-source architecture.

We denote by $s_1(\theta,k)$ and $s_2(\theta,k)$ the detector signals contributed by sources S_1 and S_2, respectively. In the overlap region, the measured intensity is modeled as:

$$I(\theta,k)=s_1(\theta,k)+s_2(\theta,k)+\eta \quad (2)$$

where each component follows the Beer–Lambert law, $s_i(\theta,k)=I_0e^{-P_i(\theta,k)}$, initial exposure intensity be defined I_0, $P_i(\theta,k)$ is the line integral of the attenuation coefficient along the ray associated with source $\boldsymbol{S_i}$ at (θ,k) ,η represents measurement noise.

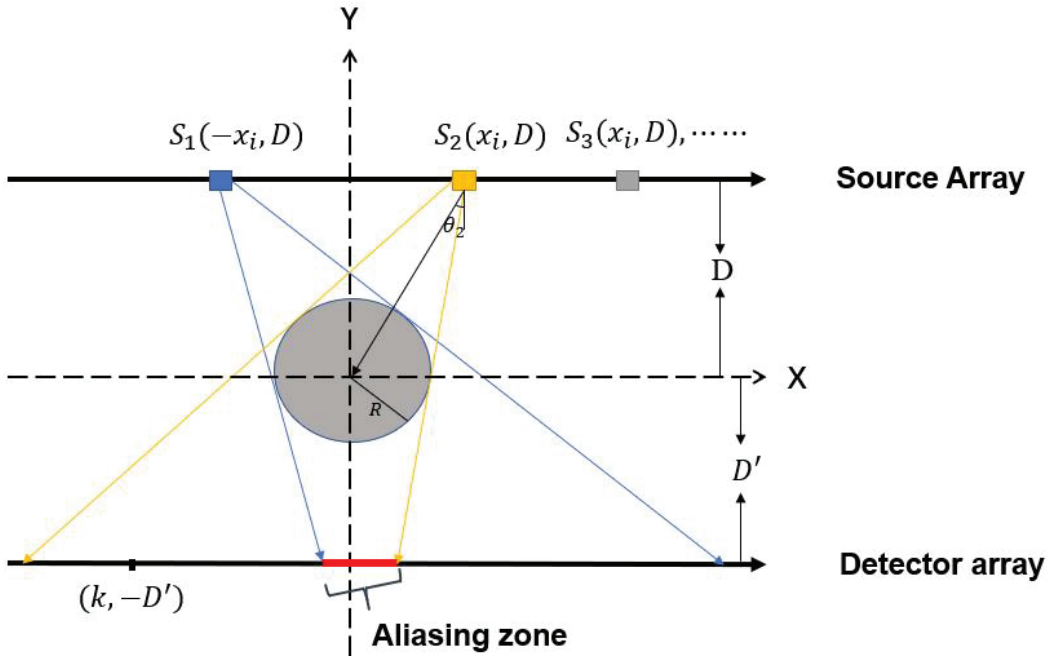


Fig. 1 System geometry of the dual-source linear-array CT. The schematic shows the coordinate system, source positions with lateral offsets, linear detector array, key parameters (DSO, DDO, fan-beam angles), overlap region on the detector, and projection aliasing from simultaneous multi-source exposure.

2.2 Unified Geometric Parameter Domain

Direct operations in the detector index domain are generally invalid for dual-source linear-array CT, because the same detector channel k corresponds to different ray directions for sources at different lateral positions ($x_1\neq x_2$). To establish a source-invariant representation, we introduce the relative fan-beam angle γ as a unified geometric parameter. For source $\boldsymbol{S_i}$, $\gamma_i(\theta,k)$ denotes the angle between its central ray and the ray hitting channel k at view θ. Since $\gamma_i(\theta,k)$ is strictly monotonic with respect to k under standard fan-beam geometry, an inter-source mapping $\kappa_{1\to2}$ can be defined by angular consistency:

$$k_1=\kappa_{1\to2}(k_2)\ s.t\ .\gamma_1(\theta,k)=\gamma_2(\theta,k) \quad (3)$$

This mapping re-parameterizes projections from different sources onto a common angular grid, enabling geometry-consistent comparison and subsequent separation. In practice, $\kappa_{1\to2}(k)$ is generally non-integer and is implemented via interpolation, which can be equivalently written as a sparse rebinning/transport operator $T_{1\to2}$ acting on the sampled detector data.

2.3 Decoupled Aliasing Model

According to formula (2), Using the geometric transport $T_{1\to2}$, the contribution of $\boldsymbol{S_1}$ can be expressed on the sampling grid of $\boldsymbol{S_2}$, yielding a geometry-aligned observation model:

$$I=M_1s_1+M_2s_2+\eta \quad (4)$$

where M_1 and M_2 are diagonal support masks indicating the illuminated channels for each source (and thus define the overlap and non-overlap regions). We then estimate (s_1, s_2) by solving a regularized least-squares problem:

$$\min_{s_1,s_2} J(s_1,s_2)=\frac{1}{2}|W\odot(I-M_1s_1-M_2s_2)|_2^2 +\mu|s_2-Ts_1|_2^2+\frac{\lambda}{2}\left(s_1^TLs_1+s_2^TLs_2\right) \quad (5)$$

subject to physical constraints $s_i\in[\epsilon,1]$. Here, W weights valid measurements, L is a 1D difference operator along the detector coordinate to promote smoothness, and (μ,λ) control

cross-consistency and regularization strength. The objective is convex in (s_1, s_2) and is efficiently minimized by alternating updates: fixing one component and solving a linear system for the other, followed by projection onto $[\epsilon,1]$. The recovered intensities are finally converted to attenuation projections via the log transform and reconstructed using a standard CT solver.

III. Experiments

In this work, we conduct simulation study to demonstrate the efficiency of PGD. In the simulation study, the geometry is listed in Table 1. Moreover, to simulate the data aliasing effects, We analyze the impact of the relative offset between two sources on the algorithm.

All algorithms were implemented in Python using the ASTRA toolbox and executed on a high-performance computing platform equipped with an AMD R7-5800H processor and an NVIDIA GeForce RTX 3060 GPU to ensure computational efficiency.

TABLE I
SCAN SIMULATION PARAMETER CONFIGURATION

Parameter type	Parameter value
DSO	800mm
DDO	500mm
Detector Pixel Array	512×512
Detector Pixel Size	1mm×1*m*m
Scanning angle range	0~180°
Number of projections	180
X-ray source offset	40,80,120mm

IV. Results

We first validate the applicability of the proposed framework within a dual-energy configuration (Fig. 2). The unmixed reconstructions for both the low- and high-energy channels preserve the dominant structural features of the phantom while substantially attenuating artifacts induced by signal mixing. Quantitatively, both channels achieve comparable fidelity (low-energy: PSNR ≈ 18.88 dB, SSIM ≈ 0.815; high-energy: PSNR ≈ 17.90 dB, SSIM ≈ 0.843), indicating that the proposed unmixing strategy effectively generalizes across energy channels and remains stable despite energy-dependent signal variations.

Furthermore, we evaluate the robustness of the method with respect to source separation distance by testing three distinct offsets (40, 80, and 120 mm) (Fig. 3). Across all tested offsets, direct reconstruction from mixed data consistently yields poor image quality (PSNR ≈ 12.7 dB, SSIM ≈ 0.51), suggesting that geometric disparity alone is insufficient to resolve projection aliasing. In contrast, reconstructions based on the proposed unmixing algorithm maintain high structural fidelity, exhibiting substantial quantitative gains (Offset 40 mm: PSNR ≈ 23.33 dB, SSIM ≈ 0.861; Offset 80 mm: PSNR ≈ 21.25 dB, SSIM ≈ 0.862; Offset 120 mm: PSNR ≈ 19.62 dB, SSIM ≈ 0.792). These results confirm the robustness of the proposed method under varying acquisition geometries.

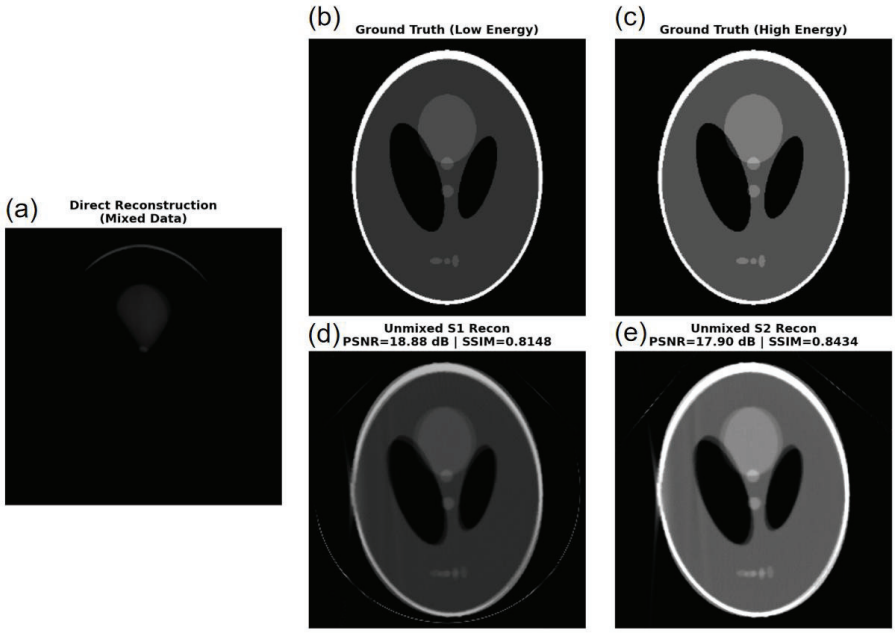


Fig. 2 Dual-energy dual-source unmixing results using the proposed PGD framework. (a) Direct reconstruction from mixed (aliased) projections showing severe overlapping artifacts. (b) Reconstructed low-energy channel after PGD unmixing (PSNR ≈ 18.88 dB, SSIM ≈ 0.815). (c) Reconstructed high-energy channel after PGD unmixing (PSNR ≈ 17.90 dB, SSIM ≈ 0.843). (d) Reference image (low-energy channel). (e) Reference image (high-energy channel).

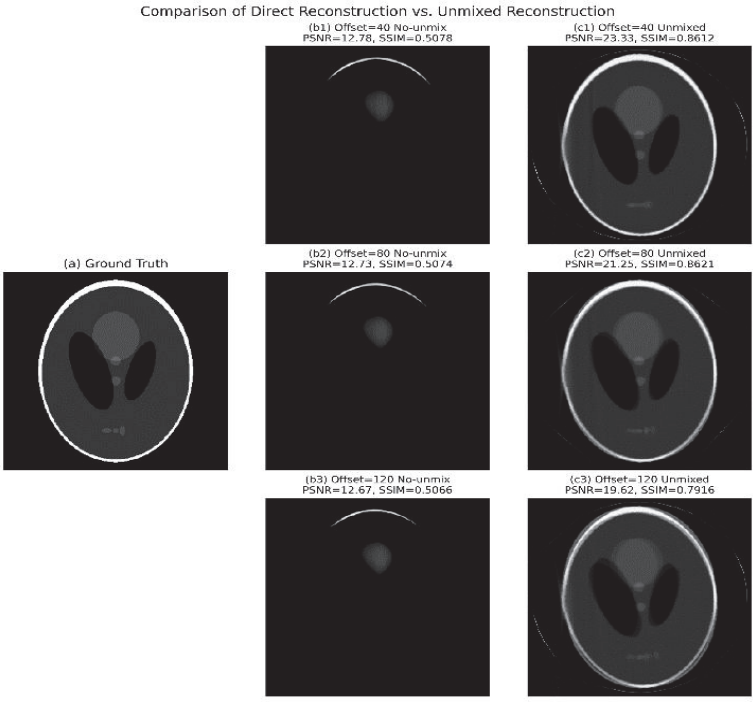


Fig. 3 Reconstruction comparison under different source offset distances (40 mm, 80 mm, 120 mm) in dual-source linear-array CT. (a) Direct reconstruction from mixed projections (severe aliasing artifacts, PSNR ≈ 12.7 dB, SSIM ≈ 0.51). (b1)–(b3) Direct reconstructions at 40 mm, 80 mm, and 120 mm offsets, respectively. (c1)–(c3) PGD-unmixed reconstructions at 40 mm, 80 mm, and 120 mm offsets, respectively (PSNR ≈ 23.33 / 21.25 / 19.62 dB; SSIM ≈ 0.861 / 0.862 / 0.792). The proposed PGD method effectively suppresses artifacts and maintains high image quality across increasing offsets, while direct reconstruction from mixed data shows consistent degradation.

V. Discussion and Conclusion

This study presents an analytical unmixing and reconstruction framework designed for dual-source linear CT to mitigate projection aliasing during simultaneous acquisition. The approach maps the two divergent scanning geometries into a common angular domain and introduces a geometry-consistent transfer operator, thereby reformulating the mixed measurements as a linear separation problem that is solved through regularized convex updates.

Simulations under dual-energy conditions reveal that direct reconstruction from multiplexed projections produces prominent high-intensity arc artifacts, which substantially obscure structural details of the phantom. In contrast, the proposed framework successfully recovers separated sinograms consistent with a single-source geometry, yielding reconstructions with suppressed artifacts and clearly resolved anatomical features. These findings indicate that the geometry-driven framework offers a scalable solution to projection aliasing in dual-source linear CT. The work provides a practical foundation for extending the approach to multi-source and fully stationary CT configurations.

References

[1] A. Cramer, J. Hecla, D. Wu, *et al*., "Stationary computed tomography for space and other resource-constrained environments," *Scientific Reports*, vol. 8, Art. no. 14195, 2018.

[2] D. Spronk, *et al*., "Evaluation of carbon nanotube x-ray source array for stationary head computed tomography," *Medical Physics*, 2021.

[3] Luo Y, Spronk D, Lee YZ, Zhou O, Lu J. Simulation on system configuration for stationary head CT using linear carbon nanotube x-ray source arrays. *Journal of Medical Imaging*. 2021; 8(5): 052114.

[4] J. Hsieh, J. Heanue, T. Funk, L. Hinshaw, E. Wilfley, and N. Solomon, "The feasibility of an inverse geometry CT system with stationary source arrays," *Medical Physics*, vol. 40, no. 3, Art. no. 031904, 2013.

[5] M. An and Y. Xie, "A Novel CT Imaging System with Adjacent Double X-Ray Sources," *Computational and Mathematical Methods in Medicine*, vol. 2013, Art. ID 391212, 2013.

[6] A. E. Becker, A. M. Hernandez, P. R. Schwoebel, and J. M. Boone, "Cone beam CT multisource configurations: Evaluating image quality, scatter, and dose using phantom imaging and Monte Carlo simulations," *Phys Med Biol,* 2020 Dec 18; 65(23): 235032.

[7] L. Neculaes, P. F. Edic, G. Frontera, B. De Man, R. Caiafa, and A. Wang, "Multisource X-Ray and CT: Lessons Learned and Future Outlook," *IEEE Access*, vol. 2, pp. 1568–1585, 2014.

[8] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "LEARN: Learned Experts' Assessment-Based Reconstruction Network for Sparse-Data CT," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1333–1347, Jun. 2018.

[9] Li, D., et al.Robust image reconstruction with real-world noise modeling for low-dose photon-counting detector CT. Pattern Recognition, 174, 112942. https://doi.org/10.1016/j.patcog.2025.112942

Understanding Radiologist Performance and Missed Lesion Detection

Parmita Mondal,[a] Akitoshi Inoue,[a] Rickey Carter,[b] Andrew D Missert,[a]
Joel G. Fletcher,[a] Hao Gong,[a] Shuai Leng,[a] Lifeng Yu,[a] Cynthia H. McCollough,[a] and Scott S. Hsieh[a]

[a]Mayo Clinic, Department of Radiology, Rochester, Minnesota, United States
[b]Mayo Clinic, Department of Quantitative Health Sciences Research,
Jacksonville, Florida, United States

***Abstract*: Our aim is to understand *why* radiologists miss liver metastases, especially at lower radiation doses. We know that detection performance drops with dose reduction, but the misses are heterogeneous. Previous eye-tracking work has characterized missed lesions into two categories: search-errors i.e., eyes never land on the lesion, and decision-errors i.e., the lesion is seen but not recognized as malignant. While the number of missed metastases increases at lower doses, the proportion of search vs. decision-errors is not known. To answer this, we integrated three prior reader studies. In all studies, radiologists interpreted the same set of 40 contrast-enhanced abdominal CT exams containing 91 liver metastases, where they marked hepatic metastases. In two studies, the workstation recorded the gaze and eye movements. Using eye dwell times, metastases were classified as search-error-dominant or decision-error-dominant using gaze time thresholds: 0.5, 1, or 2 seconds. In the third study, ten radiologists interpreted the exams, both at 120 and 200 quality reference mAs (QRM). This 3rd study did not include eye-tracking. In this study, 17% (163/910) of metastases were missed at 120 QRM and 12% (112/910) at 200 QRM. Using eye-tracker study, missed metastases were categorized into search or decision-error-dominant metastases. Decision-error-dominant metastases were missed at the same frequency at 120 QRM as in 200 QRM, but search-error-dominant metastases showed a trend for being more frequently missed at lower dose. Hence, even in a low contrast liver metastases detection, we find that many misses reflect decision-error-dominant, that show less sensitivity to reduced radiation dose.**

***Index Terms*: Hepatic metastases, Computed Tomography (CT), reader performance.**

I. INTRODUCTION

A central task for the radiologist is lesion detection. Lesion detection is impacted by several factors, including lesion contrast and conspicuity, the radiation dose of the acquisition protocol, the use of advanced reconstruction to reduce noise, and the performance of the individual radiologist. It is known that when radiation dose is reduced, image noise increases, and this obscure subtle lesions such as low-contrast metastases. For example, Goenka et al. showed that in a phantom model with 36 model metastases and 18 readers, detection performance was noninferior with 25% dose reduction, but inferior at 50% dose reduction even with iterative reconstruction [1]. In studies like these, the phantom lesions are unambiguous, and the only task is visual search. In clinical practice, lesions must further be classified into benign or malignant entities, and this adds another point of failure.

Missed lesions can thus be classified as having two main types of errors: search errors and decision errors. A search error occurs when the radiologist's eyes never actually land on the lesion, or when the eye gaze is so brief that there is no conscious recognition of a possible lesion. In contrast, a decision error occurs when the lesion is fixated upon for a period of time, but the radiologist chooses to categorize it as benign when it is malignant (or vice versa), reflecting a cognitive or interpretive failure rather than a perceptual one [2]. In this study, the lesions were marked as search-error-dominant, if the search error was more than the decision error and decision-error-dominant, if the decision error was more than the search error for the specific lesion. Eye tracking measures precisely where and for how long a reader's gaze lingers on specific image areas and thus provides a way to classify missed lesions into these categories. Eye tracking requires the experimenter to select a time threshold to perform this classification [3]. A reasonable threshold is 1 second and missed lesions with gaze times less than this would be coded as a search error, while longer gaze times would be coded as a decision error; but search and decision errors lie on a spectrum, and other thresholds could be considered. Eye-tracking provides an objective, measurable link between visual behavior, cognitive decision-making, and diagnostic performance, helping explain why even experienced radiologists can differ in accuracy.

To better understand the effects of dose, reader variability, and search versus decision errors, our institution has conducted several reader performance studies. We integrated three prior reader studies. In all studies, each reader reviewed the same set of 40 challenging CT cases (32 containing 91 liver metastases and eight normal). The first two studies classify the lesions into search and decision-error-dominant metastases. The third study with no eye-tracking data was used to detect low and high dosage. Figure 1 shows example metastases and provides examples of detected lesions by all readers, search-dominant misses,

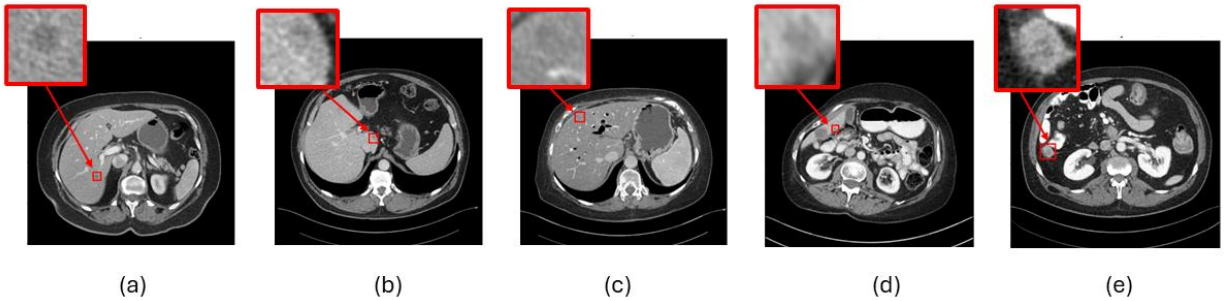


Figure 1. Liver metastatic lesions reshaped to 512 x 512 pixels. Example metastases (a) detected by all readers. (b) search dominant lesions (c) decision dominant lesions (d) five radiologists detected this lesion (R2, R3, R4, R5, R6) (e) five radiologists detected this lesion (R0, R2, R3, R4, R8)

decision-dominant misses, and lesions identified by only a subset of radiologists. Fletcher et al. examined reader performance at different dose levels with a cohort of ten abdominal subspecialists and found that performance diminished at lower dose. Hsieh et al.[4] examined reader variability at routine radiation dose using eye-tracking hardware and found that abdominal subspecialists demonstrated better detection (area under the jackknife alternative free-response receiver operating characteristic (JAFROC) curve= 0.77) than trainees (JAFROC AUC = 0.71) or non-abdominal radiologists (JAFROC AUC = 0.69), though sensitivity was similar across groups. They also found that higher sensitivity correlated with longer interpretation time and greater use of coronal images. In a follow-up investigation, Hsieh et al. [5] tested the effect of a training program on 31 readers using the same 40 CT cases, showing that search errors decreased significantly (10.8% to 8.1%, $p = 0.01$), but classification errors and overall diagnostic accuracy did not improve. Gains were most pronounced among trainees and non-abdominal radiologists, suggesting that targeted training can reduce oversight errors, whereas classification skill likely requires longer or more intensive experience. These are the three datasets that were synthesized together in this present study.

II. MATERIAL AND METHODS

Acquisition of the data used in this study was approved by our institutional review board.

A. Imaging protocol

The imaging protocol has been described in previous work [4, 7, 8]. Briefly, iodine-enhanced abdominal CT scans were performed using 128-section Siemens scanners (Definition, Definition Flash, or Definition AS+; Siemens Healthcare) at routine radiation dose levels of 200 quality reference milliampere-seconds (QRM). In the multidose study, projection-level noise insertion was used to synthesize reduced-dose 120 QRM reconstructions. Abdominal CT images from the 91 patients were obtained in the portal venous phase and reconstructed with 3 mm slice thickness at 2 mm intervals using filtered back projection and a medium-smooth (B30f) kernel. Ground truth for liver metastases was confirmed by histopathological analysis or radiologic evidence of progression. The 40 CT examinations with 91 metastases were selected from a larger group of 102 examinations to balance task difficulty and to fit within a 4-hour scheduling block. Selection was done by two gastrointestinal radiologists (27 and 7 years of experience, J.G.F. and A.I., respectively) who identified seven lesion attributes linked to detection difficulty—small size, low contrast, unclear margins, proximity to vessels or liver edges, and location in segments I/III. Cases exhibiting at least three such features were included in this study. Of the 40 CT examinations, 8 had no metastases. Many of our cases included benign lesions but these were not analyzed in this work.

B. Image analysis

In this study, we used a 500 Hz eye tracker with high accuracy (about 1°, equivalent to 15 mm in anatomy) to record gaze positions mapped onto DICOM coordinates. Readers were asked to mark all suspected metastases. A lesion was classified as detected when their marking overlapped with the ground-truth annotation. For each missed metastasis, the total gaze time was determined by summing all fixations within a 40-pixel radius and within ±2 image slices of the metastasis center point. If the total gaze time exceeded 0.5, 1 and 2 seconds respectively, the reader was coded to have made a decision error; otherwise, it was coded to be a search error.

III. RESULTS

Figure 2 shows misses for each lesion together with linear regression to predict the number of misses at 120 QRM versus 200 QRM. The intercept of the trend line was fixed at 0. In general, there are more misses at 120 QRM than 200 QRM, as expected when dose is reduced. However, this behavior was not observed for decision error-dominant lesions; the trend line for decision dominant lesions, across all gaze time thresholds was about 1. Pairwise comparisons of category-specific regression slopes were performed using t-tests. Significant differences between search and decision

dominant lesions were observed at gaze-time thresholds of 0.5 s ($p = 0.045$) and 1 s ($p = 0.050$), but not at 2 s ($p = 0.552$).

While the slope of the scatterplot is suggestive, a more direct way to test the hypothesis that search-dominant (but not decision-dominant) metastases are missed more frequently at lower radiation dose is using the Wilcoxon signed-rank test. Table 1 summarizes this analysis. As in Figure 2, we considered three gaze time thresholds: 0.5 s, 1 s, and 2 s. When considering all 91 lesions, the test shows a statistically significant increase in missed lesions at 120 QRM compared to 200 QRM ($p = 0.002$), with the mean number of misses increasing from 1.23 at 200 QRM to 1.79 at 120 QRM – a 45% increase in misses because of reduced radiation dose. However, when lesions are classified by error type, search-dominant errors (Category 0) consistently demonstrate a significant dose effect across all thresholds ($p \leq 0.002$), with substantially higher mean misses at 120 QRM than at 200 QRM. In contrast, decision-dominant errors (Category 1) do not show a statistically significant difference at any threshold ($p = 0.200$, 0.330, and 0.355 for 0.5 s, 1 s, and 2 s, respectively), although there is a trend towards more misses at low dose. These decision-dominant lesions make up between 27% to 46% of the total population depending on the threshold used – they are not exceptions, but common.

These findings indicate that even in liver metastasis detection, the clinical task that is sometimes seen as the "poster child" of low contrast lesion detection, there is a sizable group of lesions that do not suffer much degradation in lesion detection performance at reduced dose, at least for the 40% reduction used here. These lesions can be categorized using eye tracking gaze times. Our finding remains robust across different gaze time thresholds used to define search versus decision dominance.

Table 1: Wilcoxon signed test was conducted for 0.5, 1 and 2 sec respectively where Mean = mean number of missed metastases at 200 and 120 QRM, respectively.

Threshold	Category	p-value	n	Mean (200 QRM misses)	Mean (120 QRM misses)
0.5 sec	All	**0.002**	91	1.23	1.79
	Search-dominant	**0.001**	49	0.95	1.61
	Decision-dominant	0.200	42	1.54	2.00
1 sec	All	**0.002**	91	1.23	1.79
	Search-dominant	**0.001**	58	0.96	1.62
	Decision-dominant	0.330	33	1.75	2.09
2 sec	All	**0.002**	91	1.23	1.79
	Search-dominant	**0.002**	66	1.25	1.92
	Decision-dominant	0.355	25	1.16	1.41

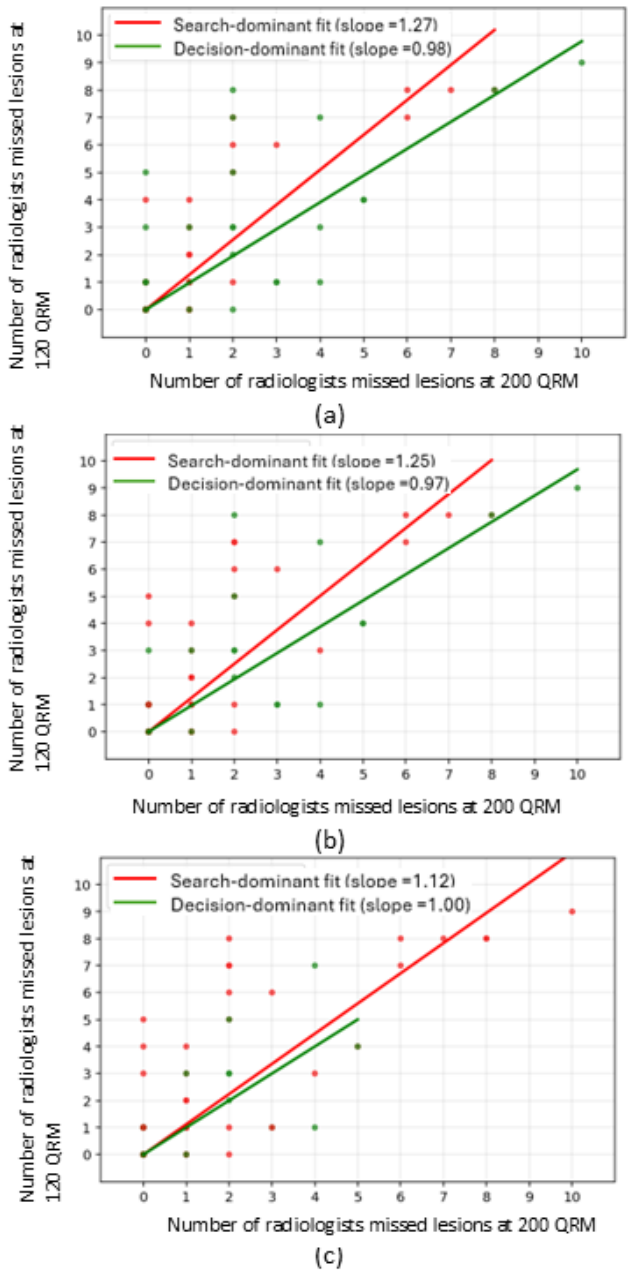


Figure 2. Plot of number of radiologists who missed the lesions at 200 QRM vs number of radiologists who missed lesions at 120 QRM. The points on the plot are color coded based on search dominant errors: red with decision dominant errors: green with threshold (a) 0.5 sec (b) 1 sec (c) 2 sec

IV. Conclusion

In this study, we demonstrated that missed liver metastases at lower radiation dose can be explained by separating errors into search versus decision failures using eye-tracking data. We observe that search-error-dominant lesions, which are characterized by low contrast and limited conspicuity, are significantly affected by dose reduction, whereas decision-error-dominant lesions are missed at similar rates across dose levels. This indicates that lower dose primarily impairs the visual search process rather than cognitive recognition. By linking eye-movement behavior with multi-dose detection outcomes, our study helps in understanding why radiologist performance declines at lower dose.

References

[1] A. H. Goenka, B. R. Herts, N. A. Obuchowski *et al.*, "Effect of reduced radiation exposure and iterative reconstruction on detection of low-contrast low-attenuation lesions in an anthropomorphic liver phantom: an 18-reader study," Radiology, 272(1), 154-63 (2014).

[2] S. S. Hsieh, D. R. Holmes Iii, R. E. Carter *et al.*, "Peripheral liver metastases are more frequently missed than central metastases in contrast-enhanced CT: insights from a 25-reader performance study," Abdom Radiol (NY), 50(2), 668-676 (2025).

[3] T. T. Brunyé, T. Drew, D. L. Weaver *et al.*, "A review of eye tracking for understanding and improving diagnostic interpretation," Cogn Res Princ Implic, 4(1), 7 (2019).

[4] S. S. Hsieh, D. A. Cook, A. Inoue *et al.*, "Understanding Reader Variability: A 25-Radiologist Study on Liver Metastasis Detection at CT," Radiology, 306(2), e220266 (2023).

[5] S. S. Hsieh, A. Inoue, M. Yalon *et al.*, "Targeted Training Reduces Search Errors but Not Classification Errors for Hepatic Metastasis Detection at Contrast-Enhanced CT," Acad Radiol, 31(2), 448-456 (2024).

[6] S. S. Hsieh, A. Inoue, M. Yalon *et al.*, "A training program to reduce reader search errors for liver metastasis detection in CT," Proc SPIE Int Soc Opt Eng, 12467, (2023).

[7] H. Gong, S. S. Hsieh, D. R. Holmes, 3rd *et al.*, "An interactive eye-tracking system for measuring radiologists' visual fixations in volumetric CT images: Implementation and initial eye-tracking accuracy validation," Med Phys, 48(11), 6710-6723 (2021).

[8] P. S. Pillai, D. R. Holmes, 3rd, R. Carter *et al.*, "Individualized and generalized models for predicting observer performance on liver metastasis detection using CT," J Med Imaging (Bellingham), 9(5), 055501 (2022).

A Theoretical Prediction of CT-based Radiomics Feature Variance to Improve the Performance of Classification Models

Huay Din, Yijie Yuan, Grace Hyun Kim, Michael McNitt-Gray, J. Webster Stayman, and Grace J. Gang

***Abstract*—Radiomics involves the high-throughput extraction of quantitative image biomarkers for analysis in machine learning models. While this field holds immense potential for personalized medicine, the clinical translation of radiomic models is severely hampered by their intrinsic dependency on image acquisition properties, degrading downstream predictive performance. In this work, we propose a statistical framework to explicitly account for stochastic-based feature uncertainty. We derive an analytic variance estimator through a novel use of noise power spectrum (NPS) information to linearly propagate image noise to radiomic features. The accuracy of this estimator was validated against Monte Carlo simulations. The predicted and measured feature variance were compared through z-score normalized percentage error with the predicted variance achieving a prediction error of $<$20% across the majority of histogram and gray level co-occurrence features. We subsequently integrated these variance estimates to an uncertainty aware support vector machine (TSVC [1]) and compare the results to the standard support vector machine model. We simulated a simple classification problem with known per-sample variance estimates and show that TSVC has the capacity to either perform just as well, or better than the standard SVM. These results demonstrate that 1) variance estimates of features can be reliably obtained and 2) illustrate the potential utility of incorporating variance information to enhance model performance.**

I. Introduction

Radiomics is a technique for extracting quantitative image biomarkers from medical images [2], [3]. These radiomic features can describe the shape, statistical distributions, and higher-order textural patterns of an image. The aim of radiomics is to identify a radiomic signature within the set of extracted radiomic features to aid in clinical decision support systems. Computed Tomography (CT), owing to its widespread clinical use, is a key modality in radiomics research. However, in practice, radiomics features are intrinsically coupled to the quality of the source image. In this work, we highlight the impact of image noise.

As medical images are stochastic, so are the measured radiomic features. That is, a radiomics feature is not just a deterministic value, but a sample drawn from a distribution. Furthermore, the spatial resolution and noise characteristics of medical images are also coupled such that sharper images tend to have higher levels of noise, and smoother images tend to have lower levels of noise. Subsequently, features extracted from noisier data will have higher stochastic variability and vice versa.

In practice, highly variable features are often left out of radiomics analysis. Furthermore, despite the irremovable presence of noise in medical images, all uncertainty information is generally ignored during the construction and training of the model such that all samples and radiomic features are treated as true point values. We hypothesize that a model informed by feature variance will allow researches to capture all possible information, even for features typically discarded. We furthermore propose that a variance-aware radiomic model will perform just as well or better than standard variance-unaware models.

In this work, we employ differentiable definitions of radiomics features and prior knowledge of CT image noise to theoretically predict radiomic feature variance using first order approximations. We then investigate the performance benefit of incorporating variance in a Support Vector Machine for a binary classification task.

II. Methods

A. Measurement Model

We focus our investigation on systems that are linear and shift-invariant. This assumption holds for local regions of interest in filtered backprojection (FBP) or locally linearizable model-based reconstructions. [4] We model the measured image data, $\tilde{\mu}$, as a perturbation of the ground truth image μ given by

$$\tilde{\mu} = b_1 * \mu + b_2 * n \tag{1}$$

where b_1 is the system blur, n is independent Gaussian noise, and b_2 is a kernel that imparts correlation. In practice, the kernels b_1 and b_2 can be empirically measured or predicted using analytical image quality models. In this above model, b_1 can be determined for each reconstruction kernel and b_2 can be estimated for each local ROI.

B. Variance Prediction of radiomics features

In this work, we theoretically predicted the variance of radiomics features using gradients derived from a differentiable implementation of radiomics computation and prior knowledge of image noise. With these two pieces of information, variance of radiomics features can be easily obtained using first order approximation and error propagation. In this work, we focus on radiomics features derived from the histogram and gray-level co-occurrence matrix (GLCM). Conventional definitions

Huay Din is in the Department of Bioengineering at the University of Pennsylvania. Grace J. Gang is in the Department of Radiology at the University of Pennsylvania. Yijie Yuan and J. Webster Stayman are in the Biomedical Engineering at Johns Hopkins University. Grace Hyun Kim and Michael McNitt-Gray are in the Department of Radiology at the University of California Los Angeles, e-mail: grace.j.gang@pennmedicine.upenn.edu

of histogram and GLCM are not differentiable due to the typical practice of discretization of images into discrete gray levels and counting of all voxels at a certain gray level (or in the case of GLCM, voxel pairs).

To address these issues, we approximated the histogram and GLCM using kernel density estimation. Specifically, we used a Gaussian kernel with the largest $\sigma = 0.7$ such that the majority of radiomic features are within 5% error to feature values computed using Pyradiomics with sufficiently smooth gradients for our variance prediction method. All operations were implemented in PyTorch. Denoting radiomics feature computation as $R = f(\mu)$, the Jacobian of f, $\mathbf{J}$, can be obtained via the automatic differentiation function of PyTorch.

The covariance of the image $\tilde{\mu}$ is given by $\mathbf{B}^T\Sigma_n\mathbf{B}$, where $\mathbf{B}$ is a Toeplitz matrix representing convolution with b_2. Using a first order approximation, the variance of radiomics feature R is given by:

$$\text{Var}[R] = \mathbf{J}^T\mathbf{B}\Sigma_n\mathbf{B}^T\mathbf{J} \tag{2}$$

When n is independent and stationary, the above equation can be simplified to:

$$\text{Var}[R] = \sigma^2\mathbf{J}\mathbf{B}\mathbf{B}^T\mathbf{J}^T \tag{3}$$

C. Experimental Dataset

To validate the proposed framework, we utilized 3D printed phantoms emulating trabecular bone textures developed in previous work [5]. The phantoms were generated using a procedural method wherein spherical voids are packed within a cylindrical volume of 25mm diameter and 8mm height. Based on the size and overlap of these voids, distinct texture classes can be generated. For this study, we utilized two texture classes characterized by void diameters 1.4mm and 1.6mm as illustrated in Fig. 1. The phantoms were fabricated on a commercial stereolithography (SLA) printer (Peopoly Moai, Hong Kong) with a 70 μm laser spot and 25 μm layer height. The phantoms were scanned on a microCT scanner (Bruker SkyScan 1172, Billerica, MA) at 50 kVp and 200 μA. Images were reconstructed using the FBP algorithm at an isotropic voxel size of 27 μm. All voxel values were than scaled to the range of linear attenuation coefficients from 0 to 0.02 mm^{-1}. These scans provides the ground truth texture for all subsequent experiments.

We simulated CT noise using a cascaded system analysis model [6]. Different noise magnitudes and correlations were simulated by varying the tube current and pulse width products, mAs, and the cutoff frequency c_0 in the Hann apodization filter in FBP. We derived b and the noise magnitude directly from the model and applied them to the ground truth microCT image according to eq. 1 to simulate CT reconstructions. For initial investigations in this work, only correlated noise was added and b_1 was set to be an impulse image.

D. Validation of the Variance Prediction Method

We first evaluated the accuracy of the analytic variance prediction against empirically obtained estimates. We computed empirical feature variance using 1000 noise realizations at different dose levels. We extracted an example ROI from the 1.4mm phantom and created six noisy datasets. We applied correlated noise using the kernel associated with $c_0 = 1, mAs = 1$) and subsequently scaled the noise magnitude to 0.0015, 0.0047, 0.0150, 0.0168, 0.0237, and 0.0375. As each dose level, we simulated $N = 1,000$ noise realizations and computed the empirical variance of radiomics features. The percentage differences between the predicted and empirical variance were computed for comparison.

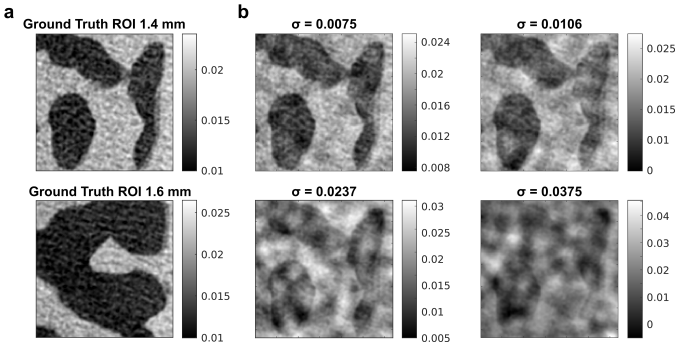


Fig. 1: (a) Example ROIs corresponding to the 1.4mm and 1.6mm texture classes are shown. (b) The 1.4mm ROI is shown with correlated noise ($c_0 = 1, mAs = 1$) scaled to four different noise magnitudes (σ). The displayed range bounds the noise levels utilized in the classification models

E. Application Towards Classification Models

1) Classification Models: To demonstrate the clinical utility of the proposed method, we emulated a classification task distinguishing between two idealized texture classes (1.4, and 1.6 mm voids). We hypothesized that incorporating feature variance would improve classification performance. In this work, we compare the performance of the standard support vector machine (SVM) with the uncertainty-aware SVM model proposed by Bi and Zhang [1] (TSVC).

The SVM is a supervised machine learning method that performs binary classification through finding the hyperplane that maximizes its distance between the two classes. To perform nonlinear classification, the data is transformed into a higher dimension, through the kernel trick, in which a hyperplane may exist to separate between the two classes.

Any hyperplane can be written as the set of points x that satisfies $w \cdot x - b = 0$, where w is the normal vector to the hyperplane, and b is some offset. The hyperplane that maximizes its distance between the two classes can be solved through solving the following optimization problem:

$$f(w^*) = \min_{w,b,\xi} \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{\ell}\xi_i \tag{4}$$

with the condition that $y_i(w \cdot x_i + b) \geq 1 - \xi_i, \forall i \in \{1, \ldots, \ell\}$ and ℓ is the number of training samples. Variable $\xi_i = \max\{0, 1 - y_i(w \cdot x_i + b)\}$ is introduced as a slack variable and effectively measures the degree to which an individual sample x_i can violate the "margin" ($w \cdot x + b = \pm 1$). Importantly, individual samples are classified in one of two

classes depending on whether $w \cdot x_i + b$ is greater than or less than 0. Lastly, parameter C is a regularization term that controls how strictly we would like to minimize the number and severity of misclassifications and competes against the goal of maximizing the margin $\frac{1}{2}||w||_2^2$.

The Total Support Vector Classification (TSVC) [1] solves the same optimization problem, but minimizes for w, b, ξ and Δx and has the condition that $y_i(w \cdot (x_i + \Delta x_i) + b) \geq 1 - \xi_i, ||\Delta x_i|| \leq \delta_i \forall i \in \{1, \ldots, n\}$. This condition allows TSVC to shift each training point inside a ball of radius δ_i before the final fit. While the TSVC can be recast to a second-order cone program (SOCP) and solved with a typical SOCP solver, the authors propose a two step iterative approach. Before iterating, Δx_i is set to 0. The first step is to fix $\Delta x_i, i = 1, \ldots, \ell$ to the current value, and solve 4 with the conditions for the TSVC model. Parameters w, b, and ξ are obtained from this solution. The next step is to then fix w, b to the values obtained in step 1, and then solve 4 with the conditions for the TSVC model for $\Delta x_i, i = 1, \ldots, \ell$, and ξ. These two steps are repeated until a termination criterion is met, such as the values of Δx_i sufficiently converge.

To solve for Δx_i in the second step, Bi and Zhang proved that for any given hyperplane (w, b), the solution $\Delta \hat{x}_i$ of the SVM with the conditions for the TSVC model is $\Delta \hat{x}_i = y_i \delta_i \frac{w}{||w||}, i = 1, \ldots, \ell$. In our framework with anisotropic uncertainty, we solved for w rotated by $\Sigma_i = \text{diag}\left(\text{Var}[x_{1,i}], \text{Var}[x_{2,i}]\right)$

$$\Delta \hat{x}_i = y_i \lambda \delta_i \frac{\Sigma_i w}{||\Sigma_i w||} \tag{5}$$

where $\text{Var}[x_{:,i}]$ is the predicted variance of sample x_i with respect to the first and second radiomic features. Given that the absolute values of $\text{Var}[x_{1,i}]$ and $\text{Var}[x_{2,i}]$ may dramatically differ, we have put in a second global scalar, λ, which controls the size of the uncertainty ellipsoid around each sample. When $\lambda = 0$, we solve for the standard SVM, and when $\lambda = 1$, Δx_i may shift by one standard deviation in each axis.

All calls to solve the SVM were implemented using the `scikit-learn` library in Python.

2) Dataset and Classification Problem: To evaluate the potential benefit towards incorporating variance information into the modeling procedure, we perform a simple experiment inspired by the synthetic dataset described by Bi and Zhang [1].

We considered two classification problems with the defined true decision boundaries, $x_1 - x_2 = 0$ and $x_1^2 + x_2^2 = 9$. If ground truth data was known, all data points above the linear boundary and inside the quadratic boundary would correspond to data belonging to class 1. Similarly, all data points below the linear boundary and outside the quadratic boundary would correspond to data belonging to class 2. Here, x_1 can represent any radiomic feature, such as Entropy, and x_2 can represent another radiomic feature, such as Correlation.

We created five datasets, with $\ell = 20, 50, 150$ training samples x_i from the uniform distribution sampled from $[-5, 5]^2$. All samples were contaminated by adding Gaussian noise with mean $[0, 0]$ and covariance matrix $[\sigma_1 e_1, \sigma_2 e_2]$, where $e_{1,2}$ are the first two canonical basis of $\mathbb{R}^2$. All $\sigma_{1,2}$ values were

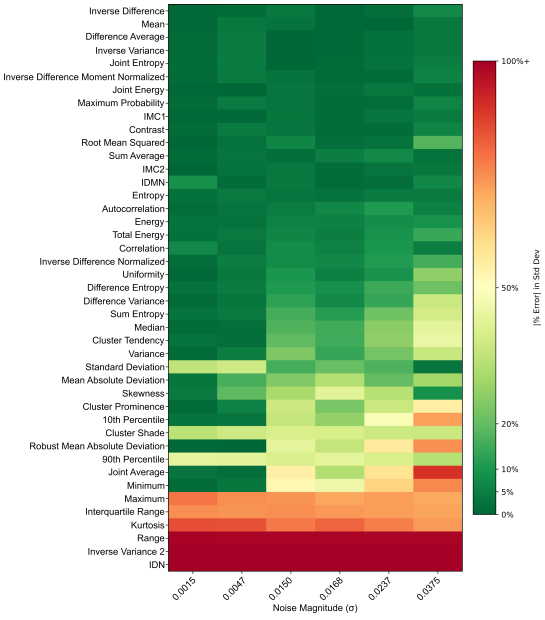


Fig. 2: Absolute percentage difference of analytical feature standard deviation estimates compared to empirical ground truth. Results are derived from one ground truth ROI, 1,000 noisy realizations at six distinct noise levels.

randomly drawn from $[0.1, 0.8]$ with the following exception. To produce an outlier effect, we first ordered all samples by ascending distance to the true decision boundary. Within the first 0.2ℓ samples of each dataset, 0.1ℓ was chosen such that $\sigma_{1,2}$ are randomly drawn from the larger range of $[0.5, 2]$.

For the first problem, models were obtained by the SVM (SVM-L) and TSVC with a linear kernel and for the second problem, models were obtained from the SVM (SVM-Q) and TSVC with a quadratic kernel $(x_i^T x_j)^2$.

The testing set for all models was constructed to be 10,000 samples uniformly sampled on $[-5, 5]^2$ and without contamination. The above experiment and evaluation was performed 50 times. The accuracy over the 50 trials are reported in Table 1.

To ensure fair comparison, we performed a hyperparameter search to identify the optimal configuration for each model. For each trial, ℓ, and problem, a validation set of 200 noisy samples was generated with $\sigma_{1,2} \in [0.1, 0.8]$. For the SVM, only parameter C was optimized over the validation training set for the highest accuracy, and for TSVC, parameters C and λ were jointly optimized over the validation training set for the highest accuracy. The regularization parameter C was optimized over a log10 array from 10^{-2} to 10^2 and parameter λ was optimized over $\{0.1, 0.3, 0.5, 0.7, 1, 1.5, 2\}$.

3) Statistical Analysis: To compare the performance differences between the baseline SVM and TSVC, we utilized a paired Student's t-test as the models were evaluated on identical cross-validation folds. We tested the null hypothesis that the mean difference in classification accuracy performance between the paired models is zero. A *p*-value of less than 0.05 was considered statistically significant for the model comparisons.

TABLE I: Model Performance Comparisons

Model Setup	**Accuracy** (mean $\pm$ std)	**p-value**
$\ell = 20$		
SVM-L	0.9440 ± 0.0258	0.0017
TSVC-L	0.9538 ± 0.0246	
SVM-Quad	0.9198 ± 0.0392	0.0020
TSVC-Quad	0.9312 ± 0.0332	
$\ell = 50$		
SVM-L	0.9504 ± 0.0248	0.0011
TSVC-L	0.9630 ± 0.0227	
SVM-Quad	0.9499 ± 0.0208	0.112
TSVC-Quad	0.9546 ± 0.0210	
$\ell = 150$		
SVM-L	0.9729 ± 0.0144	0.75
TSVC-L	0.9735 ± 0.0137	
SVM-Quad	0.9662 ± 0.0160	0.545
TSVC-Quad	0.9676 ± 0.0153	

III. Results

A. Analytic vs. Empirical Variance Prediction Validation

We validated the accuracy of the proposed variance prediction against empirical estimates computed from 1,000 noisy realizations at six different noise magnitudes that range from 0.0015 to 0.0375. Broadly, results indicate that prediction accuracy depends primarily on the mathematical definition of radiomic features with noise magnitude also playing a role in the accuracy of the variance estimation framework. Fig. 2 displays the absolute percentage difference between predicted and measured feature standard deviations for 43 radiomic features. Features are ordered by percentage difference, with the most accurate at the top and least accurate at the bottom.

From Fig. 2, we can observe that the linear error propagation method is highly effective for features with linear or second polynomial orders. This is further visualized in Figure 4. In this figure, the 1,000 measurements of Entropy and Joint Energy are shown for three of the six different noise magnitudes. The solid red cross indicates the empirical mean value of the two radiomic features. The light gray markers corresponds to a single noisy radiomic measurement. The dashed black line and solid red line corresponds to the empirical and analytical standard deviation estimates corresponding to the Entropy and Joint Energy features. While the x and y limits of the three different noise levels/subplots differ, it can be seen that there is high overlap between the empirical and analytical quantities. It can also be seen that the incidence of the two curves decreases as the noise magnitude increases. Notably, even as the noise magnitude increases a factor of 27, the percentage difference between the two features' empirical and analytical estimates are still within 20% of each other.

Continuing in Fig. 2, it can be seen that the effect of noise is more prominent for features that are higher in polynomial order, like Skewness, Kurtosis, and probabilistic features like 90th Percentile. There are also radiomic features that are consistently inaccurate, $> 50\%$ error, such as Inverse Difference Normalized (IDN), Range, Minimum, and Maximum. This feature group suggests that the analytical approximation is less accurate for features with probabilistic definitions or rely on single-values in the ROI. Overall, the majority of histogram and GLCM radiomic features achieved a percentage difference of less than 20%, suggesting that the analytical method, while being a first-order approximation, provides a simple and accurate path towards estimating noise-based feature variance.

B. Classification Performance

The classification performance of the baseline SVM versus the proposed variance-aware TSVC aggregated across 50 trials is summarized in Table I. It can be seen that at $\ell = 20$ and in both the linear and quadratic problems, the average accuracy of the TSVC models outperformed the SVM models, with the p-values showing statistical significance. When the training size more than double its size, $\ell = 50$, both TSVC models in the linear and quadratic problems outperformed its SVM counterparts with the only the findings with the linear problem being statistical significant. With the largest training set size, $\ell = 150$, the performance of the two models are effectively and statistically comparable.

These aggregate results can be visually understood through Fig. 4. In Fig. 4, we present examples for each problem type (cols) and ℓ (rows). These results were chosen to reflect cases where the TSVC models far outperformed the SVM models. The training data are plotted in red and blue markers to correspond with their class label. The per-sample $\sigma_1, 2$ are drawn as shaded ellipses. The true decision boundary is marked with a dashed black line. The SVM decision boundaries are drawn as dashed blue lines, and the TSVC decision boundaries are drawn as solid red lines. For a small training set, $\ell = 20$, we can see that the hyperplane may be highly variable due to sparse sampling of the training data. The uncertainty information here helps to reign in how "far the decision boundary wanders". For a slightly larger training set, $\ell = 50$, there are more points for both models to consider. While the TSVC models demonstrate a higher accuracy, the SVM decision boundaries are also closer to the true decision boundary. In the largest training set, $\ell = 150$, we can see that the space is well sampled, and the performance of both models are more comparable. All accuracies reported in Fig. 4 are obtained through evaluating the models on a separate testing set of 10,000 samples with no added noise.

The findings from Table I and visualization from Fig. 4 suggest that the performance of the variance-aware TSVC is equivalent to or outperforms the standard SVM for this classification task. Furthermore, we find that the variance-aware model may be of greater benefit in settings with a small training set size.

IV. Discussion and Conclusion

In this work, we demonstrated the ability to theoretically predict the variance of histogram and GLCM radiomic features. Furthermore, we implemented a variance-aware ra-

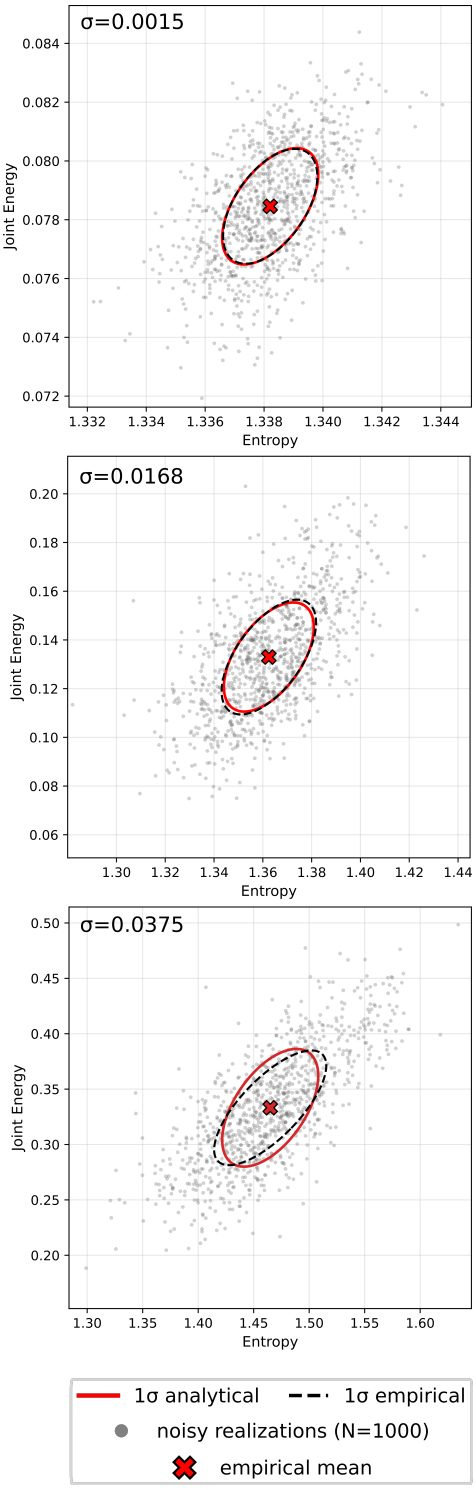


Fig. 3: Entropy and Joint Energy measurements from the 1,000 noisy instances at three noise magnitudes are plotted (gray markers). The analytical (red-solid) and empirical (black-dashed) standard deviation estimates are over-layed as ellipsoids centered around the empirical mean estimate of Entropy and Joint Energy.

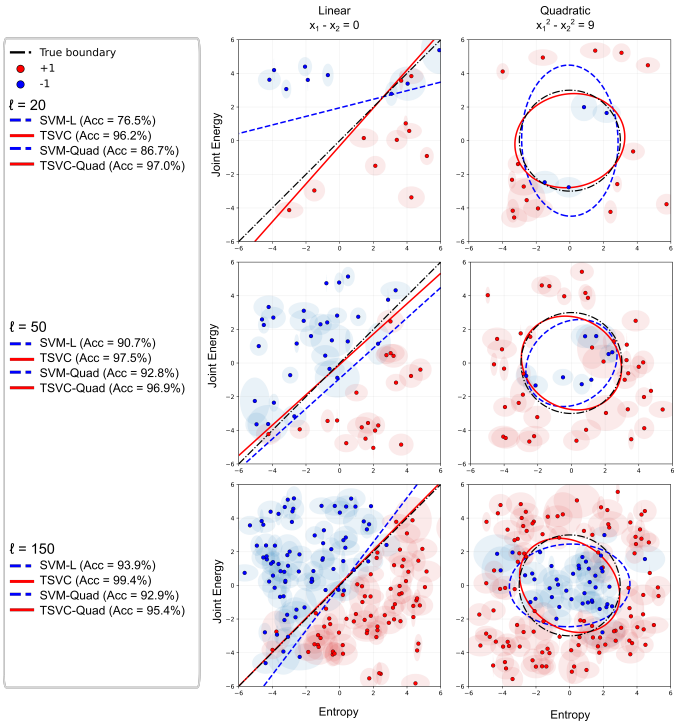


Fig. 4: For each problem (linear and quadratic) and at each training set size ($\ell = 20, 50, 150$), we show an example set of training data, the true decision boundary, and the decision boundaries of the SVM and TSVC models. In a future experiment, the training samples would not be uniformly drawn from the $[-5, 5]^2$ grid, but would be of radiomic feature measurements. For example, Entropy and Joint Energy. Accuracy measurements were obtained from a separate testing set of 10,000 samples.

diomics model that demonstrated the potential of incorporating variance information into the training of classification models. We anticipate that variance information will be more useful with smaller training set sizes. Future work includes applying this framework to a realistic clinical task, evaluating the method across a broader range of reconstruction and acquisition settings, and framing the theoretical model for variance prediction towards more complex and higher-order texture features.

Acknowledgment

This work is supported, in part, by NIH Grant R01EB031592.

References

[1] Bi, J. and Zhang, T., "Support vector classification with input data uncertainty," *Advances in neural information processing systems* **17** (2004).

[2] Lambin, P., Rios-Velazquez, E., Leijenaar, R., Carvalho, S., van Stiphout, R. G., Granton, P., Zegers, C. M., Gillies, R., Boellard, R., Dekker, A., and Aerts, H. J., "Radiomics: Extracting more information from medical images using advanced feature analysis," *European Journal of Cancer* **48**(4), 441–446 (2012).

[3] Kumar, V., Gu, Y., Basu, S., Berglund, A., Eschrich, S. A., Schabath, M. B., Forster, K., Aerts, H. J., Dekker, A., Fenstermacher, D., et al., "Radiomics: the process and the challenges," *Magnetic resonance imaging* **30**(9), 1234–1248 (2012).

[4] Gang, G. J., Stayman, J. W., Zbijewski, W., and Siewerdsen, J. H., "Task-based detectability in ct image reconstruction by filtered backprojection and penalized likelihood estimation," *Medical physics* **41**(8Part1), 081902 (2014).
[5] Shi, H., Gang, G. J., Li, J., Liapi, E., Abbey, C., and Stayman, J. W., "Performance Assessment of Texture Reproduction in High-Resolution CT," *Proceedings of SPIE–the International Society for Optical Engineering* **11316**, 113160R (Feb. 2020).
[6] Gang, G. J., Deshpande, R., and Stayman, J. W., "Standardization of histogram-and gray-level co-occurrence matrices-based radiomics in the presence of blur and noise," *Physics in Medicine & Biology* **66**(7), 074004 (2021).

Focal Spot Design Optimization for High-Resolution CT

Yue Fan, Xiao Jiang, Grace J. Gang, and J. Webster Stayman

***Abstract*—Recent advances in improved detector resolution including photon counting CT has helped to enable high-resolution diagnostic computed tomography (CT). However, the finite size of the X-ray tube focal spot has emerged as a dominant limitation on spatial resolution. While reducing focal spot size improves resolution, it also limits photon fluence. This can result in unacceptable levels of noise, particularly for larger patients. Novel x-ray tube designs may be required to overcome these fluence limitations. Previous work has shown that using structured focal spots, multiple x-ray sources, and advanced reconstruction models has the potential to overcome classic noise-resolution trade-offs. In this work, we write down general expressions for the imaging performance of a multi-focal spot CT system and use those predictors to design structured focal spots based on specific imaging tasks. Specifically, an analytic performance model is developed that explicitly links extended focal spot shape to the system modulation transfer function (MTF) and noise power spectrum (NPS) through local Fourier–based approximations. This model enables efficient prediction of a task-weighted detective quantum efficiency (DQE) and supports gradient-based optimization of focal spot emission patterns under physical fluence constraints. Focusing on high-frequency imaging tasks, we formulate the design problem as minimizing NPS within a prescribed spatial-frequency band and solve it using first-order optimization with multiple random initializations. Simulation studies demonstrate that the optimized focal spot designs substantially suppress high-frequency noise compared to a conventional small focal spot (limited fluence) configuration, yielding marked improvements in task-specific DQE. Analytical NPS predictions closely match empirical estimates from noisy reconstructions, validating the accuracy of the proposed model. These results highlight focal spot shape as a powerful and previously underutilized degree of freedom in CT system design and establish task-based, gradient-driven focal spot optimization as a promising approach for enhancing high-resolution CT performance.**

I. Introduction

HIGH–resolution computed tomography (CT) has enabled a wide range of advanced medical imaging applications, including assessment of trabecular bone microarchitecture, visualization of fine structures in the inner ear, depiction of small bronchioles and blood vessels, and characterization of subtle organ textures. Historically, improvements in spatial resolution have been driven primarily by advances in detector technology, such as reduced pixel sizes in both energy-integrating and photon-counting CT systems[1]. However, as detector resolution continues to improve, the X-ray source has increasingly become the dominant factor limiting system spatial resolution.

In particular, the finite size of the tube focal spot imposes a fundamental tradeoff between spatial resolution and photon fluence. Reducing focal spot size generally improves spatial resolution, but simultaneously limits the achievable X-ray output, leading to increased quantum noise—especially in large patients or highly attenuating anatomical regions. Although longer exposure times can partially compensate for reduced fluence, such approaches are constrained by patient motion and clinically acceptable scan times.

Several alternative X-ray source technologies have been proposed to overcome focal spot limitations, including liquid-metal jet sources, synchrotron radiation, and micro-structured x-ray targets. Although promising, these approaches are often difficult to integrate into conventional clinical CT systems or remain technologically immature. More recently, multi-focal-spot strategies—such as combining small and large focal spots within a single acquisition—have been shown to provide a favorable balance between spatial resolution and noise performance[2]. Previous studies have also suggested that the shape of the focal spot itself can be customized to improve imaging performance[3], motivating a shift from fixed hardware design to performance-driven focal spot optimization. The authors have previously proposed focal spot design[4] based on idealized approximations to the CT forward model. However, those methods do not account for sampling and make additional approximations on object uniformity, etc.

In this work, we leverage previous approaches based on local Fourier approximation to develop an analytic performance model for CT systems employing multiple focal spots and use it to formulate a gradient-based, task-driven optimization framework for focal spot design. By explicitly modeling the impact of focal spot shape on the system modulation transfer function (MTF) and noise power spectrum (NPS), the proposed framework enables efficient gradient-based optimization of focal spot shapes to enhance detective quantum efficiency (DQE) in the high-frequency regime. The framework is validated through simulation studies, demonstrating that optimized focal spot designs achieve significantly improved DQE within the desired frequency band compared to conventional focal spot configurations.

II. Methods

A. CT Forward Model with Dual Extended Focal Spots

Consider a CT system with two X-ray sources, each has an extended focal spot that is discretized into multiple sourcelets. Under the Gaussian noise assumption, the measurement model is given by:

$$\mathbf{y} \sim \mathcal{N}(\bar{\mathbf{y}}, \mathbf{K}), \quad \bar{\mathbf{y}} = \mathbf{G}\exp(-\mathbf{A}\mathbf{x}) \tag{1}$$

where

$$\bar{\mathbf{y}} = \begin{bmatrix} \bar{\mathbf{y}}_1 \\ \bar{\mathbf{y}}_2 \end{bmatrix}, \quad \mathbf{K} = \mathbf{D}\left\{ \begin{bmatrix} \bar{\mathbf{y}}_1 \\ \bar{\mathbf{y}}_2 \end{bmatrix} \right\}$$

Yue Fan, Xiao Jiang, and J. Webster Stayman are with the Department of Biomedical Engineering, Johns Hopkins University, e-mail: {yfan59, xjiang43, web.stayman}@jhu.edu.

Grace J. Gang is with Department of Radiology, University of Pennsylvania, e-mail: grace.j.gang@pennmedicine.upenn.edu.

denotes the total measurement vector and diagonal covariance matrix, respectively. Each of these is written in terms of the mean measurements from each X-ray source, $\bar{y}_1$ and $\bar{y}_2$. The operator $\mathbf{G}$ is responsible for the integration over all sourcelets associated with each X-ray source and can be written as

$$\begin{aligned}\mathbf{G}(\mathbf{s}) &= \mathbf{s} \otimes \mathbf{I} \\ &= \begin{bmatrix} s_{11}\mathbf{I} & s_{12}\mathbf{I} & \cdots & s_{1N}\mathbf{I} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & s_{21}\mathbf{I} & s_{22}\mathbf{I} & \cdots & s_{2N}\mathbf{I} \end{bmatrix}\end{aligned}$$

where s_{ij} denotes the intensity of the j^{th} sourcelet in the i^{th} source. We define an "extra-large" system matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11}^T & \mathbf{A}_{12}^T & \cdots & \mathbf{A}_{1N}^T & \mathbf{A}_{21}^T & \mathbf{A}_{22}^T & \cdots & \mathbf{A}_{2N}^T \end{bmatrix}^T$$

where $\mathbf{A}_{mn}$ represents the forward projection operator corresponding to the n-th sourcelet of the m-th focal spot. In this work, we concentrate on a 2D CT system for preliminary investigation, but 3D extensions should be straightforward.

B. Model-based Reconstruction and Property Prediction

Model-based reconstruction seeks to estimate $\mathbf{x}$ from $\mathbf{y}$ by minimizing a regularized likelihood objective function:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}) = \|\bar{\mathbf{y}} - \mathbf{y}\|^2_{\mathbf{K}^{-1}} + \beta \mathbf{R}(\mathbf{x}), \tag{2}$$

where $\mathbf{R}$ represents a quadratic regularization term. Following previous work [5], [6], [7] in the approximation local reconstructed image properties, we have derived an expression for the local modulation transfer function (MTF) and local noise power spectrum (NPS). Both of these can be written in terms of the Fisher information matrix $\mathbf{F}$ for the data likelihood, which is

$$\mathbf{F} = \mathbf{A^T D G^T K G D A}, \ \mathbf{D} = D\{\exp(-\mathbf{A}\mathbf{x})\} \tag{3}$$

The MTF and NPS for model-based reconstruction of the form in (2) can be approximated as:

$$\begin{aligned}\mathrm{MTF}_j &= \mathrm{DFT}_{2D}\left\{ \left[\mathbf{F} + \mathbf{R}\right]^{-1} \mathbf{F}\mathbf{e}_j \right\} \\ &\approx \frac{\mathrm{DFT}_{2D}\{\mathbf{F}\mathbf{e}_j\}}{\mathrm{DFT}_{2D}\{\mathbf{F}\mathbf{e}_j + \mathbf{R}\mathbf{e}_j\}}\end{aligned} \tag{4}$$

$$\begin{aligned}\mathrm{NPS}_j &= \mathrm{DFT}_{2D}\left\{ \left[\mathbf{F} + \mathbf{R}\right]^{-1} \mathbf{F} \left[\mathbf{F} + \mathbf{R}\right]^{-1} \mathbf{e}_j \right\} \\ &\approx \frac{\mathrm{DFT}_{2D}\{\mathbf{F}\mathbf{e}_j\}}{\mathrm{DFT}_{2D}\{\mathbf{F}\mathbf{e}_j + \mathbf{R}\mathbf{e}_j\}^2}.\end{aligned} \tag{5}$$

where $\mathbf{e}_j$ represents a delta function at the location j on the object. $\mathrm{DFT}_{2D}\{\}$ represents the discrete 2D Fourier transform.

C. Task-based Multi-Focal-Spot Design Optimization

Task-based optimization aims to design the focal spot pattern $\mathbf{G}(s)$ to maximize detective quantum efficiency (DQE) weighted by a task function $T(f)$:

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} \int_{f_1}^{f_2} \frac{\mathrm{MTF}^2(f; \mathbf{G}(\mathbf{s}))\, T^2(f)}{\mathrm{NPS}(f; \mathbf{G}(\mathbf{s}))}\, \mathrm{d}f, \tag{6}$$

In this work, our objective was to increase the detective quantum efficiency (DQE) within a high-frequency band from f_1 to f_2. The task function $T(f)$ was therefore defined as a binary mask in this target frequency range. Although different regularization strengths can lead to substantially different DQE characteristics, in this proof-of-concept study, we restricted our attention to lightly regularized reconstruction, for which the modulation transfer function (MTF) remains close to unity (presuming invertibility). Under this assumption, maximizing the DQE was therefore equivalent to maximizing the inverse of NPS within the target frequency band from f_1 to f_2:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \sum_{f}^{f_1 \le f \le f_2} \mathrm{NPS}(f; \mathbf{G}(\mathbf{s}))T(f), \tag{7}$$

In addition, to ensure physically realizable designs, we set constraints on the focal spot emission, enforcing nonnegativity and with an upper bound on the maximum fluence (maximum number of photons/sourcelet). The optimization was solved via the Adam optimizer. To mitigate the risk of convergence to local minimums as encountered in previous experiments, we repeat the optimization multiple times using random initializations and different learning rates.

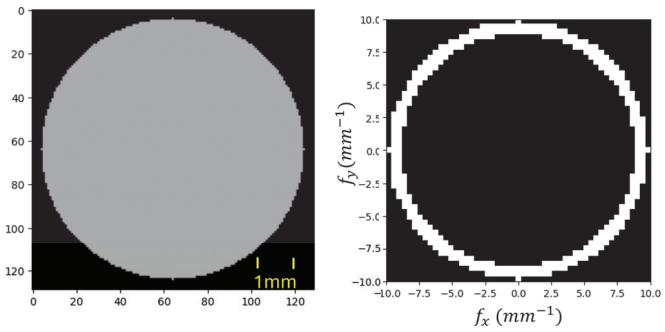


Fig. 1. (a) Disc phantom (Radius is 3 mm); (b) Frequency domain mask

D. Evaluation

We performed design optimization and validated its effectiveness on a simulated CT system. The source-detector distance (SDD) was set to 1600 mm, and the source-axis distance (SAD) was 800 mm, resulting in a magnification factor of 2 onto a detector with 301, 0.1 mm pixels. To ensure good sampling, we also shift the detector 0.025 mm. To partially model discretization effects, the projection simulations were conducted using an upsampled grid, with a voxel size of 0.01 mm and a detector pixel size of 0.025 mm. The projections were then downsampled via 4×4 binning to match the pixel sizes used in the reconstruction system.

The image volume was represented using 0.05 mm voxels (based on system magnification). All reconstructions were performed using a disc phantom (Fig. 1(a)) with a size of 129×129 voxels.

In the focal spot design, we selected the central voxel to evaluate $\mathbf{F}$, and targeted the spatial-frequency band from $f_1 = 9\ \mathrm{mm}^{-1}$ to $f_2 = 10\ \mathrm{mm}^{-1}$. This was achieved by applying a ring-shaped frequency domain mask, illustrated in Fig. 1(b), which restricted the computation of the NPS integral to the desired frequency range. The length of each extended

focal spot was 0.4 mm, and each extended focal spot was discretized into 40 sourcelets. We set the maximum number of photons of each sourcelet to be 30000. All optimization was implemented in PyTorch with differentiable projector tools CTorch[8], using Adam as the optimizer. And we repeated the optimization procedure multiple times, using different learning rate from 10^3 to 10 and initialization sampled from a uniform distribution. We selected the two designs with the lowest loss values for further empirical NPS evaluation.

We performed iterative reconstructions of the disc phantom using the designed focal spots, as well as a small single sourcelet spot for comparison.A quadratic regularization term ($\beta = 1 \times 10^2$) was incorporated to suppress aliasing near the Nyquist frequency.

Each case was repeated with eight independent noise realizations. We cropped a 51×51 flat area around the central voxel to compute empirical NPS:

$$\mathrm{NPS}(f_u, f_v) = \frac{\Delta x\, \Delta y}{N_x N_y} \cdot \mathbb{E}\{|\mathcal{F}\{\mathbf{x} - \mathbb{E}\{\mathbf{x}\}\}|^2\} \qquad (8)$$

where Δx and Δy denote the voxel size, N_x and N_y represent the number of voxels in the region of interest along each spatial dimension. $\mathbf{x}$ here denotes ROI region cropped from the reconstructed image.

III. Result

We present two of the best optimized focal spot designs in Fig. 2. The total fluence of Design 1 is 9.3×10^5, and the total fluence of Design 2 is 9.5×10^5. Although each sourcelet was allowed to take any value between zero and the specified maximum fluence, both optimized designs exhibit a strongly binary behavior, with each sourcelet converging to either zero or the maximum allowed fluence. This binarization suggests that, under the proposed objective and constraints, the optimal focal spot patterns exhibit high-contrast emission structures rather than smoothly varying fluence distributions. The two selected designs differ in their fluence maps while maintaining a comparable design objective function value, suggesting that multiple local optima exist. Fig. 3 shows reconstructed images obtained using a conventional monolithic source (single sourcelet) and the two optimized focal spot designs. There appears to be a reduced noise level in the designed cases. This is evident in the reconstruction but much more clear in the standard deviation (STD) maps at the bottom of the figure. We also note that the noise textures are also visibly different. In particular, high frequency noise is reduced in designs 1 and 2, but mid-frequency noise is evident.

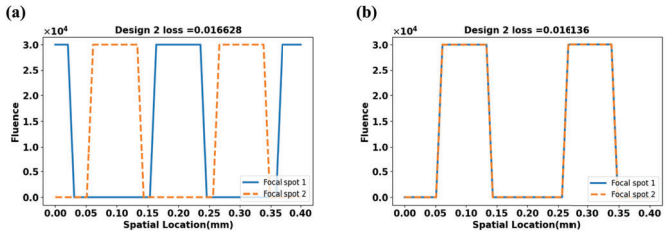


Fig. 2. (a): Focal-spot design 1; (b): Focal-spot design 2.

Fig. 4 shows both the analytically predicted and empirically estimated two-dimensional noise power spectra (NPS) for the single sourcelet and optimized focal spot designs. Importantly, the analytical NPS predictions closely match the empirical NPS estimates obtained from multiple noise realizations, validating the accuracy of the proposed analytic performance model for efficient reconstruction property prediction. The single sourcelet yields a ramp-like NPS (as one would expect for very low regulatization strength), indicating noise magnification in the high-frequency region. In contrast, both optimized designs exhibit a substantial suppression in the NPS within the target frequency band. However, both designed cases have a bright ring structure in a mid-frequency band (as we saw in the reconstructions). Thus, it appears the high-frequency noise reduction has come at the cost of an increase of mid-frequency noise.

To further illustrate the different X-ray source designs and their NPS, we computed the radial average of the NPS maps and plotted the resulting one-dimensional profiles, as shown in Fig. 5. Both optimized focal spot designs achieve substantial NPS reduction within the target frequency range $[9\ \mathrm{mm}^{-1}\ \ 10\mathrm{mm}^{-1}]$ compared to the point source. Outside this band, we see increased noise at mid-frequencies and lower noise at low frequencies. This tradeoff reflects the task-specific nature of the optimization, which can intentionally redistribute noise power away from frequencies of interest.

IV. Conclusion

We presented a task-driven, gradient-based focal spot design optimization framework for high-resolution CT based on an analytic model predictions of the properties of the imaging system. In simulation studies, we demonstrated that optimized focal spot designs can significantly suppress high-frequency noise and improve task-weighted detective quantum efficiency relative to a conventional point-source (single sourcelet). The strong agreement between analytical predictions and empirical measurements validates the proposed model for efficient, task-specific performance optimization.

Although this study focused on lightly regularized reconstruction to isolate the effects of focal spot design, the proposed framework is general and can be extended to incorporate regularized reconstruction, alternative task functions, and more complex anatomical imaging tasks. Future work will also investigate other task objectives, the robustness to model mismatch, extension to three-dimensional geometries, and integration with practical source constraints and system calibration considerations.

Acknowledgment

This work was supported, in part, by NIH grant R01EB035908.

References

[1] M. Remy-Jardin, A. Hutt, T. Flohr, J.-B. Faivre, P. Felloni, S. Khung, and J. Remy, "Ultra-high-resolution photon-counting ct imaging of the chest: a new era for morphology and function," *Investigative Radiology*, vol. 58, no. 7, pp. 482–487, 2023.

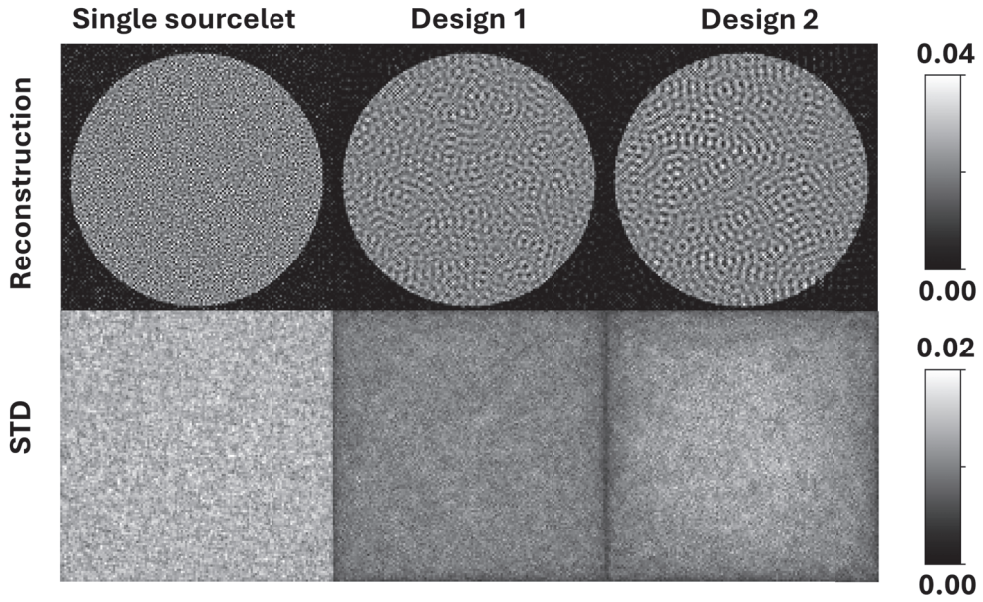


Fig. 3. Reconstructed image and Standard Deviation(STD) maps for the three focal spot scenarios. The top row shows the reconstructed images and the bottom row presents the corresponding (global) STD maps. From left to right, the columns correspond to the single-sourcelet, Design 1, and Design 2.

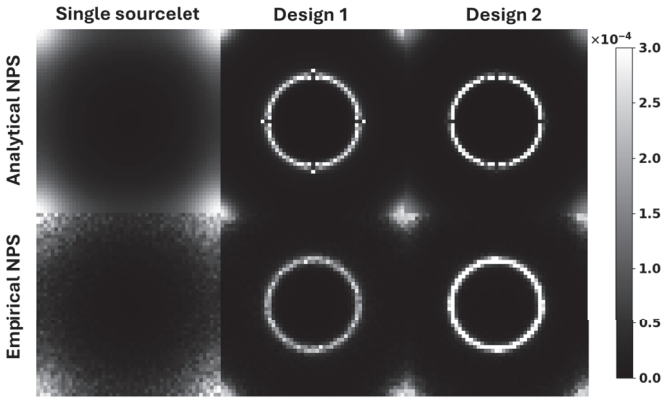


Fig. 4. NPS comparison for different focal spots. Top row is the analytical NPS;Bottom row is the empirical NPS; From left to right, the columns correspond to the single sourcelet, Design 1 and Design 2 cases.

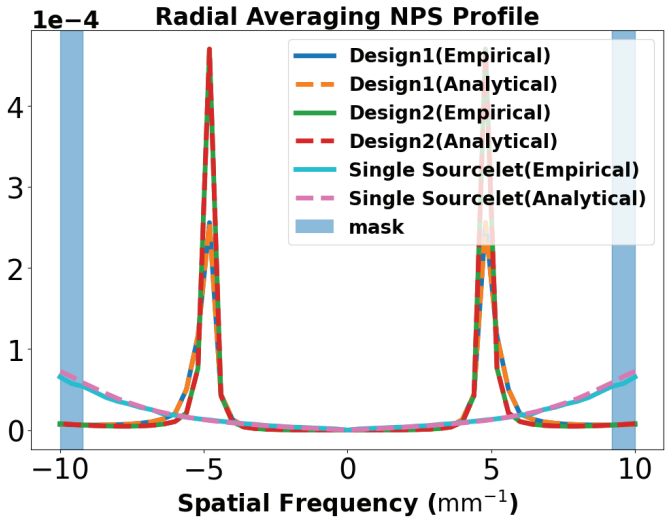


Fig. 5. 1D NPS profile. Solid lines show the empirically estimated NPS, while dashed lines indicate analytically predicted NPS. The single sourcelet NPS is also shown for comparison. The blue shaded regions denote the target spatial-frequency band.

[2] J. W. Stayman, P. B. Noël, and G. J. Gang, "Multiple focal spots for high resolution ct," in *Medical Imaging 2023: Physics of Medical Imaging*, vol. 12463. SPIE, 2023, pp. 160–165.
[3] N. J. Pelc and A. S. Wang, "Focal spot shape for computed tomography," Jan. 30 2024, uS Patent 11,883,226.
[4] J. W. Stayman, Y. Fan, X. Jiang, and G. J. Gang, "Design of multiple focal spots for high-resolution ct," in *Medical Imaging 2026: Physics of Medical Imaging*. SPIE, 2026.
[5] J. Fessler and W. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs," *IEEE Transactions on Image Processing*, vol. 5, no. 9, pp. 1346–1358, 1996.
[6] W. Wang, G. J. Gang, J. H. Siewerdsen, and J. W. Stayman, "Predicting image properties in penalized-likelihood reconstructions of flat-panel cbct," *Medical Physics*, vol. 46, no. 1, pp. 65–80, 2019. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.13249
[7] M. Tivnan, W. Wang, G. Gang, and J. W. Stayman, "Design optimization of spatial-spectral filters for cone-beam ct material decomposition," *IEEE Transactions on Medical Imaging*, vol. 41, no. 9, pp. 2399–2413, 2022.
[8] X. Jiang, G. J. Gang, and J. W. Stayman, "Ctorch: Pytorch-compatible gpu-accelerated auto-differentiable projector toolbox for computed tomography," *arXiv preprint arXiv:2503.16741*, 2025.

Scatter Characteristic and LBTE-based Correction in Multi-source Dual-ring Stationary CT

Guoxi Zhu, Hewei Gao, Li Zhang, Yuxiang Xing, Changyu Chen*, and Zhiqiang Chen*

***Abstract*—Multi-source stationary CT has strong potential for high-temporal-resolution imaging enabled by ultrafast switching among X-ray sources and simultaneous multi-source emission. However, simultaneous emission significantly increases scatter contamination which can substantially degrade image quality. In this work, we investigate the scatter characteristics of multi-source dual-ring stationary CT (MSDR-CT) and develop an efficient scatter correction method based on the linear Boltzmann transport equation (LBTE). Specifically, we employ GPU-accelerated Monte Carlo simulations to model and analyze scatter distributions in MSDR-CT under different scan strategies. Based on these observations, we propose an LBTE-based analytic framework that estimates both forward scatter and cross scatter contributions in MSDR-CT with multi-source emission. Simulation studies on a chest phantom and a water-cylinder phantom reveal distinctive scatter behaviors in MSDR-CT: scatter intensity increases with the number of simultaneously fired sources and is strongly influenced by source arrangement due to large-angle Compton scattering. Experiments further confirm the effectiveness of the proposed correction method under different scan strategies.**

***Index Terms*—Stationary CT, cross scatter, scatter correction, linear Boltzmann transport equation.**

I. Introduction

In recent years, multi-source stationary CT has attracted great attention for its potential to provide high-temporal-resolution imaging, which is achieved by combining fast switching of X-ray sources and simultaneous multi-source emission [1], [2]. However, simultaneous emission poses a major challenge for scatter correction: the measured signal is contaminated not only by forward scatter from the primary source but also by cross scatter from other sources fired at the same time. Moreover, stationary CT employs a non-standard imaging geometry with complicated projection relationship between sources and detectors. Each X-ray source may illuminate multiple detector modules, and each detector module receives primary and scatter signals from multiple sources which restricts the application of hardware-based scatter rejection techniques such as anti-scatter grids or collimators. Among software-based scatter estimation methods, kernel-based approaches are attractive due to their practicality and low computational cost. But their accuracy is often insufficient, and their effectiveness for estimating cross-scatter components may be limited. [3]. Monte Carlo (MC) methods, which are widely regarded as the gold standard for scatter estimation, typically require high computational cost. [4]. While deep learning-based methods have the potential to enable fast and accurate scatter estimation, their reliance on large-scale training datasets and limited physical interpretability can restrict the applicability in practice.

Linear Boltzmann Transport Equation (LBTE) has been used to describe the statistical transport behavior of particles. LBTE-based scatter correction methods offer high accuracy and general applicability, and are naturally amenable to parallel implementation, enabling fast computation on modern GPU architectures, as demonstrated in previous studies. [5], [6], [7].

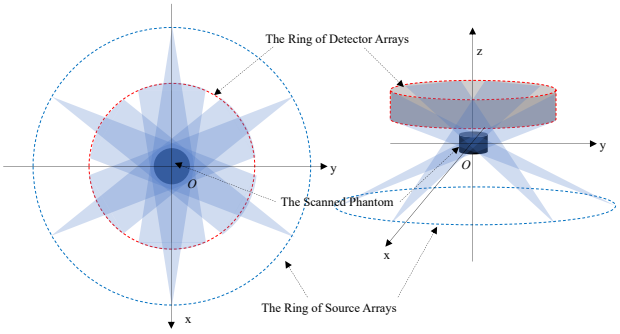


Fig. 1. Architecture of the MSDR-CT system. The overall 3D view is shown on the right and the top view on the left. The detector array (red ring) and source array (blue ring) are axially offset along the z-axis, with the scanned phantom positioned between the two rings.

In this work, we investigated the scatter distribution characteristics in Multi-source Dual-ring Stationary CT (MSDR-CT) [2] (see Fig. 1) under different scan strategies. We propose a GPU-accelerated, LBTE-based scatter correction method and conduct simulation experiments to evaluate its performance under different scan strategies.

II. Methods

A. Scatter Signal Analysis in MSDR-CT

In MSDR-CT, the primary signal from an X-ray source is not only affected by its own forward scatter signal but also by cross scatter signals originating from other simultaneously active X-ray sources. For the scatter signal produced by a single source, its distribution across the entire detector ring exhibits a characteristic pattern: strong forward scatter at 0°

This work was supported in part by the National Natural Science Foundation of China under Grant 12327809, in part by the Beijing Natural Science Foundation No. L252021, and in part by China Postdoctoral Science Foundation No. 2025M780861

Guoxi Zhu, Hewei Gao, Li Zhang, Yuxiang Xing, Changyu Chen, and Zhiqiang Chen was with the with Key Laboratory of Particle and Radiation Imaging (Tsinghua University), Ministry of Education, China and Department of Engineering Physics, Tsinghua University, Beijing 100084, China. (Corresponding authors: Changyu Chen, Email: ccy@mail.tsinghua.edu.cn; Zhiqiang Chen, Email: czq@mail.tsinghua.edu.cn)

(mainly attributed to Rayleigh scattering), weak scatter at 90° , and strong cross scatter at 180° (mainly attributed to Compton scattering), as illustrated in Fig. 2(b). Consequently, the multi-source scatter signals are influenced by both the number of simultaneously active sources and their spatial distribution. Without loss of generality, we investigated four representative scan strategies in this work (see Fig. 2):

(1) $\mathbf{Src1}$: each source is fired individually.

(2) $\mathbf{Src3Adja}$: Three adjacent sources on the ring are fired simultaneously.

(3) $\mathbf{Src3Equi}$: Three sources separated by equivalent angular interval (120° apart) on the ring are fired simultaneously.

(4) $\mathbf{Src6}$: Six sources are fired simultaneously.

The Monte Carlo simulation software package GGEMS was adapted for the MSDR-CT geometry to simulate different scan strategies. Each X-ray source was sequentially fired individually in simulations, and all primary and scatter signals were recorded on the entire detector ring. There are no anti-scatter grids or mini-bowties in our system used. The projection data were then combined according to the respective scan strategies.

B. Scatter Estimation

In CT scans, the X-ray radiation field distribution can be described by LBTE, which takes the following formula,

$$\overrightarrow{\Omega}\cdot\nabla\phi(\overrightarrow{r},E,\overrightarrow{\Omega})+\mu_t(\overrightarrow{r},E)\phi(\overrightarrow{r},E,\overrightarrow{\Omega})=S(\overrightarrow{r},E,\overrightarrow{\Omega})+\int_E^{E_{max}}dE'\int_{4\pi}d\overrightarrow{\Omega}'[\mu_s(\overrightarrow{r},E'\to E,\overrightarrow{\Omega}'\to\overrightarrow{\Omega})\phi(\overrightarrow{r},E',\overrightarrow{\Omega})] \tag{1}$$

Here, $\overrightarrow{r}$ represents spatial position, $\overrightarrow{\Omega}$ represents flight direction, E represents energy of photons, ϕ represents the photon fluence distribution, S represents the photon source distribution, μ_t represents the linear attenuation coefficient distribution, and μ_s represents the linear scatter coefficient distribution, E_{max} represents the maximum energy of photons. Within the X-ray energy range used in medical imaging, the scattered photons are almost entirely contributed by the Compton effect and the Rayleigh effect, which were considered in the μ_s. This integro-differential equation can be solved using the iterative source method. First, we expand the photon fluence and source distribution as an infinite summation according to times of photon scattered: $\phi(\cdot)=\sum_{k=0}^{\infty}\phi^{(k)}(\cdot)$ and $S(\cdot)=\sum_{k=0}^{\infty}S^{(k)}(\cdot)$, where k represents scattering order, indicating that $S^{(k)}$ and $\phi^{(k)}$ are contributed by the photons that have been scattered k times. Then Eq. (1) can be separated into the following equations,

$$\overrightarrow{\Omega}\cdot\nabla\phi^{(k)}(\cdot)+\mu_t(\cdot)\phi^{(k)}(\cdot)=S^{(k)}(\cdot), \tag{2a}$$

$$S^{(k+1)}(\cdot)=\int_E^{E_{max}}\int_{4\pi}\mu_s(\cdot)\phi^{(k)}(\cdot)dE'd\overrightarrow{\Omega'}, \tag{2b}$$

where, $\phi^{(k)}(\cdot)$ and $S^{(k)}(\cdot), k\geq 1$ represent the higher-order scattered photon fluence distribution and source distribution in the phantom. While $S^{(0)}(\cdot)$ term is an exception, it indicates the distribution of X-ray sources position and their emission intensity in each direction (only the photons generated by the X-ray source is zero-order photon without scattering). By repeatedly iterating from $S^{(k)}(\cdot)\to\phi^{(k)}(\cdot)\to S^{(k+1)}(\cdot)$ for K iterations, we can numerically estimate $S(\cdot)$ with K-th-order accuracy. So we can calculate the scatter signal detected according to the following process:

$\mathbf{Step1}$, obtain $\phi^{(0)}(\cdot)$ from $S^{(0)}(\cdot)$ based on Beer-Lambert law,

$\mathbf{Step2}$, obtain $S^{(k+1)}(\cdot)$ from $\phi^{(k)}(\cdot)$ based on Eq. (2b), then obtain $\phi^{(k+1)}(\cdot)$ form $S^{(k)}(\cdot)$ based on Eq. (2a), and continue until reaching the maximum K,

$\mathbf{Step3}$, obtain scatter signals from $\sum_{k=1}^{K}S^{(k)}(\cdot)$ based on Beer-Lambert law.

After discretization [8], the calculation of $\mathbf{Step2,3}$ can be written in matrix form as:

$$\overrightarrow{\phi^{(k)}}=\mathbf{T}\overrightarrow{S^{(k)}}, k\geq 1, \tag{3a}$$

$$\overrightarrow{S^{(k+1)}}=\mathbf{C}\overrightarrow{\phi}^{(k)}, k\geq 0, \tag{3b}$$

$$\overrightarrow{I^{(S)}}=\mathbf{R}\sum_{k=1}^{N_S}\overrightarrow{S^{(k)}}, \tag{3c}$$

where, $\mathbf{T},\mathbf{C},\mathbf{R}$ represent transport, scattering and reception matrix, $\overrightarrow{\phi^{(k)}}$ and $\overrightarrow{S^{(k)}}$ represent discrete form of $\phi^{(k}(\cdot)$ and $S^{(k)}(\cdot)$, $\overrightarrow{I^{(S)}}$ represent detected scatter signals. Step 2 corresponds to Eqs. 3a and 3b, and Step 3 corresponds to Eqs. 3c. Combining these equation, one gets:

$$\overrightarrow{I^S}=\mathbf{RC}\left(\sum_{k=1}^{K}(\mathbf{TC})^{k-1}\right)\overrightarrow{\phi^{(0)}}, \tag{4}$$

From Eq. 4, once $\overrightarrow{\phi^{(0)}}$ is obtained following $\mathbf{Step1}$, based on the position of the X-ray source, the scatter signals can be calculated. Compared to the calculation process for single-source emission, the process for simultaneous multi-source emission differs only in the calculation of $\phi^{(0)}$, while the rest of computations remain unchanged. This highlights the advantage of scatter estimation based on LBTE: it efficiently calculates the scatter field distribution for the entire scanned object in a unified manner, regardless of the X-ray source configuration. The distribution of X-ray sources position and their emission intensity only provides an initial value for $\phi^{(0)}$, allowing the method to seamlessly adapt to both single-source and multi-source simultaneous emission scenarios without altering the core computational framework. For single source emission ($\overrightarrow{S^{(0)}}$), $\overrightarrow{\phi^{(0)}}=\mathbf{T_0}\overrightarrow{S^{(0)}}$; for multi-source simultaneous emission ($\overrightarrow{S_i^{(0)}}, i=0,1,...$), $\overrightarrow{\phi^{(0)}}=\sum_i\mathbf{T_{0,i}}\overrightarrow{S_i^{(0)}}$, where $\mathbf{T_{0;i}}$ represent emission matrix of $\overrightarrow{S_i^{(0)}}$.

C. Scatter Correction and Image Reconstruction

After the scatter signals have been estimated based on LBTE, they are subtracted from the measured signals. However, the presence of noise and residual errors may result in non-physical negative values after subtraction, which can cause anomalies in the subsequent logarithmic calculations. To address this issue, a upper limit for the scatter-to-primary ratio (SPR) is imposed. In this work ,the limit $SPR_{\mathbf{max}}$ is set as 19.

$$\overrightarrow{I_{\mathbf{sc}}^P}=\overrightarrow{I_{\mathbf{m}}^P}-\overrightarrow{I_{\mathbf{m}}^P}\cdot\min(\frac{\overrightarrow{I^S}}{\overrightarrow{I_{\mathbf{m}}^P}},\frac{1}{1/SPR_{\mathbf{max}}+1}), \tag{5}$$

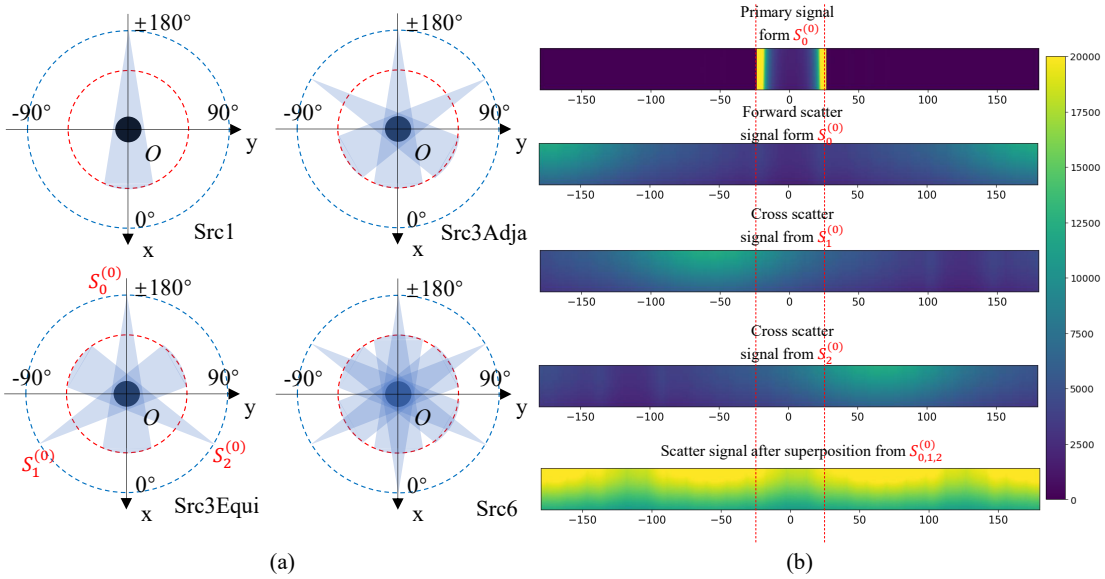


Fig. 2. (a) Illustrations of four representative scan strategies; (b) primary, forward scatter, and cross scatter distribution on the detector ring (using **Src3Equi** as an example).

where $\overrightarrow{I_{\mathbf{m}}^{P}}$ represents measured signal, $\overrightarrow{I_{\mathbf{sc}}^{P}}$ denotes the measured signal after scatter correction. After scatter correction, we employ the SIRT iterative algorithm to obtain the reconstruction images.

III. Result

A. Scatter Characteristics

To explore the scatter characteristics of MSDR-CT, we performe simulations with a water cylinder phantom and a human chest phantom. The primary and scatter signals for these phantoms under four scan strategies (**Src1**, **Src3Adja**, **Src3Equi** and **Src6**) are shown in Fig. 3. Compared to single-source emission, simultaneous multi-source emission leads to the superposition of scatter signals from multiple X-ray sources, significantly increasing the SPR. For the chest phantom, the maximum SPR increases from approximately 0.8 under single-source emission to nearly 5 under six-source simultaneous emission. Similarly, for the water cylinder phantom, the maximum SPR increases from approximately 1 to over 15. The cross-scatter signals, primarily caused by large-angle Compton scattering, become smoother as the number of simultaneously fired sources increases. Larger angular intervals between simultaneously fired sources strengthen mutual cross-scattering interference due to Compton back-scattering. While small-angle interference from forward scattering at the edges is possible, it is avoided in our experiments by preventing overlap between primary signals.

B. Evaluation of Scatter Estimation

Fig. 4 compares the primary and scatter signals obtained from MC simulation and LBTE calculation under the **Src1** scan strategy, along with the corresponding profiles for a water cylinder phantom. Quantitatively, the mean relative error

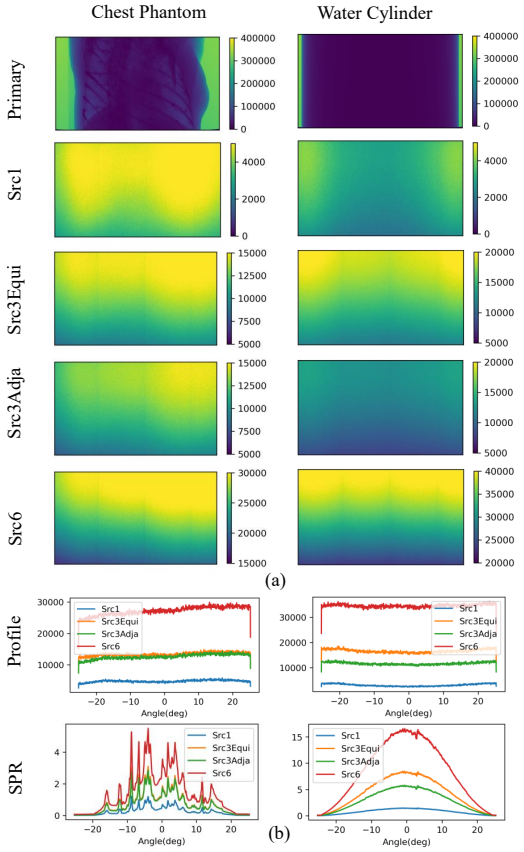


Fig. 3. (a) The primary and scatter signals for the chest phantom and water cylinder phantom under four different scan strategies with colorbar unit in keV. (b) Scatter signal and SPR profiles by different scan strategies with the y-axis unit of scatter profile figure is keV.

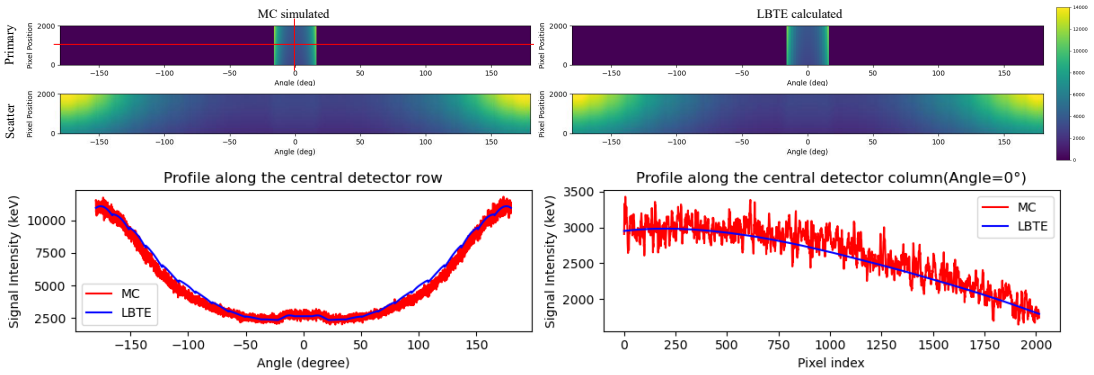


Fig. 4. The comparison of primary and scatter signals of water cylinder phantom estimated by MC methods and our implemented LBTE-based method, and their profile. The window is [0, 14000], and the unit is keV

(MRE) across the entire detector ring is 7.05%. The scatter signals estimated by LBTE show good agreement with the MC simulation results around 0° and 180°, while larger deviations are observed around 90°. Although this deviations have no impact under single-source emission, it reduces the accuracy of cross-scatter signal estimation during multi-source simultaneous emission, leading to residual scatter artifacts after correction, as shown in Fig.5.

C. Reconstruction Results

Fig. 5 presents the reconstructed images of the water cylinder phantom before and after scatter correction under different scan strategies. As the number of simultaneously fired sources increases, scatter signals rise significantly, leading to severe artifacts in the reconstructed images of uncorrected data. Our scatter correction method performs well for **Src1**. For **Src3Adja** and **Src3Equi**, most scatter effects are effectively removed, with minor residual errors observed in the central region. For **Src6**, the performance is degrades, mainly due to reduced scatter estimation accuracy in the large-angle region (about 90°), as shown in Fig. 4.

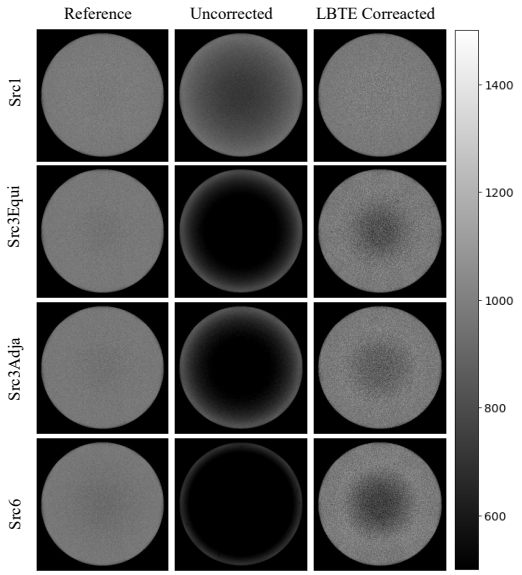


Fig. 5. Reconstruction images of water cylinder phantom with or without scatter correction under different scan strategies, and their profiles. The window is [500HU, 1500HU]

IV. Discussion

In MSDR-CT, simultaneous multi-source emission can significantly reduce scan time of stationary CT scans, which fully utilizing its multi-source and multi-detector architecture. However, cross-scattering from multi-source emission severely degrades image quality. Based on MC simulation and SIRT algorithm, we analyzed scatter distributions in MSDR-CT, and evaluated the effectiveness of our LBTE-based scatter correction methods under different scan strategies. The results show that our correction method performs effectively under **Src1**. Under **Src3Adja**, **Src3Equi**, and **Src6**, there are still residual artifacts, which indicates that the current method lacks accuracy in cross-scatter estimation.

This LBTE-based scatter correction method has serveral limitations. First, the current scatter estimation is not accuracy enough at large angles, especially around 90° as shown in Fig. 4, which may be an issue with discretization in the code implementation and we are working to analyze and resolve it. Second, only the mean values of scatter signals are removed after scatter subtraction, and residual scatter noise is superimposed on primary signals, there by reducing the image quality even if scatter estimation is very accurate.

Future work will involve further improving the computing speed and accuracy, more phantom studies, suppressing image degradation caused by scattering noise.

V. Conclusion

In this work, we investigated the characteristics of scatter signals in the architecture of MSDR-CT. We implemented a scatter estimation and correction method based on LBTE for this system, and preliminarily verified the effectiveness of this method.

References

[1] C. Chen, Y. Xing, L. Zhang, and Z. Chen, “Static CT With Sources and Detectors Distributed in a Multi-Segment Manner: System Analysis

and Analytical Reconstruction," *IEEE Transactions on Computational Imaging*, vol. 11, pp. 251–264, 2025.

[2] Y. Xia, L. Zhang, Y. Xing, Z. Chen, and H. Gao, "ComptoNet: A Compton-map guided deep learning framework for multi-scatter estimation in multi-source stationary CT," *Physics in Medicine & Biology*, vol. 70, no. 14, p. 145022, Jul. 2025.

[3] M. Sun and J. M. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Physics in Medicine & Biology*, vol. 55, no. 22, p. 6695, Oct. 2010.

[4] A. Badal and A. Badano, "Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit," *Medical Physics*, vol. 36, no. 11, pp. 4878–4880, 2009.

[5] A. Maslowski, A. Wang, M. Sun, T. Wareing, I. Davis, and J. Star-Lack, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - Part I: Core algorithms and validation," *Medical Physics*, vol. 45, no. 5, pp. 1899–1913, May 2018.

[6] T. Niu, L. Xu, Q. Ren, Y. Gao, C. Luo, Z. Teng, J. Du, M. Ding, J. Xie, H. Han, and Y. Jiang, "UBES: Unified scatter correction using ultrafast Boltzmann equation solver for conebeam CT," *Computers in Biology and Medicine*, vol. 170, p. 108045, Mar. 2024.

[7] G. Zhu, Z. Chen, and H. Gao, "Unified scatter estimation in x-ray spectral cone-beam CT using linear Boltzmann transport equation with labels on energy groups," in *Medical Imaging 2025: Physics of Medical Imaging*, vol. 13405. SPIE, Apr. 2025, pp. 346–349.

[8] G. Zhu, L. Zhang, Z. Chen, and H. Gao, "Matrixed-Spectrum Decomposition Accelerated Linear Boltzmann Transport Equation Solver for Fast Scatter Correction in Multi-Spectral CT," arXiv preprint arXiv:2508.20524, Aug. 2025.

Measurement-based Deep Scatter Estimation

Lukas Hennemann, Julien Erath, Andreas Heinkele, Eric Fournié, Martin Petersilka, Karl Stierstorfer, and Marc Kachelrieß

***Abstract*—Photon-counting computed tomography (PCCT) provides energy-resolved detection with the potential for improved image quality and enhanced quantitative accuracy. However, scatter correction remains a major challenge, particularly due to high-frequency artifacts caused by coarse anti-scatter grids and the use of sharp reconstruction kernels. In this work, we propose a measurement-based deep scatter estimation (mbDSE) approach that is trained exclusively on measured data from a photon-counting CT system, thereby eliminating the domain gap typically introduced by Monte-Carlo-based training. Experiments on a pelvis phantom demonstrate that mbDSE substantially reduces both low-frequency scatter artifacts and high-frequency ring-like artifacts, outperforming a kernel-based scatter estimation (KSE) method and producing nearly artifact-free images. Furthermore, quantitative evaluation using the mean absolute error (MAE) on an anthropomorphic thorax phantom shows consistently lower errors for mbDSE compared to KSE. Importantly, mbDSE generalizes well to detector regions near the z-edge, despite being trained solely on data acquired near the detector center. These results indicate that measurement-based training markedly improves the robustness and clinical relevance of deep-learning-based scatter correction for PCCT.**

***Index Terms*—scatter, deep learning, photon-counting CT.**

I. Introduction

WITHIN the last years photon-counting computed tomography (PCCT) has established itself as a promising technology, as its energy-resolved detection enables improved material decomposition, a potential dose reduction compared to conventional energy integrating CT as well as overall better image quality [1]–[6]. Despite these advantages, scatter correction remains one of the biggest challenges, as scattered photons significantly impair image quality and quantitative accuracy. Monte Carlo (MC) simulations, which are considered the gold standard for modeling scatter processes, are traditionally used to develop scatter correction methods. However, these simulations often do not adequately reflect the complex physical and geometric properties of clinical CT systems, which limits their transferability to real-world applications. The requirements for scatter correction methods are particularly high in PCCT. Due to the significantly smaller pixels sizes compared to conventional energy-integrating systems, several pixels can lie between the lamellae of the anti-scatter grid (ASG) [4], [7], [8]. This geometry favors the occurrence of high-frequency artifacts, such as partial ring patterns, which are image-degrading in clinical imaging, highlighting the fact that such artifacts require a correction strategy specifically adapted to PCCT. Nevertheless, most of these approaches are based on MC simulated training data, which limits their robustness against real-world system variations. In this paper, we therefore propose a deep learning-based approach, measurement-based deep scatter estimation (mbDSE), that is trained on measured data from a PCCT system. By using real measurements, we address the discrepancy between simulation and practice and improve the clinical relevance of the model. While established works such as the original proposed deep scatter estimation [9], [10] were trained on MC data, we deliberately pursue a practice-oriented approach that considers the specific properties of photon-counting systems.

Lukas Hennemann, Andreas Heinkele are with the Division of X-ray Imaging and CT of the German Cancer Research Center (DKFZ), Heidelberg, Germany, the Computed Tomography Division of Siemens Healthineers AG, Forchheim Germany, and the Medical Faculty Heidelberg of the Heidelberg University, Heidelberg Germany.

Dr. Julien Erath, Eric Fournié, Dr. Martin Petersilka and Dr. Karl Stierstofer are with the Computed Tomography Division of Siemens Healthineers AG, Forchheim, Germany.

Prof. Dr. Marc Kachelrieß is with the Division of X-ray Imaging and CT of the German Cancer Research Center (DKFZ), Heidelberg, Germany, and the Medical Faculty Heidelberg of the Heidelberg University, Heidelberg Germany.

Corresponding author: Lukas Hennemann (lukas.hennemann@dkfz.de)

II. Methodology

A. Data acquisition

Considering that scatter radiation is proportional to the z-collimation, it follows that a scan with very narrow collimation produces almost no scatter. This assumption is leveraged in this work. To train the network, we require projections as input that contain both primary intensity and scatter. These projections are acquired from scans with full collimation (144×0.4 mm=57.6 mm). For the network labels, scatter intensity profiles are needed. Assuming that a scan with narrow collimation (6×0.4 mm=2.4 mm) captures only primary intensity, we compute the scatter intensities by subtracting the narrowly collimated scan from the fully collimated scan (see Figure 1).

For the dataset used in this study, scans were acquired with a NAEOTOM Alpha.Peak (Siemens Healthineers AG, Forchheim, Germany) system. Various phantoms were included, such as simple water phantoms, QA multi-energy phantoms, and anthropomorphic patient-like phantoms, using the method described above. The phantoms were scanned sequentially with an automated script that moved the patient

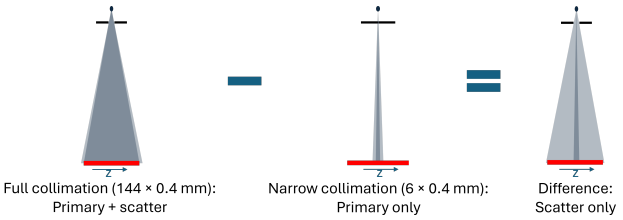


Fig. 1. Methodology to obtain scatter intensities.

table in predefined steps along the z-axis. At each z-position, the table height was randomly varied around the isocenter. Furthermore, 36 projections were acquired per z-position (every 10°). To improve photon statistics, the scans were performed without rotation (static scan), and the projections recorded for each scan were averaged before computing the difference between the different collimations. Scans were conducted using the default bowtie filter and energy thresholds (20 and 65 keV). The tube voltage was set to 140 kV with 100 mAs per scan. In total, the training dataset for this study consists of 5,760 projections (36 projections per z-position, 160 z-positions). As the test dataset, scans of an anthropomorphic thorax and an anthropomorphic pelvis phantom were used. The scan parameters were chosen equally to the training dataset.

B. Network architecture

The network architecture used in this work is the well-known DSE model from the literature [9], [10]. Most implementations of DSE are based on MC simulated data [9]–[14]. In this study, however, DSE is trained using measured data obtained with the method described above. To adapt the DSE to the coarse anti-scatter grids of the photon-counting detector, both the network input and output are modified [15]. Instead of using one input channel per energy threshold, the signals are divided into sub-signals according to their pixel position relative to the ASG. This results in six sub-signals per energy threshold for both input and output. After passing through the network, these sub-signals are recombined into a single detector signal. Each input and output channel has a size of 688×2 pixels. (The two pixels in the z-direction are a result of the narrow collimation. Since only six detector rows are illuminated, the scatter distribution is available only for these six rows. When these signals are then divided into sub-signals for each macro pixel, this leads to the two pixels.) Since two energy thresholds are used, the total number of channels for both input and output is twelve. The architecture is illustrated in Figure 2, which also shows the detector layout with the different pixel positions.

C. Evaluation

To evaluate the performance of our approach, the scatter-corrected images are reconstructed using the filtered back-projection algorithm. Two reconstruction kernels are employed: a softer qualitative kernel (Qr40) and a sharper reconstruction one (Qr76), the latter of which may exhibit partial ring artifacts if correction is not applied. Our method is compared to the kernel-based approach (KSE) proposed by Ohnesorge et al. [16] In addition, a quantitative assessment of the reduction in the amplitude of high-frequency rings is performed. Reconstructions of both central slices and slices located at the periphery of the cone beam were analyzed. For the quantitative evaluation of CT values, only slices for which a slit scan is available (i.e., slices within the detector rows of the narrowly collimated scan) were considered. The image quality in this case was assessed using the mean absolute error (MAE) relative to the ground truth (slit scan, GT). Only the region within the patient was included in the evaluation.

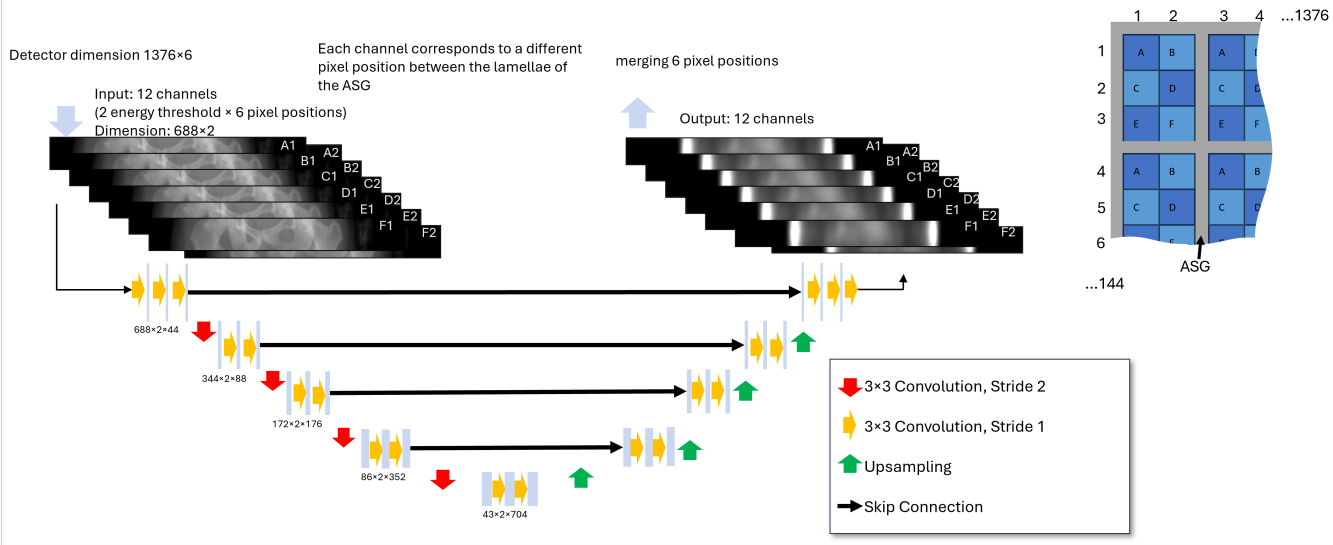


Fig. 2. The U-Net has six input and output channels per energy threshold. Those positions correspond to the different pixel positions (A-F) within the ASG.

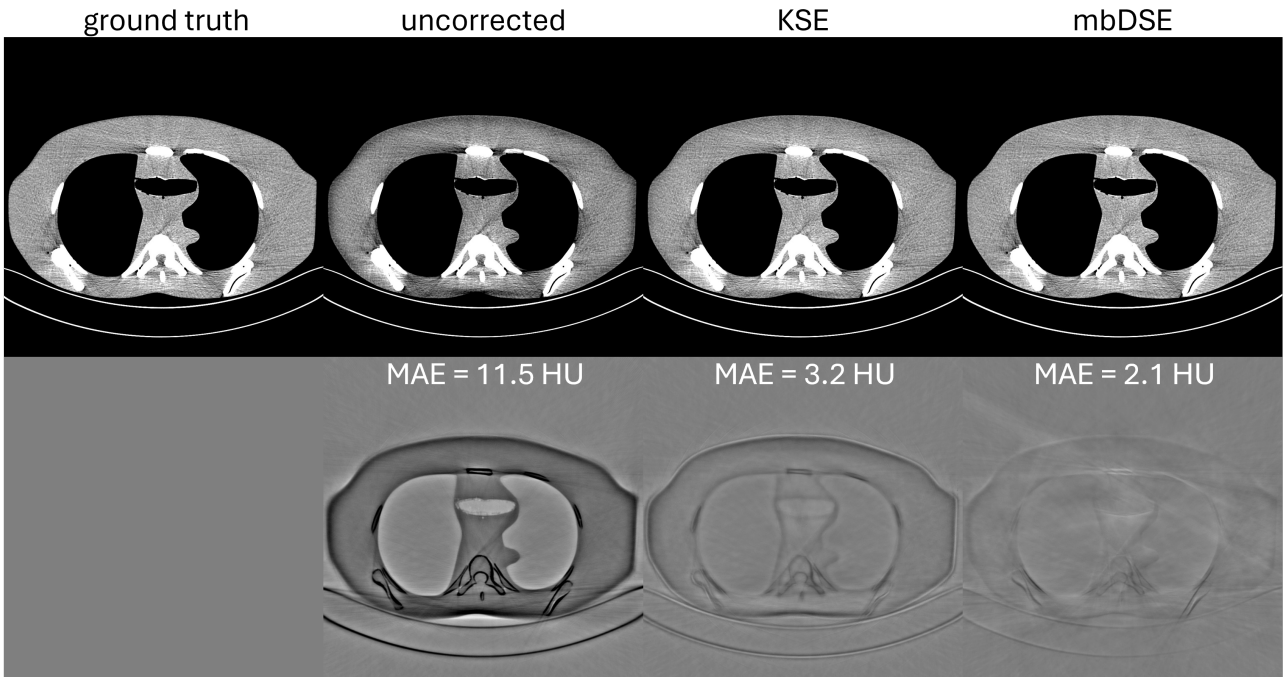


Fig. 3. Example of scatter correction for 20 keV threshold images of an anthropomorphic thorax phantom. Two different scatter correction methods are shown as well as the ground truth (row 1) and the uncorrected image. Each MAE excluding air is shown in the difference to ground truth images (row 2). The reconstructed slices are within the detector rows of the narrow collimated scan. The reconstruction was performed with a Qr40 kernel and a 380 mm FOV. Reconstructed images: $C = 40$ HU, $W = 120$ HU. Difference images: $C = 0$ HU, $W = 100$ HU.

III. Results

A. Reconstructions with smooth kernel

Figure 3 compares the proposed mbDSE with the KSE approach for a centered slice of a scan of a thorax phantom (20 keV threshold image). It can be observed that both methods are able to clearly reduce the image MAE compared to the uncorrected reconstruction. Without scatter correction the MAE in the patient is 11.5 HU. KSE reduces the error to 3.2 HU, while our approach leads to a MAE of 2.1 HU. Results from a pelvis phantom demonstrate that the proposed mbDSE scatter correction substantially reduces artifacts compared to the uncorrected images. The displayed slices at opposite edges of the detector (rows 1 and 2), which were not included in the training that used only the central six detector slices, show consistently improved image appearance, with smoother CT-value distributions in soft tissue and reduced edge darkening caused by bowtie scatter (see Figure 4). Overall, mbDSE outperforms the KSE method in terms of artifact suppression and visual image quality.

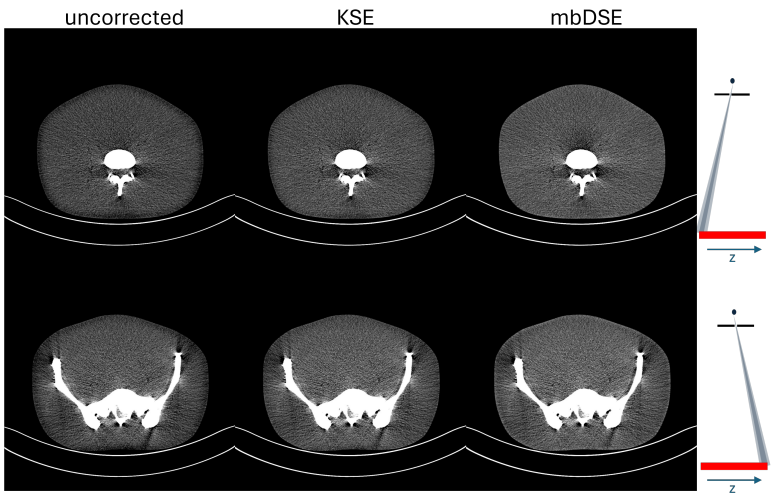


Fig. 4. Example of scatter correction for 20 keV threshold images of an anthropomorphic pelvis phantom. Two different scatter correction methods as well as the uncorrected images are shown. Both rows show images reconstructed from data at the edge of the detector. The reconstruction was performed with a Qr40 kernel and a 380 mm FOV. Reconstructed images: $C = 20$ HU, $W = 120$ HU.

B. Reconstructions with sharp kernel

The performance of the proposed method in correcting high-frequency scatter artifacts arising from coarse ASGs in photon-counting detectors, in combination with reconstruction using a very sharp kernel, is illustrated in Figure 5. While the KSE method effectively compensates low-frequency scatter artifacts, it has no observable impact on the ring-like high-frequency artifacts. In contrast, the proposed mbDSE

approach successfully corrects both low-frequency scatter components and high-frequency artifacts. Quantitative evaluation of the ring amplitude along a line profile through the isocenter clearly demonstrates the reduction of these artifacts: whereas the uncorrected and KSE-corrected images exhibit minimum and maximum CT-values of -762 and 820 HU, respectively, the amplitude is reduced to -131 and 183 HU when applying mbDSE. This leads to a reduction in the mean absolute error from approximately 109 HU to about 33 HU, corresponding to an overall improvement of roughly 70 %.

IV. Discussion and Conclusion

In this work, DSE was trained using measured data from a photon-counting CT scanner rather than relying on Monte-Carlo simulations. The results demonstrate that mbDSE not only outperforms the KSE method but also produces images that are nearly free of scatter-related artifacts. In addition to providing effective correction in central detector regions near the isocenter, mbDSE generalizes well to detector areas at the z-edges, despite the absence of corresponding training data. Furthermore, mbDSE is capable of almost completely removing high-frequency ring-shaped artifacts that arise from the combination of coarse anti-scatter grids and sharp reconstruction kernels. This highlights the methods ability to address both low- and high-frequency scatter components, which is a key requirement for modern photon-counting CT systems. In addition, it has been shown that using multiple energy thresholds simultaneously for scatter correction can further improve performance, potentially enabling even more accurate correction with mbDSE [14]. Currently, the training dataset is limited to projections acquired near the detector center in the z-direction. Future work will focus on extending the dataset to include ground-truth measurements across the full detector width as well as increasing the variability of acquisition parameters, such as tube voltage and tube current. These extensions aim to further improve the robustness, generalizability, and clinical applicability of the mbDSE approach.

References

[1] S. Sawall, L. Klein, C. Amato, E. Wehrse, S. Dorn, J. Maier, S. Heinze, H.-P. Schlemmer, C. Ziehner, M. Uhrig, and M. Kachelrieß, "Iodine contrast–to–noise ratio improvement at unit dose and contrast media volume reduction in whole–body photon–counting CT," *European Journal of Radiology*, vol. 126, p. 108909, May 2020.

[2] M. M. Lell and M. Kachelrieß, "Recent and upcoming technological developments in computed tomography: high speed, low dose, deep learning, multienergy," *Investigative Radiology*, vol. 55, no. 1, pp. 8–19, Jan. 2020.

[3] S. S. Hsieh, S. Leng, K. Rajendran, S. Tao, and C. H. McCollough, "Photon–counting CT: clinical applications and future developments," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 4, p. 441–452, Jul. 2021.

[4] K. Rajendran, M. Petersilka, A. Henning, E. R. Shanblatt, B. Schmidt, T. G. Flohr, A. Ferrero, F. Baffour, F. E. Diehn, L. Yu, P. Rajiah, J. G. Fletcher, S. Leng, and C. H. McCollough, "First clinical photon-counting detector CT system: technical evaluation," *Radiology*, vol. 303, no. 1, p. 130–138, Apr. 2022.

[5] M. M. Lell and M. Kachelrieß, "Computed tomography 2.0: New detector technology, AI, and other developments," *Investigative Radiology*, vol. 58, no. 8, pp. 587–601, Aug. 2023.

[6] A. Meloni, E. Maffei, V. Positano, A. Clemente, C. D. Gori, S. Berti, L. L. Grutta, L. Saba, E. Bossone, C. Mantini, C. Cavaliere, B. Punzo, S. Celi, and F. Cademartiri, "Technical principles, benefits, challenges, and applications of photon counting computed tomography in coronary imaging: a narrative review," *Cardiovascular Diagnosis and Therapy*, vol. 14, no. 4, 2024.

[7] T. Flohr, M. Petersilka, A. Henning, S. Ulzheimer, J. Ferda, and B. Schmidt, "Photon–counting CT review," *Physica Medica*, vol. 79, p. 126–136, Nov. 2020.

[8] T. Flohr and B. Schmidt, "Technical basics and clinical benefits of photon–counting CT," *Investigative Radiology*, vol. 58, no. 7, pp. 441–450, Jul. 2023.

[9] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (DSE): accurate real–time scatter estimation for x–ray CT using a deep convolutional neural network," *Journal of Nondestructive Evaluation*, vol. 37, no. 3, p. 57 (9 pages), Jul. 2018.

[10] J. Maier, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. Kachelrieß, "Real–time scatter estimation for medical CT using the deep scatter estimation: method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation," *Med. Phys.*, vol. 46, no. 1, pp. 238–249, Nov. 2019.

[11] J. Erath, T. Vöth, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep learning-based forward and cross-scatter correction in dual-source CT," *Med. Phys.*, vol. 48, no. 9, pp. 4824–4842, Sep. 2021.

[12] A. Heinkele, J. Erath, L. Hennemann, J. Maier, E. Fournié, J. Sunnegaardh, C. Hofmann, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep scatter estimation for static CT using multiple projections," *SPIE Medical Imaging Proc.*, vol. 13405, Apr. 2025.

[13] L. Hennemann, J. Erath, A. Heinkele, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep bowtie and patient scatter correction applied to clinical photon–counting CT," *SPIE Medical Imaging Proc.*, vol. 13405, Apr. 2025.

[14] L. Hennemann, J. Erath, A. Heinkele, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Spectral deep learning-based patient and bowtie scatter correction for clinical photon-counting CT," *Med. Phys.*, 2026, accepted.

[15] J. Erath, J. Magonov, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep scatter estimation for coarse anti-scatter grids as used in photon-counting CT," *SPIE Medical Imaging Proc.*, vol. 12304, Feb. 2025.

[16] B. Ohnesorge, T. Flohr, and K. Klingenbeck-Regn, "Efficient object scatter correction algorithm for third and fourth generation CT scanners," *European Radiology*, vol. 9, pp. 563–569, 1999.

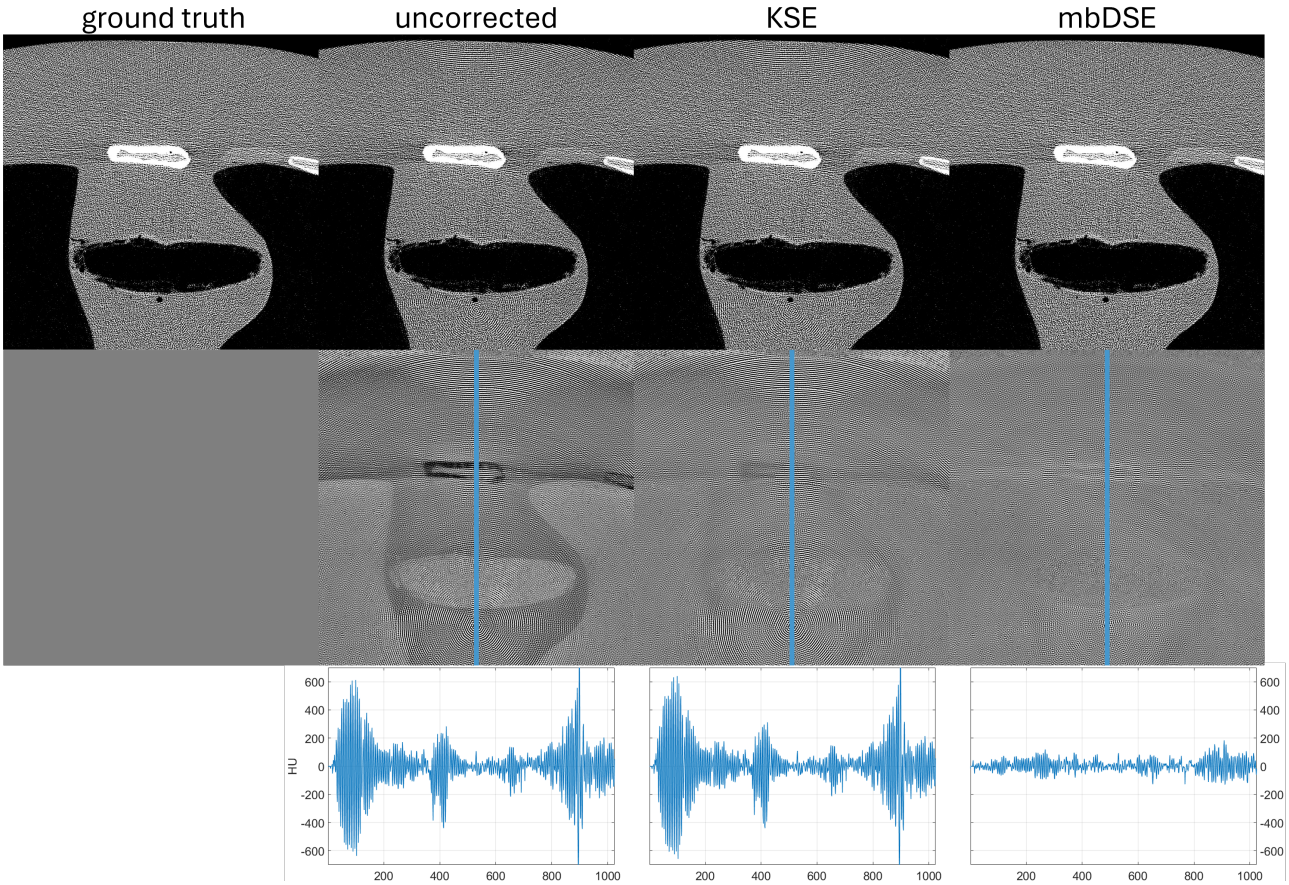


Fig. 5. Example of scatter correction for 20 keV threshold images of an anthropomorphic thorax phantom. The reconstructed slices are within the detector rows of the narrow collimated scan. The reconstruction was performed with a sharper Qr76 kernel and a 120 mm FOV. The third row displays the plot along the line marked in blue. The amplitude of the ring artifacts in CT values can be seen on the y-axis. Reconstructed images: $C = 60$ HU, $W = 800$ HU. Difference images: $C = 0$ HU, $W = 250$ HU.

Preliminary investigation of motion-artifact suppression in image-guided radiation therapy of prostate cancer

Zheng Zhang, Bulent Aydogan, Erik Pearson, Dan Xia, Emil Y. Sidky, and Xiaochuan Pan

***Abstract*—Patient-organ motion remains a challenge for accurate image reconstruction for cone-beam computed tomography (CBCT) in image-guided radiation therapy. In this work, we investigate an approach to suppressing the motion artifacts in images reconstructed from CBCT data. In the approach, the organ of interest in the subject scanned is assumed to be approximately static over short temporal intervals during data acquisition, and data collected at the projection views over a limited angular range (LAR) within this time interval are affected minimally by the motion. We then seek to reconstruct images from the LAR data collected. As the problem of image reconstruction from LAR data is highly ill-conditioned, we apply an optimization approach to addressing the problem by including both data and image constraints. Specifically, the image reconstruction is formulated as an optimization problem in which, in addition to an ℓ_2 data-fidelity term, the directional total variations (DTVs) of the image are included for effective suppression of LAR artifacts in the reconstruction. Subsequently, we developed and used a primal-dual-based iterative algorithm to solve the optimization problem for achieving image reconstruction from the LAR data. In the study, the proposed algorithm is evaluated with simulated and real data of digital and physical phantoms mimicking imaging conditions of radiation therapy of prostate cancer from the Hypersight imaging system on a Varian Ethos platform. The numerical results in the evaluation study demonstrate that the algorithm proposed has the potential to effectively suppress LAR artifacts, preserving anatomical structures of interest. This work reveals the potential of algorithm development for addressing accurate image reconstruction from LAR data in CBCT imaging in the radiation therapy, and possibly in other CBCT applications.**

I. Introduction

In image-guided radiation therapy, patient-organ motion during cone-beam computed tomography (CBCT) imaging can lead to artifacts, including blurring, streaking, and image-value bias, which can degrade image utility for precision radiation therapy. To suppress motion-induced artifacts in CBCT, a variety of approaches have been investigated, including gated acquisition strategies or registration-based methods [1], [2]. However, these methods typically rely on accurate motion estimation or additional hardware, which can be challenging, especially in the presence of complex and irregular internal organ motion.

In this work, we investigate an approach to suppressing the motion artifacts in images reconstructed from CBCT data. In the approach, the organ of interest in the subject scanned is assumed to be approximately static over short temporal intervals during data acquisition, and data collected at the projection views over a limited angular range (LAR) within this time interval are affected minimally by the motion, where the LAR is significantly smaller than the full-angular range (i.e., 360°.) We then seek to reconstruct images from the LAR data collected.

As the problem of image reconstruction from LAR data is highly ill-conditioned, we apply an optimization approach to addressing the problem by including both data and image constraints. Specifically, the image reconstruction is formulated as an optimization problem in which, in addition to an ℓ_2 data-fidelity term, the directional total variations (DTVs) of the image are included for effective suppression of LAR artifacts in the reconstruction [3], [4]. Subsequently, we developed and used a primal-dual-based iterative algorithm [5], [6] to solve the optimization problem for achieving image reconstruction from the LAR data.

The proposed algorithm is evaluated in the study with simulated and real data of digital and physical phantoms mimicking imaging conditions of radiation therapy of prostate cancer from the Hypersight imaging system on a Varian Ethos platform. The numerical results in the evaluation study demonstrates that the algorithm proposed has the potential to effectively suppress LAR artifacts, preserving anatomical structures of interest. This work reveals the potential of algorithm development for addressing accurate image reconstruction from LAR data in CBCT imaging in the radiation therapy, and possibly in other CBCT applications.

II. Methods and Materials

In this work, we develop the reconstruction algorithm and conduct numerical experiments mimicking the CBCT imaging in radiation therapy of prostate cancer, using simulated data based on a digital patient phantom in which the prostate is assumed to be static during the data acquisition over a LAR, as well as real data acquired from a physical phantom mimicking the prostate.

A. Optimization problem and reconstruction algorithm

We formulate the reconstruction problem from the LAR data as a convex optimization program incorporating a data-ℓ_2-minimization term and image DTV constraint, defined as:

$$\mathbf{f}^\star = \Phi(\mathbf{g}^{[\mathcal{M}]}, \mathcal{H}\mathbf{f}) \quad \text{s.t.} \quad \Psi(\mathbf{f}), \tag{1}$$

Z. Zhang, D. Xia, and E. Y. Sidky are with the Department of Radiology, The University of Chicago, Chicago, IL 60637, USA.

X. Pan is with the Departments of Radiology & Radiation and Cellular Oncology, The University of Chicago, Chicago, IL 60637, USA.

B. Aydogan and Erik Pearson are with the Department of Radiation and Cellular Oncology, The University of Chicago, Chicago, IL 60637, USA.

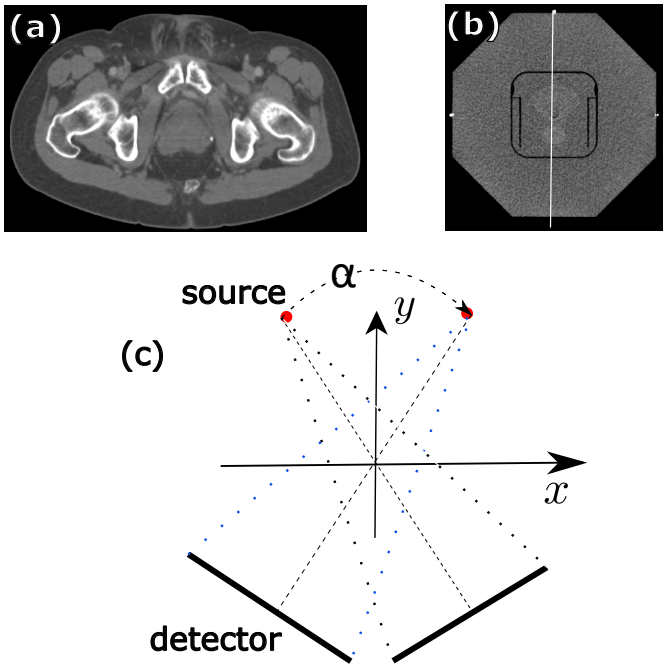


Fig. 1. (a) Digital patient phantom, (b) physical phantom reconstructed by use of FBP algorithm, and (c) illustration of the scanning configuration over LAR α in the studies with simulated data.

where vector $\mathbf{f}$ denotes the image; $\Phi(\mathbf{g}^{[\mathcal{M}]}, \mathcal{H}\mathbf{f})$ represent data ℓ_2 norm between measured data $\mathbf{g}^{[\mathcal{M}]}$ and model data $\mathcal{H}\mathbf{f}$; the matrix $\mathcal{H}$ corresponds to the discrete 2D X-ray transform; $\Psi(\mathbf{f})$ denotes the image constraint, which consists of two terms: the image DTV constraint and the image non-negativity constraint.

We develop the DTV algorithm as an instance of a primal-dual method to solve Eq. (1). The performance of the reconstruction depends on selection of DTV parameters. For simulated data, these parameters are computed based on the ground truth phantom. For real-data reconstruction, the parameters are determined based on evaluation metrics such as visual inspection.

B. Study design and data

1) Simulated data: We conduct initial experiments using simulated data generated from a digital patient phantom (see Fig. 1(a)). In the simulated study, a fan-beam X-ray geometry with a linear detector configuration is used for generating data. The distances from the X-ray source to the rotation center and to the detector are 100 cm and 150 cm, respectively; and the detector consists of 800 bins, each with a width of 0.51 mm. Projection data are acquired at 720 uniformly distributed views over a full angular range of 360°.

The digital patient phantom, which includes a prostate structure, is discretized on a 235×350 image grid with a pixel size of 0.78 mm. Noiseless projection data are generated first for all views. From the full data, LAR data sets are extracted over angular ranges spanning from 90° to 180° (see Fig. 1(c)). Noisy data are obtained by addition of Poisson noise to the noiseless projections, corresponding to an incident photon flux of 5×10^5 photons per ray.

2) Real data: We then acquire real projection data using the Hypersight imaging system on a Varian Ethos platform, employing a physical phantom that mimics a patient's prostate (see Fig. 1(b)). The Hypersight system utilizes a cone-beam X-ray geometry with a flat-panel detector. The distances from the X-ray source to the rotation center and to the detector are 100 cm and 154 cm, respectively. In this study, we extract the detector row corresponding to the central plane and perform two-dimensional image reconstruction. The central detector row consists of 3068 bins with a bin width of 0.28 mm. Full projection data are acquired from 618 views over a total angular range of 211°.

Full-angular-range projection data of the physical phantom are first collected. From these data, LAR subsets are extracted with angular range varying from 90° to 180°. Image reconstruction is performed on a 350×550 grid with a pixel size of 1.05 mm.

III. Results

We first numerically validate the DTV algorithm and its implementation by demonstrating that reconstruction from noiseless full-angular-range data is quantitatively identical to the ground truth up to computer floating-point precision. Following this validation, the DTV algorithm is applied to investigate image reconstruction from LAR data in both simulated- and real-data studies.

A. Study with simulated noiseless data

We first apply the DTV algorithm to reconstruct images from simulated noiseless data acquired with angular ranges of 360°, 150°, 120°, and 90°, respectively. The reconstructed images are shown in Fig. 2. For all LAR cases, the DTV reconstructions exhibit no visible artifacts when compared with the ground truth shown in Fig. 1(a). For comparison, Fig. 2 also presents the corresponding FBP reconstructions obtained from the same projection data. In contrast to the DTV results, the FBP reconstructions exhibit significant streak artifacts and noticeable bias in reconstructed image values when the angular coverage is limited.

B. Study with simulated noisy data

We then apply the DTV algorithm to reconstruct images from noisy projection data acquired with angular spans of 360°, 150°, 120°, and 90°, respectively. The reconstructed images are shown in Fig. 3. Minor artifacts are observed in the DTV reconstructions, particularly for the most severe LAR case with an angular range of 90°. For comparison, the FBP algorithm is applied to the same LAR data, and the corresponding reconstructions are shown in the bottom row of Fig. 3. It is worth noting that, the prostate can be visually identified and delineated for all DTV reconstructions. In contrast to the DTV results, the FBP reconstructions exhibit substantial streak artifacts and pronounced image degradation, thereby preventing reliable visualization of the prostate.

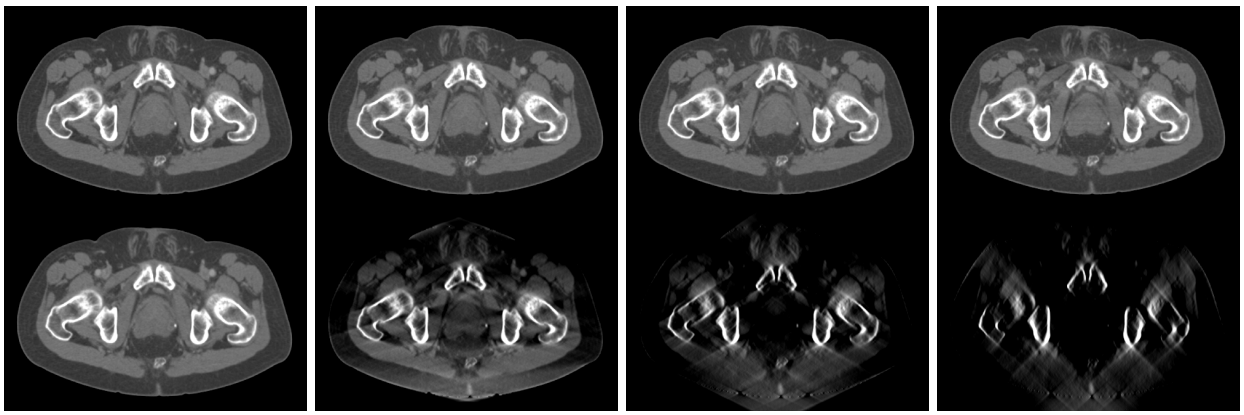


Fig. 2. Images of the digital patient phantom reconstructed by use of the DTV (top row) and FBP (bottom row) algorithms from simulated noiseless data collected over angular ranges of 360° (column 1), 150° (column 2), 120° (column 2), and 90° (column 2). The display window for the DTV reconstructions is [0.18, 0.24] cm^{-1}, while the display windows for the FBP reconstructions are individually adjusted for comparable visualization.

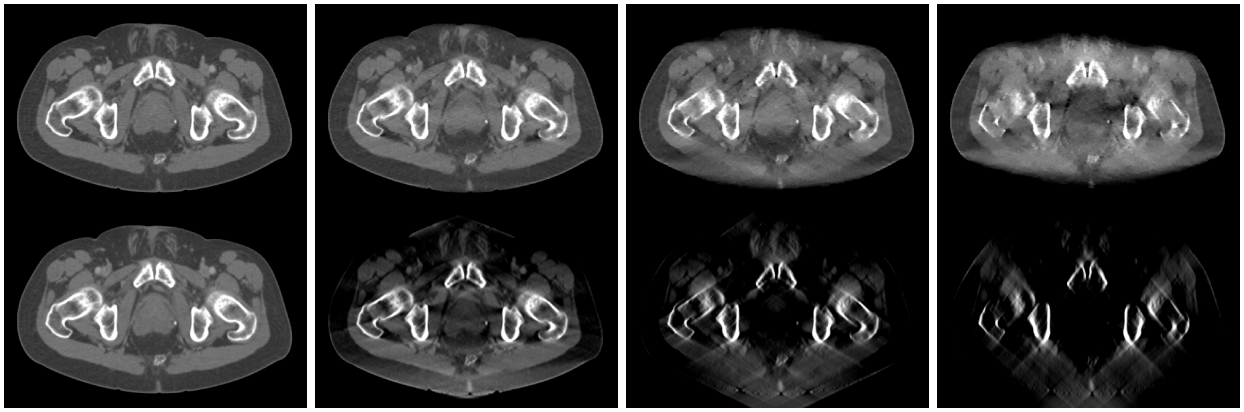


Fig. 3. Images of the digital patient phantom reconstructed by use of the DTV (top row) and FBP (bottom row) algorithms from simulated noisy data collected over angular ranges of 360° (column 1), 150° (column 2), 120° (column 2), and 90° (column 2). The display window for the DTV reconstructions is [0.18, 0.24] cm^{-1}, while the display windows for the FBP reconstructions are individually adjusted for comparable visualization.

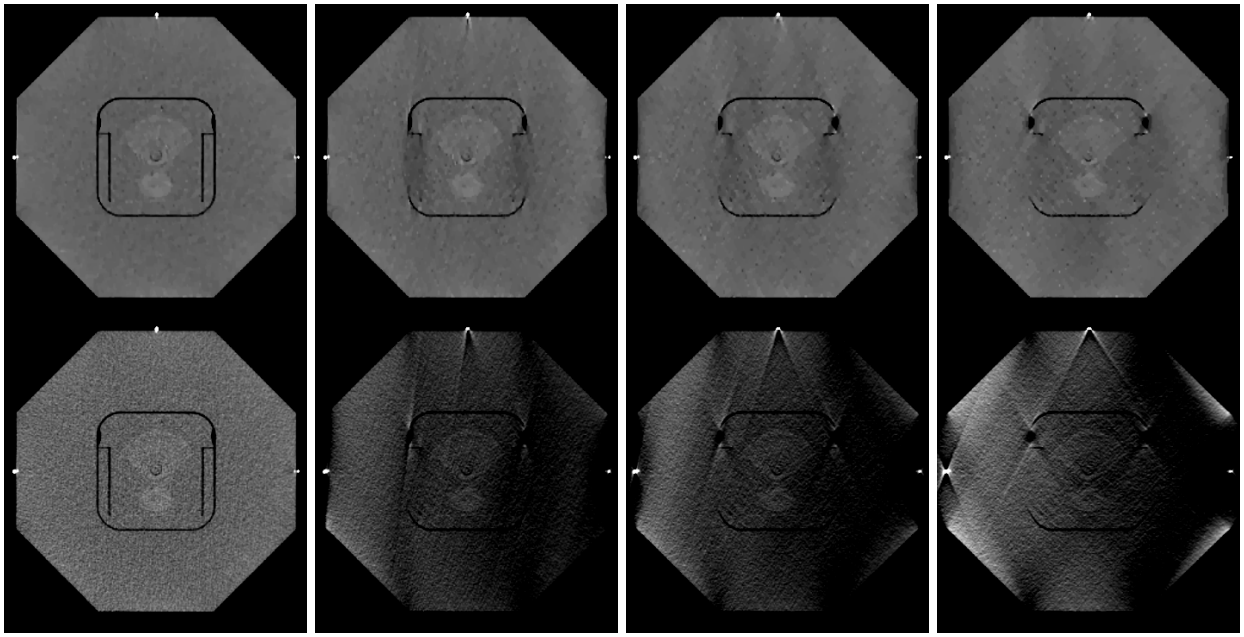


Fig. 4. Images of the physical phantom reconstructed by use of the DTV (top row) and FBP (bottom row) algorithms from real data collected over angular range of 211°, i.e., short scan (column 1), 160° (column 2), 140° (column 3), and 120° (column 4). The display window for the DTV reconstructions is [0.15, 0.25] cm^{-1}, while the display windows for the FBP reconstructions are individually adjusted for comparable visualization.

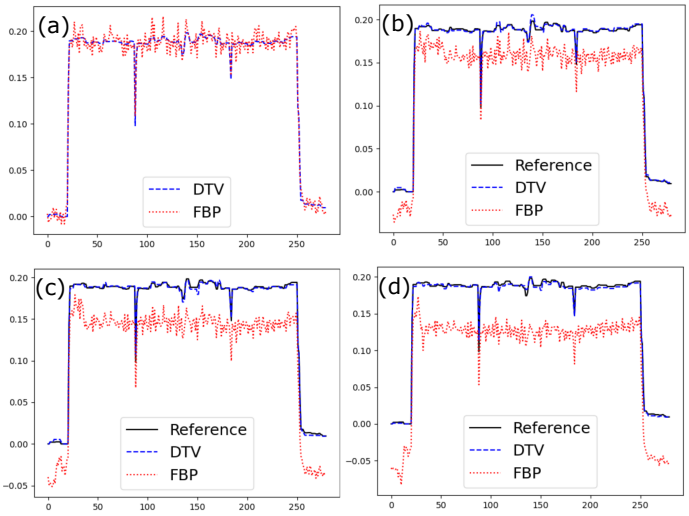


Fig. 5. Image profiles (over the white line overlaid in the physical phantom image in Fig. 1(b)) of the images reconstructed by use of the DTV algorithm (dashed, blue), and the FBP algorithm (dotted, red), from real LAR data collected over angular ranges of 160° (b), 140° (c), and 120° (d), respectively, along with the profiles of the reference image reconstructed from the full data by use of the DTV algorithm.

C. Study with real data

Finally, we perform image reconstruction using real phantom data acquired under LAR configurations with angular ranges of 211°, 160°, 140°, and 120°, respectively. The reconstructed images are shown in Fig. 4. Minor artifacts are observed in the DTV reconstructions, particularly for the lowest LAR of 120°. However, the simulated prostate remains clearly visible in all DTV reconstructions. For comparison, FBP reconstructions from the same LAR data are presented in the bottom row of Fig. 4. In contrast to the DTV results, the FBP reconstructions exhibit pronounced streak artifacts and substantial image degradation.

To further assess reconstruction accuracy, line profiles are plotted in Fig. 5 along the white line indicated in Fig. 1(b). These profiles further demonstrate the accuracy and robustness of the DTV reconstruction.

IV. Discussion

In practical applications, the organ of interest within the subject scanned may be static approximately over short temporal intervals during data acquisition, yielding LAR data that are affected minimally by the motion. In the work, we investigate an approach to minimizing the motion artifacts by reconstruction of the images only from the LAR data collected, and develop and evaluate the DTV algorithm for enabling the approach by using LAR data relevant to CBCT imaging in radiation therapy. Performing numerical experiments using both simulated and real phantom data, we demonstrate that the DTV algorithm proposed has the potential to reconstruct accurate images from LAR data, thus enabling an approach to suppressing motion-induced artifacts in CBCT imaging. The DTV constraints play a key role in ensuring reconstruction accuracy and effectiveness. Future work will focus on optimizing the DTV constraint parameters and extending the proposed framework to real patient data with motion.

We note that patient motion in CBCT often occurs internally and is confined to specific regions, such as gastrointestinal motion, while surrounding anatomical structures remain relatively static throughout the scan. In such scenarios, regions affected by motion may be illuminated over an LAR, whereas regions outside the area effectively receive FAR data. Therefore, it is of interest to investigate region-specific reconstruction strategies that account for spatially varying angular coverage.

Acknowledgments

This work was supported in part by NIH Grant R01CA287302 and by a Varian-UC research grant (Varian Medical Systems (Palo Alto, CA)). The contents of this article are solely the responsibility of the authors and do not necessarily represent the official NIH views.

References

[1] C. J. Ritchie, J. Hsieh, M. F. Gard, J. D. Godwin, Y. Kim, and C. R. Crawford, "Predictive respiratory gating: a new method to reduce motion artifacts on CT scans," *Radiology*, vol. 190, no. 3, pp. 847–852, 1994.

[2] B. Kostyrko, K. Rubarth, C. Althoff, M. Zibell, C. A. Neizert, F. Poch, G. F. Torsello, B. Gebauer, K. Lehmann, S. M. Niehues *et al.*, "Evaluation of different registration algorithms to reduce motion artifacts in CT-thermography (CTT)," *Diagnostics*, vol. 13, no. 12, p. 2076, 2023.

[3] Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan, "Directional-TV algorithm for image reconstruction from limited-angular-range data," *Medical Image Analysis*, p. 102030, 2021.

[4] ——, "Image reconstruction from data over two orthogonal arcs of limited-angular ranges," *Med. Phys.*, pp. –In Press, 2022.

[5] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, pp. 1 – 26, 2011.

[6] E. Y. Sidky, J. H. Jørgensen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Phys. Med. Biol.*, vol. 57, no. 10, pp. 3065–3091, 2012.

Dynamic CT Chest Respiratory Motion Phantom with Lifelike 3D-Printed Deformable Lungs

Jessica Y. Im, Michael Geagan, Sebastian Meyer, Kai Mei, Peter B. Noël

Abstract— **Respiratory motion compensation technologies for computed tomography (CT) should be evaluated with realistic respiratory motion phantoms (RMPs) during quality assurance testing. However, most existing RMPs are overly simplistic: lacking realistic features and limited to rigid motion in a single axis. This work describes the development of a patient-based deformable chest RMP that exhibits realistic lung structures and a dynamic multi-axis motion mechanism. The lungs, mediastinum, chest wall, and diaphragm were separately 3D-printed based on a patient's 4DCT. The lungs were 3D-printed with a flexible thermoplastic polyurethane (TPU) using *PixelPrint* software, which enabled accurate replication of CT texture and attenuation. A motorized superior–inferior (SI) and anterior–posterior (AP) compression device was developed and programmed to operate in two modes: (1) a static mode that sets and maintains discrete compression levels, and (2) a dynamic mode that simulates continuous breathing. Scans of the RMP were obtained at compression levels matched to each phase of the patient's 4DCT. Structural similarity index (SSIM) between the patient and phantom images was > 0.9. Furthermore, deformation vectors of the RMP closely matched those of the patient, with mean displacement errors of -1.7 ± 4.2 mm and -0.7 ± 1.7 mm in the SI and AP directions, respectively. Finally, scans were obtained with the RMP in dynamic mode to mimic free-breathing scans. Images exhibited motion artifacts near the diaphragm. In conclusion, this work presents a dynamic CT chest RMP with lifelike deformable lungs, which can serve as a quality assurance testing device for motion compensation technologies.**

Index Terms— **Respiratory motion phantom, Deformable lung phantom, 3D-printed phantom, CT phantom, 4DCT**

I. Introduction

Patient respiratory motion in medical imaging often results in image artifacts which reduce diagnostic confidence and quantitative accuracy. Respiratory motion also complicates radiation therapy: treatment plans must account for movement of the tumor volume. Many techniques have been developed to address respiratory motion in diagnostic imaging and radiation therapy, including motion artifact reduction algorithms [1], deformable image registration [2], and tumor tracking techniques [3]. It is necessary to evaluate and validate the effectiveness of these techniques both prior to clinical use and regularly during routine validation. To avoid unnecessary radiation exposure to patients and human subjects during such validation experiments, respiratory motion phantoms (RMPs) have been developed as quality assurance testing environments for computed tomography (CT) motion compensation technique development and evaluation.

Previously developed RMPs include a variety of design paradigms. Some are composed of rigid components which undergo bulk motion [1, 3] or slide relative to each other [4, 5], while others utilize compressible foam or 3D-printed phantoms to mimic deformable lungs[6-9]. However, most of these examples have simplified geometries, lacking the complex textures and fine structures which are found in patient lungs. Furthermore, the motion profiles of these RMPs are simplified, some demonstrating only rigid translation and some only incorporating a single axis of compression, which is contrary to the complex and multi-axial motion of patient respiration. The RMPs which come the closest to replicating true clinical scenarios are *ex vivo* phantoms [10], but these are not suitable for repeatable long term use due to tissue degradation and lack of reproducibility. Finally, digital motion phantoms have been developed to simulate breathing motion *in silico* [11], but are limited to use in the assessment of software-based motion compensation and cannot be used in the direct testing and validation of physical imaging systems.

To address these limitations, we previously introduced *PixelPrint*^{4D}, a method of 3D printing deformable lung phantoms [12]. The patient-based *PixelPrint*^{4D} lung phantom not only demonstrated voxel-by-voxel precision in textural and structural accuracy within the lung volume, but also lifelike deformation properties including deformation vector fields (DVFs), attenuation changes, and Jacobian measurements. In this work, we expanded upon *PixelPrint*^{4D} to develop a full deformable chest RMP prototype with dynamic multi-axis breathing motion.

J. Y. Im is with the Department of Bioengineering, University of Pennsylvania, PA, USA and Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. M. Geagan, K. Mei, and P. B. Noël are with the Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. S. Meyer is with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. (e-mail: imjess@seas.upenn.edu).

TABLE I
IMAGE ACQUISITION AND RECONSTRUCTION PARAMETERS

	Patient 4DCT scan	Phantom scans
CT Scanner	SOMATOM Drive (Siemens Healthineers)	Spectral CT 7500 (Philips Healthcare)
Tube voltage [kVp]	120	120
$CTDI_{vol}$ [mGy]	31.0	31.1
Scan mode	Helical	Helical
Pitch	0.14	1.0
Rotation time [s]	0.14	1.0
Collimation	64 x 0.6 mm	64 x 0.625 mm
Slice thickness [mm]	1	0.67
Field of view [mm]	400	200
Matrix size [pixels]	512 x 512	512 x 512
Pixel spacing [mm]	0.7812	0.3906
Reconstruction filter	Br32	B

II. METHODS

A. Phantom Design

The phantom was designed based on a patient chest 4DCT scan and manufactured at 0.5x scale (Fig. 1). Segmentations of the lungs, mediastinum, chest wall, and diaphragm/liver were obtained from the end inhalation phase image of the patient 4DCT. Each component was fabricated separately. First, 3D-printer instructions were generated for the lungs using *PixelPrint* software[13, 14], which converted CT attenuation data from each voxel into a target printed infill ratio [12]. The target infill was continuously varied to mimic high resolution patient attenuation profiles. The lung phantoms were printed out of a flexible 3D-printing material, thermoplastic polyurethane (TPU) (Ninjaflex (85A), Ninjatek, PA, USA) using fused deposition modeling (FDM) 3D-printing technology. Polylactic acid (PLA) was used to print support material below the lung and was later removed. Next, the segmentations of the mediastinum, chest wall, and diaphragm/liver were converted into standard tessellation language (STL) file format. The mediastinum was 3D-printed using material jetting 3D-printing technology with the "liver – moderately stiff" material blend (J850 Digital Anatomy 3D Printer, Stratasys, MN, USA). The majority of the chest wall was printed with a rigid PLA material (Bright White PLA Filament, Atomic Filament, IN, USA) at 86% infill to achieve a soft-tissue-like attenuation. To enable the chest wall to deform, a ~15 mm coronal slab of the chest wall was 3D-printed using flexible TPU at a low infill of 10%. Finally, the diaphragm/liver was 3D-printed using PLA with an infill of 90% to mimic liver attenuation. The lungs, chest wall, and diaphragm were 3D-printed using an FDM 3D-printer (Original PrusaXL 2-toolhead, Prusa Research, Czech Republic).

B. Compression device design

The RMP was designed to mimic both superior-inferior (SI) and anterior-posterior (AP) compression of the lungs and mediastinum while minimizing the presence of unnatural

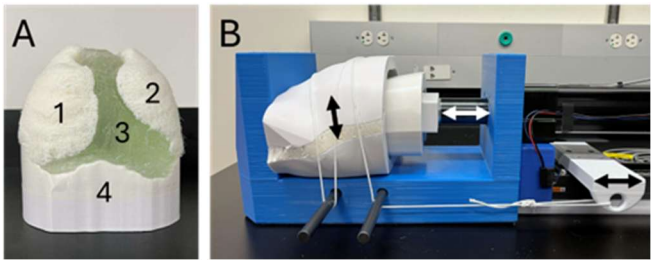


Fig. 1. RMP design. (A) 3D-printed phantom components including right lung (1), left lung (2), mediastinum (3), and diaphragm/liver (4) which were fitted into the chest wall shell. (B) The assembled RMP consisted of an SI motion mechanism (white arrow) and AP motion mechanism (black arrows).

structures, especially metal components, within the expected scan range (Fig. 1B). SI compression was achieved by driving the phantom diaphragm/liver with an electric rod actuator (RSA32 Electric Rod Actuator, Tolomatic, MN, USA). AP compression was simulated by a cinching mechanism whereby a high tensile-strength ultra-low-stretch synthetic cable was laced across the anterior chest wall and attached at both ends to a positioning slide. The cable was pulled by the slide to cinch down on the chest wall. Additional mounts and bases for the phantom and actuators were 3D-printed with PLA and polyethylene terephthalates glycol (PETG) using FDM 3D-printers. Both the actuator and positioning slide were driven by Nema 23 stepper motors with DM542T stepper drivers (StepperOnline, Nanjing, China) and controlled by an Arduino Uno R3 (Arduino, Monza, Italy). Limit switches were used to ensure repeatable positioning of both SI and AP motion mechanisms, and a homing step was included in the motion procedures. The RMP was programmed with two compression modes: static and dynamic. The static mode allowed the RMP to be driven to any specified displacement level and hold the position while a scan is performed. This allowed simulation of different respiratory phases of a 4DCT series when scanned in a "step and shoot" manner. The dynamic mode drove the RMP continuously in a sinusoidal trace and was used to simulate free-breathing scans with various respiratory rates or amplitudes.

C. Image Acquisition

The RMP was scanned using a dual-layer CT (Spectral CT 7500, Philips Healthcare, OH, USA). Acquisition parameters were chosen to mimic the patient's CT scan settings (Table I). Since the phantom was fabricated at 0.5x scale, the reconstruction field of view was set to half that of the patient scan, and the slice thickness was set to the minimum allowed on the scanner. Pitch was set to 1.0 and rotation time to 1.0 s to simulate a free-breathing CT used for radiotherapy planning.

To generate pseudo-4DCT phantom data, SI displacements of the patient's diaphragm and AP displacements of the patient's anterior chest wall were measured from the patient's 4DCT at each phase of the respiratory cycle. Each

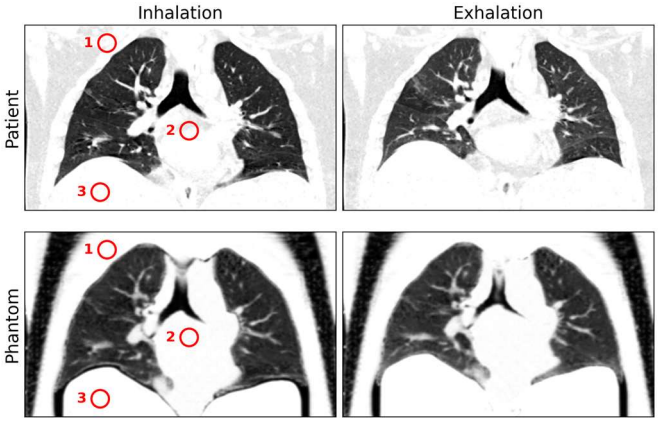


Fig. 2. CT scan images of the patient (top) and phantom (bottom) at end inhalation (left) and end exhalation (right). ROIs used for measuring mean attenuation in the chest wall (1), mediastinum (2), and liver (3) are indicated with red circles in the end inhalation phase images. Note that phantom images are shown at 2x scale compared to the patient (WL/WW: - 500/1000 HU).

displacement was programmed into the RMP at 0.5x scale. The RMP was then driven in static mode to each position, and a CT scan was acquired at each phase.

To match the patient's 4DCT motion at 0.5× scale, the RMP was driven in dynamic mode to generate free-breathing phantom CT images, with SI and AP motion amplitudes set to 8.6 mm and 1.8 mm, respectively. Free-breathing CTs were acquired with the RMP set at 12 and 20 breaths per minute (bpm) to mimic typical and fast respiratory rates, respectively.

D. Image Analysis

First, the attenuation and structural accuracy of the phantom were assessed. Lung segmentations were performed using an automated deep-learning tool [15]. The mean attenuation within the lung segmentation was measured at end inhalation and end exhalation in both the patient's 4DCT and phantom pseudo-4DCT images. Next, 1 cm radius circular ROIs (Fig. 2) were placed to measure mean attenuation and noise (standard deviation) in the chest wall, mediastinum, and liver of the patient and phantom images at the end inhalation phase. Additionally, the structural similarity index measure (SSIM) was calculated between the patient and phantom images using the scikit-image implementation. The images were masked to include only the segmented lung region for the SSIM calculations. Rigid registration was performed on phantom data to ensure alignment with patient data during data processing [16].

TABLE II
MEAN ATTENUATION MEASUREMENTS

Anatomical region	Attenuation [HU]	
	Patient	Phantom
Lung (inhale)	-702 ± 177	-706 ± 163
Lung (exhale)	-652 ± 176	-670 ± 158
Chest wall	0 ± 45	30 ± 20
Mediastinum	-21 ± 60	-8 ± 6
Liver	38 ± 34	38 ± 7

To quantify motion accuracy, deformable image registration (DIR) [17] was performed on both patient 4DCT and phantom pseudo-4DCT data to obtain DVFs, which indicated the displacement that each voxel underwent in each phase. All DVFs were calculated with respect to the end inhalation phase of the respiratory cycle. The SI displacements of the diaphragm were then measured at the apex of the right diaphragm in both the patient 4DCT and phantom pseudo-4DCT images, across all phases of the respiratory cycle. AP displacements were also measured at the anterior chest wall in the same manner. The mean displacement errors were calculated across the respiratory cycle for the diaphragm and chest wall. In addition, the mean vector errors throughout the full lung volume, defined by the lung segmentation volumes, were also calculated for both the SI and AP directions. In all analyses, the displacements from the phantom were doubled to compensate for the 0.5x scaling of the phantom. Finally, the free-breathing scans at 12 bpm and 20 bpm were qualitatively compared to the static scans to identify and assess motion artifacts.

III. RESULTS

A video showing the RMP in motion alongside the pseudo-4DCT scan images can be viewed at the following link: https://youtu.be/fKBO-HFJF7E. Fig. 2 shows representative coronal images of the patient and phantom at end inhalation and end exhalation phases with ROIs marked where attenuation data was collected. Mean attenuation of the fully segmented lungs at end inhalation and end exhalation, and different anatomic regions (Fig. 2) at end inhalation are listed in Table II. The phantom lung attenuation at end inhalation (-706 ± 163 HU) closely matched the patient lung attenuation (-702 ± 177 HU) since the lung attenuations at this phase were explicitly mapped to printed infill levels during the *PixelPrint* printing process. The change in attenuation between end inhalation and end exhalation in the phantom lungs (~36 HU), however, is less than the change in attenuation observed in the patient lungs (~50 HU). The SSIM comparing the patient and phantom lung volumes at end inhalation and end exhalation were 0.91 and 0.93, respectively, indicating that the phantom lungs closely emulated the structural details in the patient lungs.

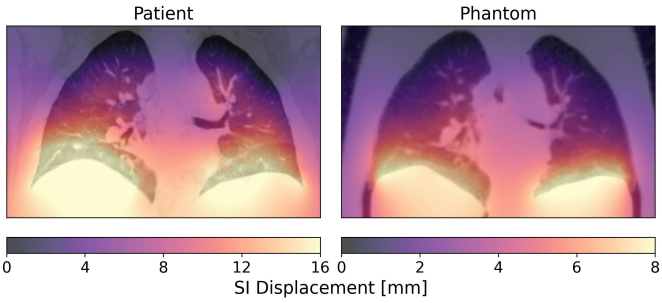


Fig. 3. SI displacement vectors from the end exhalation phase overlaid on coronal CT images of the patient and phantom. Note that the phantom displacement scale bar is halved compared to the patient displacement scale bar due to the 0.5x scaling. (Base CT image WL/WW: - 500/1000 HU).

Fig. 3 shows the SI-component of the DVFs at end exhalation overlaid on the patient and phantom images as a heatmap, which visually demonstrates the similarity in motion patterns throughout the image. The displacements of the diaphragm and chest wall at each phase of the respiratory cycle are shown in Fig. 4, where the phantom displacements (doubled to compensate for the 0.5x scale) closely follow the path of the patient displacements. The mean errors in diaphragm SI motion and chest wall AP motion measured at the indicated locations (Fig. 4) were -0.3 ± 0.5 mm and 0.1 ± 0.2 mm, respectively, across the respiratory cycle. Notably, these errors are less than the dimensions of a single voxel in the patient image (0.7812 x 0.7812 x 1 mm). The mean displacement errors throughout the full lung volume at end exhalation were -1.7 ± 4.2 mm and -0.7 ± 1.7 mm in the SI and AP directions, respectively.

The free-breathing scans showed mild to moderate motion artifacts near the diaphragm. Compared to the static scan, the inferior border of the lung phantom appears blurry and thickened in the 12 bpm scan, while there is a clear doubling artifact of the bottom border of the lung phantom in the 20 bpm scan (Fig. 5).

IV. Discussion

This work presents a prototype chest RMP that includes lifelike deformable 3D printed lungs. The RMP can apply simultaneous SI and AP compression to the lungs and mediastinum, mimicking human respiratory motion, and can be driven in either a static or a dynamic free-breathing mode. This RMP demonstrated accurate lung details on CT imaging, with SSIM of 0.91 and 0.93 compared to the patient reference image, and realistic tissue-like attenuation in structures including the mediastinum, chest wall, and liver. Moreover, motion vector fields closely matched those seen in the reference patient 4DCT, with diaphragm and chest-wall errors

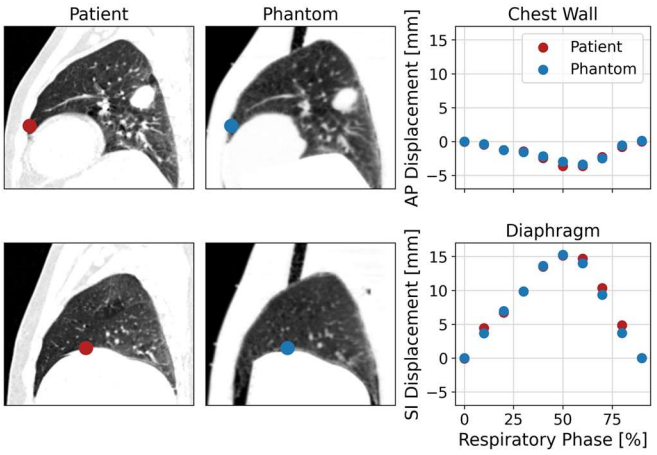


Fig. 4. Chest wall and diaphragm displacements. The AP displacements of the chest wall and the SI displacements of the diaphragm are plotted across the different phases of the respiratory cycle for both the patient and phantom. Phantom displacements were doubled to account for the 0.5x fabrication scale. Locations where the displacements were measured are indicated with circular markers in the sagittal view CT scan images shown on the left (WL/WW: - 500/1000 HU).

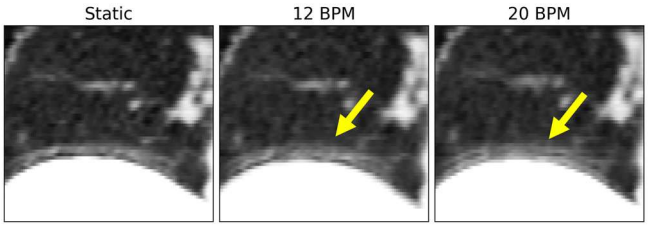


Fig. 5. Comparison of motion artifacts. CT scan images of a static scan and 2 free breathing scans at different breathing rates: 12 and 20 bpm. Motion artifacts indicated by the yellow arrows can be appreciated near the diaphragm in both the free breathing scans, showing blurring and doubling of the bottom border of the lung phantom (WL/WW: - 500/1000 HU).

smaller than one voxel of the patient image. Images taken in the free-breathing dynamic mode at different respiratory rates demonstrated varying degrees of motion artifacts.

There are a few limitations to this study. First, this initial prototype was fabricated at half scale, thus limiting the RMP's ability to simulate realistic patient sizes and resulting in some loss of finer lung details. Future iterations will be fabricated at full scale. Another limitation is that the chest wall shell contained a low-density slab of TPU to enable AP compression, resulting in an unnatural low attenuation region. Different printing materials will be investigated to find a solution which readily allows for compression while maintaining tissue-like attenuation. Additionally, the mediastinum, diaphragm, and chest wall were printed with uniform infill and thus did not include realistic attenuation variations for different tissue types. Future iterations will include more detailed differentiation of tissue types, including bone, muscle, and fat, which can be printed with a combination of calcium-doped filaments and PLA. In addition, the motion analyses in this study were primarily based on DIR outputs, and thus were susceptible to inaccuracies introduced by the registration algorithm. Future studies will include landmark and feature-based displacement analyses and comparisons of organ boundaries via metrics such as mean surface distance to present a more comprehensive motion analysis. Finally, this study only investigated regular sinusoidal breathing patterns for the free-breathing dynamic mode, while realistic human respiration is more complex and irregular. Our next steps involve incorporating realistic patient-based breathing patterns.

In summary, we developed a respiratory motion phantom (RMP) that reproduces lung anatomy and deformation patterns with greater physiological realism than prior phantom technologies. This enhanced realism enables more clinically relevant and quantitative benchmarking of CT respiratory motion compensation methods, supporting both preclinical validation and routine performance assessment. Such benchmarking is critical for both radiation oncology - where motion management affects target delineation and localization as well as dose delivery - and diagnostic imaging, where respiratory motion can degrade image quality and quantitative accuracy.

ACKNOWLEDGMENT

The authors acknowledge support through the National Institutes of Health (R01EB035092, R01HL166236) and thank Dr. Boon-Keng Kevin Teo for selecting the reference patient 4DCT for this study.

REFERENCES

[1] D. Kim *et al.*, "Motion correction for routine X-ray lung CT imaging," *Sci Rep,* vol. 11, no. 1, p. 3695, Feb 12 2021, doi: 10.1038/s41598-021-83403-w.

[2] S. Miyakawa, H. Tachibana, S. Moriya, T. Kurosawa, T. Nishio, and M. Sato, "Design and development of a nonrigid phantom for the quantitative evaluation of DIR-based mapping of simulated pulmonary ventilation," *Med Phys,* May 28 2018, doi: 10.1002/mp.13017.

[3] J. Jung *et al.*, "Verification of Accuracy of CyberKnife Tumor-tracking Radiation Therapy Using Patient-specific Lung Phantoms," *Int J Radiat Oncol Biol Phys,* vol. 92, no. 4, pp. 745–53, Jul 15 2015, doi: 10.1016/j.ijrobp.2015.02.055.

[4] P. Steidl *et al.*, "A breathing thorax phantom with independently programmable 6D tumour motion for dosimetric measurements in radiation therapy," *Phys Med Biol,* vol. 57, no. 8, pp. 2235–50, Apr 21 2012, doi: 10.1088/0031-9155/57/8/2235.

[5] R. K. Chaudhary, R. Kumar, S. D. Sharma, S. Bera, V. Mittal, and S. Deshpande, "Performance Validation of In-House Developed Four-dimensional Dynamic Phantom," *J Med Phys,* vol. 44, no. 2, pp. 99–105, Apr–Jun 2019, doi: 10.4103/jmp.JMP_114_18.

[6] J. Laidlaw *et al.*, "Design and CT imaging of casper, an anthropomorphic breathing thorax phantom," (in English), *Biomed Phys Eng Expr,* vol. 9, no. 2, Mar 1 2023, doi: 10.1088/2057-1976/acb7f7.

[7] D. S. Shin *et al.*, "Development of a deformable lung phantom with 3D-printed flexible airways," *Med Phys,* vol. 47, no. 3, pp. 898–908, Mar 2020, doi: 10.1002/mp.13982.

[8] D. Lustermans *et al.*, "Development of a novel 3D-printed dynamic anthropomorphic thorax phantom for evaluation of four-dimensional computed tomography," *Physics and Imaging in Radiation Oncology,* vol. 32, p. 100656, 2024/10/01/ 2024, doi: https://doi.org/10.1016/j.phro.2024.100656.

[9] S. Abdollahi *et al.*, "Dynamic anthropomorphic thorax phantom for quality assurance of motion management in radiotherapy," (in eng), *Phys Imaging Radiat Oncol,* vol. 30, p. 100587, Apr 2024, doi: 10.1016/j.phro.2024.100587.

[10] J. Biederer *et al.*, "Reproducible Simulation of Respiratory Motion in Porcine Lung Explants," *Rofo,* vol. 178, no. 11, pp. 1067–72, Nov 2006, doi: 10.1055/s-2006-927149.

[11] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4D XCAT phantom for multimodality imaging research," *Medical physics,* vol. 37, no. 9, pp. 4902–4915, 2010.

[12] J. Y. Im *et al.*, "PixelPrint4D: A 3D Printing Method of Fabricating Patient-Specific Deformable CT Phantoms for Respiratory Motion Applications," *Investigative Radiology,* April 2 2025, doi: 10.1097/rli.0000000000001182.

[13] K. Mei *et al.*, "Three-dimensional printing of patient-specific lung phantoms for CT imaging: Emulating lung tissue with accurate attenuation profiles and textures," *Med Phys,* vol. 49, no. 2, pp. 825–835, Feb 2022, doi: 10.1002/mp.15407.

[14] N. Shapira *et al.*, "Three-dimensional printing of patient-specific computed tomography lung phantoms: a reader study," *PNAS Nexus,* vol. 2, no. 3, p. pgad026, Mar 2023, doi: 10.1093/pnasnexus/pgad026.

[15] J. Hofmanninger, F. Prayer, J. Pan, S. Rohrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *Eur Radiol Exp,* vol. 4, no. 1, p. 50, Aug 20 2020, doi: 10.1186/s41747-020-00173-2.

[16] P. A. Yushkevich *et al.*, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *Neuroimage,* vol. 31, no. 3, pp. 1116–28, Jul 1 2006, doi: 10.1016/j.neuroimage.2006.01.015.

[17] S. Kabus and C. Lorenz, "Fast elastic image registration," *Proceedings of the Medical Image Analysis For The Clinic: A Grand Challenge, MICCAI, Beijing, China,* pp. 81–89, 2010.

Tunable Metal-Artifact CT Phantoms via Dual-Filament 3D Printing

Pouyan Pasyar[1,2], Kai Mei[1], Jessica Y. Im[1,2], Leonid, Roshkovan[1], Michael Geagan[1], Peter B. Noël[1]

Abstract— **Metal artifacts from orthopedic implants remain a persistent challenge in Computed Tomography (CT), degrading image quality and obscuring adjacent anatomy. Evaluation of metal artifact reduction (MAR) algorithms is limited by the scarcity of anatomically realistic phantoms with controlled implant properties for quantitative benchmarking. We present a dual-filament *PixelPrint* method to fabricate patient-specific phantoms with tunable metal artifact severity by varying printed implant density. Using clinical CT data from a bone-healthy patient, we fabricated lifelike cervical spine phantoms featuring six orthopedic screws spanning C4–C6, with screw geometry defined from CAD models. Three implant configurations were manufactured using calcium-doped (tissue/bone surrogate) and metal-doped PLA filaments: 0% (no metal), 50% (medium), and 85% (high) infill. Phantoms were scanned at a tube voltage of 120 kVp and current of 1000 mAs, then reconstructed as virtual monoenergetic images from 50 to 190 keV. The printed phantoms reproduced vertebral anatomy, screw geometry, and soft-tissue contrast; HU measurements agreed closely with patient values for soft tissue and bone. Artifact severity increased systematically with implant infill and was quantified using a Gumbel-based p-index. Higher-energy monoenergetic reconstructions reduced streaking and improved screw visualization, consistent with findings in real patients. Dual-filament *PixelPrint* can reliably produce anatomically realistic CT phantoms with tunable metal artifact severity, creating a practical testbed for quantitatively evaluating new CT technologies - especially MAR algorithms and spectral CT methods.**

Keywords—3D printing, spectral CT, metal artifact reduction, image quality

I. Introduction

Orthopedic implants introduce severe metal artifacts in computed tomography (CT), primarily due to beam hardening, photon starvation, and scatter effects [1-3]. These artifacts obscure the visualization of surrounding soft tissue, bone, and implant-tissue interfaces, limiting the assessment of postoperative complications such as loosening, infection, fracture healing, or adjacent soft-tissue pathology [2, 4, 5]. In addition, metal artifacts can distort implant geometry and local attenuation values, complicating both qualitative interpretation and quantitative image analysis [3, 6]. Together, these effects significantly reduce diagnostic confidence and hinder reliable clinical evaluation in patients with metallic hardware [4, 7].

Despite substantial advances in reconstruction methods and metal artifact reduction (MAR) techniques, accurate assessment of image quality near metallic implants remains challenging [1-3]. Robust evaluation of MAR algorithms is hindered by the lack of anatomically realistic phantoms with known ground truth, particularly those that simultaneously reproduce patient-specific anatomy and implant geometry [6-8]. Current MAR validation studies often rely on simplified phantoms (e.g., uniform water-equivalent backgrounds with metal rods). These designs do not capture the heterogeneous attenuation and complex anatomy seen in clinical imaging [8-10]. As a result, MAR performance is frequently assessed using metrics that do not reflect realistic implant–tissue interactions, potentially limiting algorithm optimization and generalizability to patient data [7]. These limitations highlight the need for patient-derived, anatomically realistic phantoms that enable controlled, quantitative evaluation of metal artifact behavior under clinically relevant conditions.

Recent advances in three-dimensional (3D) printing have enabled fabrication of patient-specific phantoms. *PixelPrint* [13] is a voxel-based printing framework that directly converts DICOM images into printer instructions, preserving anatomical texture and attenuation without having to perform segmentation or surface meshing to the DICOM input data. Previous *PixelPrint* studies have demonstrated accurate reproduction of lung, soft tissue, and bone attenuation [11-13].

Dual-filament *PixelPrint* [14] prints tissue and implant materials in the same patient-specific phantom, enabling lifelike anatomy with independently tunable metal content. Using calcium-doped PLA for bone/soft tissue and metal-doped PLA for screws, we can vary metal-doped infill to systematically adjust artifact severity while keeping anatomy identical - something single-filament printing cannot provide. Here, we introduce and demonstrate this approach to cervical spine phantoms with screws spanning C4–C6, creating a reproducible platform for quantitative evaluation of CT technology.

II. Materials and Methods

A. Phantom design

Six real-world orthopedic screw CAD models were digitally inserted into the clinical CT dataset from a bone-healthy patient. Bilateral screws were placed in the C4, C5, and C6 vertebrae along clinically realistic trajectories through the pedicles and vertebral bodies, as confirmed by a board-certified radiologist. Three phantom variants were designed to enable controlled evaluation of metal artifacts. First, an artifact-free reference ("ground truth") phantom preserved the screw geometry but used 0% metal infill, yielding identical anatomy without metal artifacts. Second, a 50% metal-infill phantom was chosen to approximate the effective radiological density of clinical titanium screws and produce clinically realistic artifacts. Third, an 85% metal-infill phantom generated severe artifacts to test algorithm performance under challenging conditions.

This study was funded by the National Institutes of Health (R01EB035092, R01EB035908)

P. Pasyar and J. Y. Im, are with the Department of Bioengineering and Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA.
K. Mei, L. Roshkovan, M. Geagan, and P. B. Noël are with the Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA.

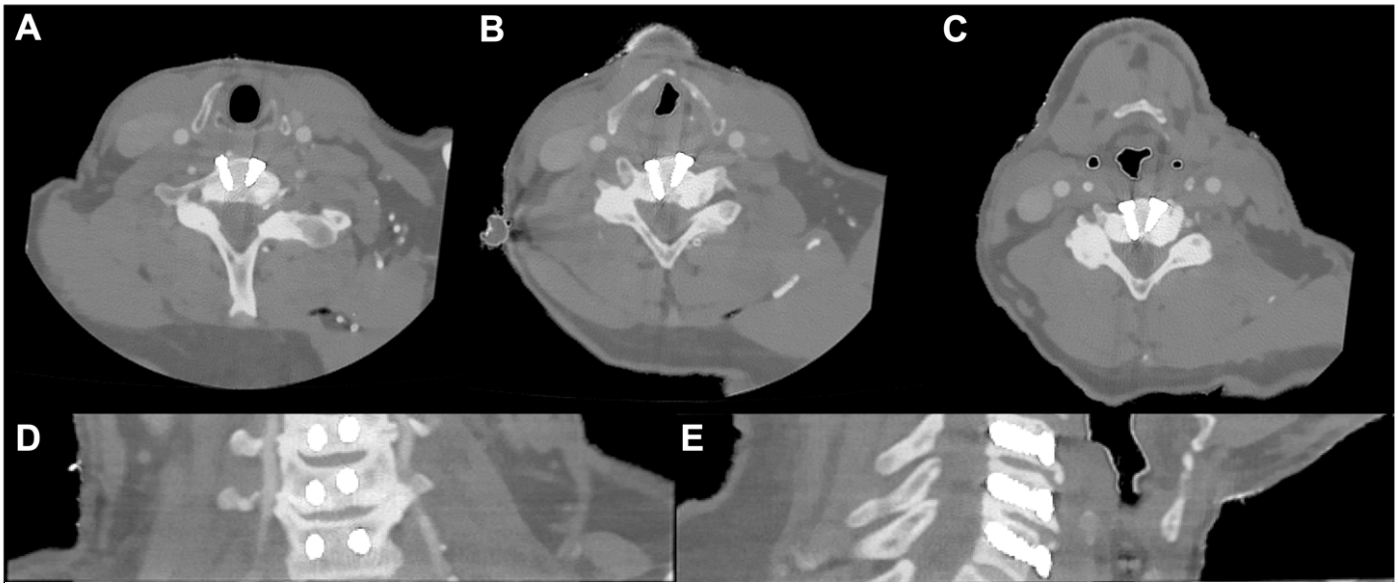


Figure 1. Representative CT images of the *PixelPrint* cervical spine phantom (medium metal infill) acquired at VMI 70 keV without MAR, demonstrating accurate anatomical reproduction of the vertebral column and embedded screws. (A–C) Axial, (D) coronal, and (E) sagittal images through the C4, C5, and C6 vertebrae, depicting the full screw trajectories and preserved cortical and trabecular bone boundaries. WL/WW: 0/1500 HU.

B. Phantom fabrication

Phantoms were fabricated using the *PixelPrint* dual-filament framework, which converts DICOM image data directly into G-code. *PixelPrint* generates two separate 3D arrays corresponding to calcium- and metal-doped PLA. Filament lines were interleaved within each layer using a fixed 0.5 mm spacing, with alternating line orientations between layers to minimize grid artifacts. Calcium-doped PLA filament (custom made filament) was used to represent bone and soft tissue, while metal-doped PLA (Ultrafuse 17-4, BASF, Ludwigshafen, Germany) simulated orthopedic implants. Phantoms were printed on a dual-extruder fused deposition modeling system (Prusa XL 2-toolhead, Prusa, Prague, Czech Republic) with 0.25 mm nozzles. Extrusion temperatures were 210 °C for calcium-doped PLA and 230 °C for metal-doped PLA. We varied filament line widths to modulate local attenuation: calcium-doped PLA ranged from 0.2-0.9 mm, while metal-doped PLA used 0.5 mm (50%) and 0.85 mm (85%) widths for medium and high infill screws, respectively.

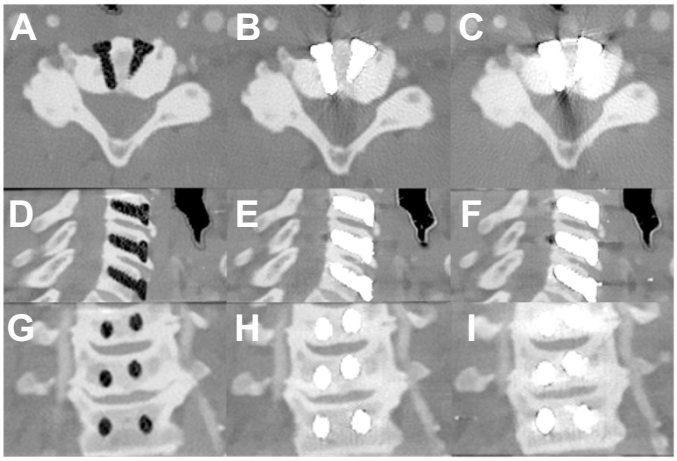


Figure 2. Comparison of three *PixelPrint* cervical spine phantoms imaged at VMI 70 keV without MAR. Axial views at the C6 level for (A) 0%, (B) 50%, and (C) 85% metal infill, respectively. (D–F) Sagittal and (G–I) coronal reconstructions of the same phantoms. Increasing metal infill produces progressively stronger streak and photon-starvation artifacts while anatomical structure and screw positioning remain unchanged. WL/WW: 0/1500 HU.

C. CT Acquisition and Analysis

All phantoms were scanned on a dual-layer spectral CT (Spectral CT 7500, Philips Healthcare, Amsterdam, Netherlands) using helical acquisition at a tube voltage of 120 kVp and exposure of 1000 mAs. In addition to conventional images, virtual monoenergetic images (VMIs) were reconstructed at 50, 70, 90, 110, 150 and 190 keV. Reconstructions were performed without vendor-specific MAR to retain artifact behavior. Quantitative evaluation was performed to assess both attenuation accuracy and metal artifact severity. First, region-of-interest (ROI)-based Hounsfield Unit (HU) analysis was used to evaluate the fidelity of *PixelPrint* attenuation reproduction. Circular ROIs were placed in representative soft-tissue and bone regions throughout the phantom to sample anatomically distinct areas with different expected attenuation values. These ROIs enabled direct comparison of mean HU and standard deviation between the printed phantom and the reference patient data, providing quantitative assessment of tissue-specific attenuation accuracy and image noise characteristics.

Metal artifact severity was then quantified using a Gumbel-based extreme-value analysis framework adapted from Cammin [15]. Rectangular ROIs were placed in regions with artifact adjacent to, but excluding, the implanted screws, while corresponding background ROIs were selected in homogeneous regions unaffected by metal artifacts. For each ROI, we extracted intensity profiles perpendicular to the dominant streak direction and computed the maximum absolute difference between adjacent pixels along each profile. The resulting extreme-value distributions were then fit with Gumbel models to estimate the location parameter, which quantifies artifact-

driven intensity functions This metric enables objective, quantitative comparisons of streaking and photon-starvation artifacts across phantoms with different metal infill levels and reconstruction settings.

III. Results

Representative 70 keV VMI images of the medium-metal phantom show lifelike vertebral anatomy, screw geometry, and soft-tissue contrast (Figure 1), with well-defined cortical margins, visible trabecular structure, and screws aligned with the intended trajectories. ROI-based HU measurements agree closely with patient data in soft tissue and moderate-density bone, with mean differences within ±23 HU. The largest discrepancy occurs in >1000 HU cortical bone, consistent with limits of current calcium-doped PLA, while overall agreement remains strong for tissues relevant to artifact evaluation. Across phantoms reconstructed at 70 keV without MAR, streaking and beam-hardening artifacts increase stepwise with metal infill: the ground-truth phantom remains uniform, while the medium- and high-metal phantoms show progressively greater photon starvation near the screws (Figure 2).

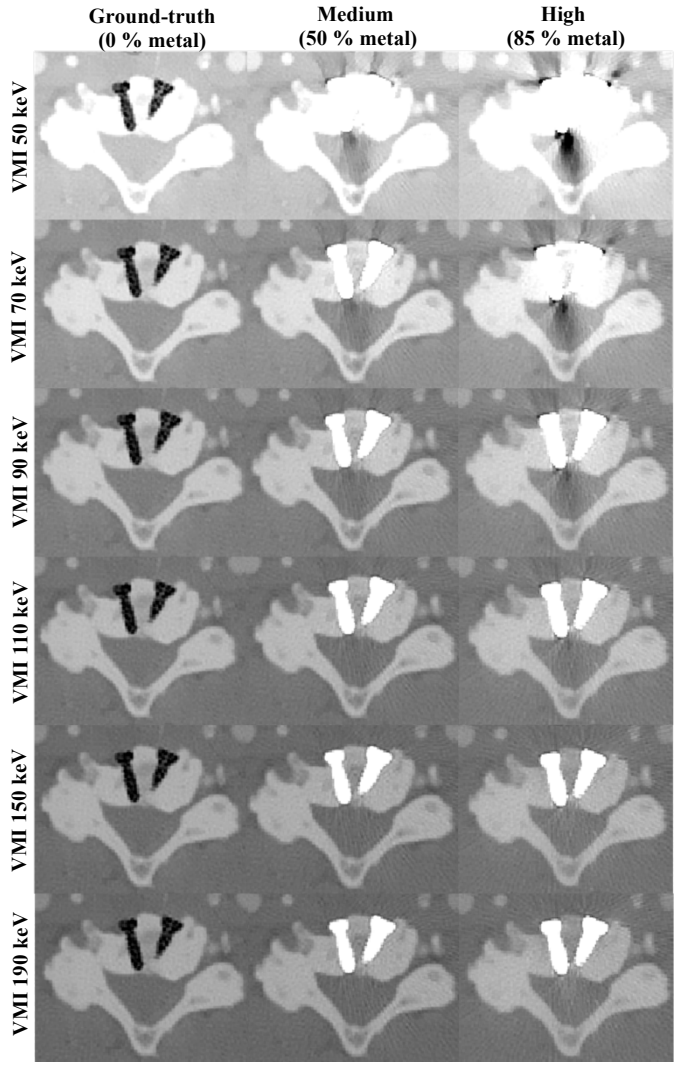


Figure 3. Axial C6 images of the three *PixelPrint* cervical spine phantoms without MAR, shown across VMI energies from 50 to 190 keV (rows) and metal infill levels of 0 %, 50 %, and 85 % (columns). Increasing metal infill produces stronger artifacts at low energies, while higher VMI reconstructions reduce streaking and improve anatomical visibility. WL/WW: 0/1500 HU.

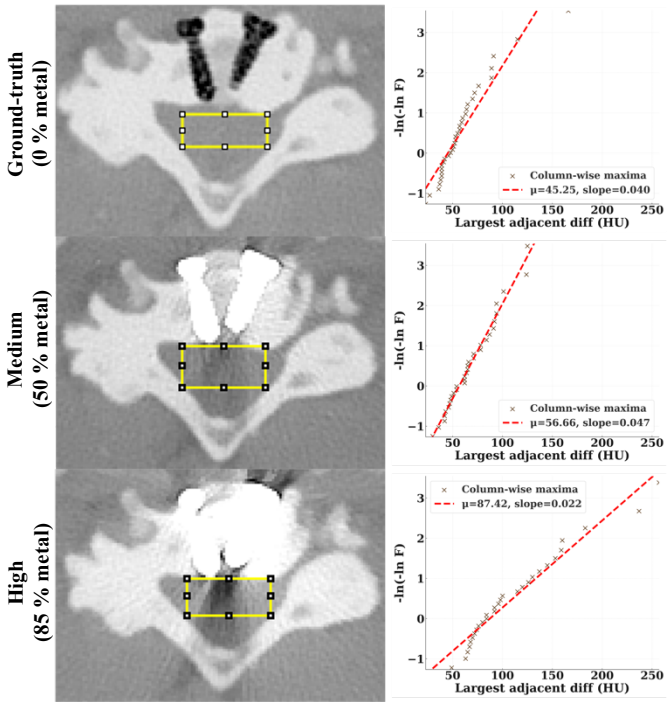


Figure 4. Quantitative evaluation of metal artifact severity using a Gumbel-based extreme-value analysis. Axial images show artifact evaluation ROIs for the ground-truth (0% infill), medium metal (50% infill), and high metal (85% infill) phantoms. Corresponding Gumbel distribution fits demonstrate progressively increasing artifact-induced intensity variations with increasing metal infill, as reflected by a monotonic increase in the Gumbel location parameter (μ). WL/WW: 0/1500 HU.

Across higher virtual monoenergetic levels, metal artifacts are progressively reduced in all phantoms, improving visualization of the screw–bone interface (including shafts and threads) while preserving bone and soft-tissue detail (Figure 3).

Metal artifact severity was quantified using a Gumbel-based extreme-value analysis of rectangular ROIs placed in artifact-prone tissue adjacent to the screws (Figure 4). The Gumbel location parameter (μ); reflecting extreme adjacent pixel-intensity differences from streaking and photon starvation, increased monotonically with metal infill: 45.25 HU (0% infill), 56.66 HU (50% infill), and 87.42 HU (85% infill).

IV. Discussion

Dual-filament *PixelPrint* enables patient-specific CT phantoms that preserve anatomy while allowing independent, reproducible control of metal artifact severity. By pairing an artifact-free reference phantom (screw geometry present but 0% metal infill) with phantoms containing defined metal content, we create a true baseline with identical anatomy but no metal artifacts, enabling objective quantification of artifact generation and reduction. Unlike conventional MAR phantoms with simplified shapes, this approach maintains realistic vertebral anatomy, implant geometry, and clinically relevant tissue attenuation, while systematically tuning artifact strength via the metal-doped filament infill percentage. The marked artifact reduction at higher virtual monoenergetic energies aligns with

expected spectral CT behavior, further supporting the phantoms' realism in attenuation response and energy-dependent artifact characteristics. Although soft tissue and moderate-density bone closely match patient HU values, reproducing the very high attenuation of dense cortical bone remains limited by the achievable range of calcium-doped PLA. This limitation has minimal impact on MAR benchmarking, since clinically relevant artifacts predominantly degrade the peri-implant soft tissues and adjacent bone rather than the highest-attenuation cortical bone itself. Overall, controllable metal infill provides a practical route to generate anatomically identical phantoms spanning a range of artifact severities - an evaluation framework not available with commercial phantoms.

V. Conclusion

Dual-filament *PixelPrint* enables reproducible fabrication of patient-specific, multi-material CT phantoms with independently tunable metal artifact severity. By co-printing calcium-doped PLA for tissue/bone surrogates and metal-doped PLA for implants, the phantoms preserve realistic anatomy and screw geometry while generating clinically relevant, energy-dependent artifact patterns that can be systematically modulated via implant infill. This approach provides a practical, quantitative benchmark for evaluating and comparing MAR algorithms and spectral CT reconstruction methods under controlled, anatomically realistic conditions.

Acknowledgment

We acknowledge support from the National Institutes of Health (**R01EB035092, R01EB035908**).

References

[1] Willemink MJ, Noël PB. The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence. *Eur Radiol*. 2019;29(5):2185-2195. doi:10.1007/s00330-018-5810-7

[2] Wellenberg RHH, Hakvoort ET, Slump CH, Boomsma MF, Maas M, Streekstra GJ. Metal artifact reduction techniques in musculoskeletal CT-imaging. *Eur Radiol*. 2018;107:60-69. doi:10.1016/j.ejrad.2018.08.010

[3] Kawahara D, Ozawa S, Yokomachi K, et al. Metal artifact reduction techniques for single energy CT and dual-energy CT with various metal materials. *BJR Open*. 2019;1(1):20180045. Published 2019 Jul 8. doi:10.1259/bjro.20180045

[4] Jeong S, Kim SH, Hwang EJ, Shin CI, Han JK, Choi BI. Usefulness of a metal artifact reduction algorithm for orthopedic implants in abdominal CT: phantom and clinical study results. *AJR Am J Roentgenol*. 2015;204(2):307-317. doi:10.2214/AJR.14.12745

[5] Van Der Bie J, Sharma SP, Hirsch A, et al. Coronary Stent Imaging with Photon-counting Detector CT. *Radiology Cardiothoracic Imaging*. 2025;7(3):e240338. doi:10.1148/ryct.240338

[6] Große Hokamp N, Eck B, Siedek F, et al. Quantification of metal artifacts in computed tomography: methodological considerations. *Quant Imaging Med Surg*. 2020;10(5):1033-1044. doi:10.21037/qims.2020.04.03

[7] Hatamikia S, et al. Impact of metal artifacts on CT-based diagnosis and quantitative imaging. Z Med Phys, 33(3), 2023.

[8] Tino R, Yeo A, Leary M, Brandt M, Kron T. A Systematic Review on 3D-Printed Imaging and Dosimetry Phantoms in Radiation Therapy. *Technol Cancer Res Treat*. 2019;18:1533033819870208. doi:10.1177/1533033819870208

[9] The Phantom Laboratory. Catphan® 700 Product Information. Available from: https://www.phantomlab.com/library. Accessed Jan. 2026.

[10] Branco D, Kry S, Taylor P, et al. Evaluation of image quality of a novel computed tomography metal artifact management technique on an anthropomorphic head and neck phantom. *Phys Imaging Radiat Oncol*. 2021;17:111-116. Published 2021 Feb 21. doi:10.1016/j.phro.2021.01.007

[11] Shapira N, Donovan K, Mei K, et al. Three-dimensional printing of patient-specific computed tomography lung phantoms: a reader study. *PNAS Nexus*. 2023;2(3):pgad026. Published 2023 Feb 1. doi:10.1093/pnasnexus/pgad026

[12] Mei K, Pasyar P, Geagan M, et al. Design and fabrication of 3D-printed patient-specific soft tissue and bone phantoms for CT imaging. Preprint. *Res Sq*. 2023;rs.3.rs-2828218. Published 2023 Apr 26. doi:10.21203/rs.3.rs-2828218/v1

[13] Mei K, Geagan M, Roshkovan L, et al. Three-dimensional printing of patient-specific lung phantoms for CT imaging: Emulating lung tissue with accurate attenuation profiles and textures. *Med Phys*. 2022;49(2):825-835. doi:10.1002/mp.15407

[14] Pasyar P, Im J, Mei K, et al. PixelPrint: Generating Patient-Specific Phantoms for Spectral CT using Dual Filament 3D Printing. *Proc SPIE Int Soc Opt Eng*. 2024;12925:1292527. doi:10.1117/12.3006512

[15] Cammin J. A robust index for metal artifact quantification in computed tomography. *J Appl Clin Med Phys*. 2024;25(8):e14453. doi:10.1002/acm2.1445

A Site-Agnostic Dual-Domain Deep Learning Reconstruction Framework for Nonstop Gated CBCT

Mitchell Yu, Noah Silverberg, Sean Berry, Wendy Harris, Jean Moran, Lei Xing, Hao Zhang

***Abstract*—Respiratory gating radiotherapy (RG-RT) allows patients to breathe freely while only delivering radiation during a specific portion of the breathing cycle, offering clear advantages for thoracic and abdominal cancers such as lung, pancreas, and liver. Accurate pretreatment imaging is essential for the safe and effective delivery of RG-RT. We previously developed nonstop gated CBCT (ngCBCT) to replace current clinical gated CBCT for patient setup, reducing half-fan scan time on C-arm linear accelerators from 2-8 minutes to just 1 minute while substantially lowering imaging dose. However, ngCBCT produces highly non-uniform and under-sampled projection data, posing a major challenge for conventional reconstruction methods. In this study, we developed deep learning-based reconstruction strategies for ngCBCT across multiple disease sites. Specifically, dual-domain convolutional neural network (DDCNN) models were trained using both site-specific datasets and pooled cross-site data from lung, pancreas, and liver cases. We demonstrate that a DDCNN trained on data from a single disease site (e.g., lung) generalizes effectively to other sites (e.g., pancreas, liver), achieving high-quality reconstructions despite substantial differences in thoracic and abdominal projection characteristics and image-domain anatomy. Furthermore, a single site-agnostic DDCNN provides image quality comparable to site-specific models across all evaluated sites. This unified reconstruction framework represents a key innovation by removing the need for disease-site-specific training, thereby simplifying clinical implementation and improving scalability. These results support a standardized ngCBCT acquisition and reconstruction paradigm for RG-RT across diverse disease sites.**

Keywords: nonstop gated CBCT, deep learning reconstruction, site-agnostic model, respiratory gating radiotherapy

I. Introduction

Respiratory motion remains a major challenge in radiotherapy for thoracic and abdominal cancers[1], [2]. Respiratory gating radiotherapy (RG-RT) mitigates motion effects by delivering radiation only during a specific portion of the breathing cycle (known as the "gating window"), reducing radiation-induced toxicity and facilitating tumor dose escalation[3], [4]. For patients undergoing RG-RT, free-breathing gated CBCT (gCBCT) is routinely used for pretreatment setup verification by imaging anatomy corresponding to the gating window[5]. However, current gCBCT acquisitions on C-arm linear accelerators are inefficient and time-consuming, typically requiring 2–8 minutes per scan for duty cycles of 30–60% and breathing periods of 3–6 s. This inefficiency arises from repeated gantry acceleration, deceleration, and "setback" maneuvers, as gantry rotation and kV acquisition are restricted to the gating window. Prolonged acquisition time increases patient on-table time, discomfort, and the risk of patient motion and breathing baseline drift—issues compounded by multiple scans per fraction and multiple treatment fractions.

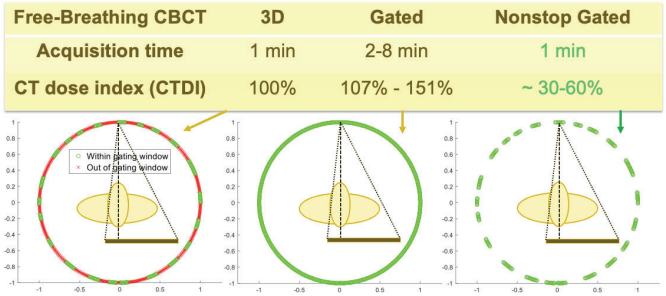


Fig. 1. Comparison of acquisition time, CTDI, and projection sampling among 3D CBCT, gated CBCT (gCBCT), and nonstop gated CBCT (ngCBCT). Compared with current gCBCT, the ngCBCT achieves substantial reductions in acquisition time and imaging dose but also produces highly non-uniform and under-sampled projection data.

To overcome these limitations, we proposed and implemented nonstop gated CBCT (ngCBCT), which enables continuous gantry rotation while activating the kV X-ray beam only during the gating window [6]. A half-fan ngCBCT scan is completed in a single 360° rotation (≈1 minute) while substantially reducing imaging dose. Figure 1 compares free-breathing 3D CBCT, gCBCT, and ngCBCT in terms of acquisition time, CT dose index, and projection distribution. Animation videos illustrating the acquisition mechanisms of gCBCT and ngCBCT are provided as a YouTube link in the Appendix for interested readers. However, ngCBCT produces highly non-uniform and under-sampled projections, posing a significantly more complex reconstruction challenge than traditional sparse-view CBCT. Conventional reconstruction methods, including Feldkamp–Davis–Kress (FDK) and penalized-likelihood iterative reconstruction, have been shown to produce suboptimal image quality for ngCBCT. Although prior-image–based iterative reconstruction can improve image quality, its reliance on high-quality patient-specific prior images limits its clinical applicability [7].

Recent advances in convolutional neural network (CNN)-based image reconstruction have created new opportunities to address the challenges posed by the highly non-uniform and under-sampled projection data in ngCBCT. We initially investigated image-domain–only CNN approaches and observed that reconstruction quality was strongly limited by the severe artifacts present in the input images, which arise from

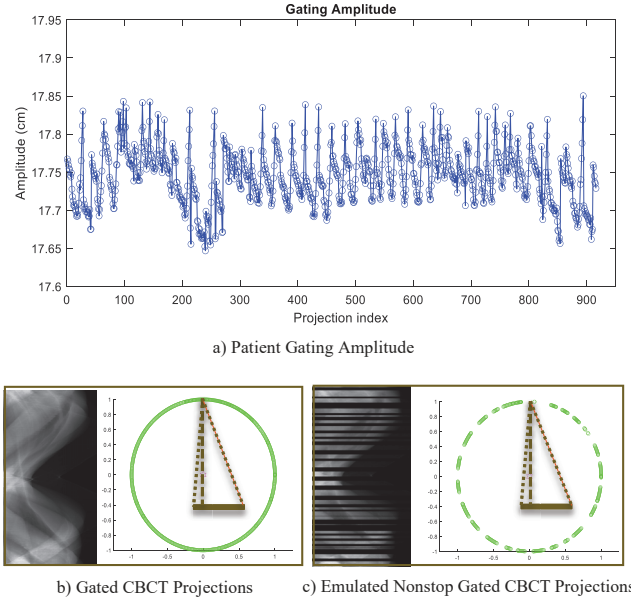
This project was funded partially through Varian Medical Systems, Inc. Noah Silverberg was partially supported by the AAPM Summer Undergraduate Fellowship. M. Yu, N. Silverberg, S. Berry, W. Harris, and J. Moran are with the Department of Medical Physics, Memorial Sloan Kettering Cancer Center, New York, NY 10065. L. Xing and H. Zhang* are with the Department of Radiation Oncology, Stanford University School of Medicine, Stanford, CA 94305 (correspondence: hzhang96@stanford.edu)

nonuniform sampling. To overcome this limitation, we developed a dual-domain CNN (DDCNN) framework that jointly exploits information in both the projection and image domains. In DDCNN, a projection-domain network synthesizes missing projections, while an image-domain network refines the reconstructed image, with the FDK algorithm serving as the link between the two domains. This dual-domain strategy enables high-quality ngCBCT reconstruction with rapid inference (<1 min), supporting real-time clinical deployment. Despite these promising results, widespread clinical adoption remains challenging due to deployment complexity across different disease sites affected by respiratory motion (e.g., lung, pancreas, and liver). A key unanswered question is whether a single, universal reconstruction model can generalize across multiple anatomical sites without sacrificing image quality. Addressing this question and evaluating the feasibility of a site-agnostic DDCNN for ngCBCT reconstruction is the focus of this work.

II. Methods

A. Data Acquisitions

The nonstop gated CBCT (ngCBCT) data used in this study were emulated from clinical gated CBCT (gCBCT) scans from respiratory gating radiotherapy patients on Varian TrueBeam linear accelerators. Figure 2 illustrates an example of the projection data emulation process. All projections acquired with gCBCT are within the gating window and by selecting projections from alternating patient breathing cycles, we were able to emulate the non-uniformly under-sampled ngCBCT projection data.



a) Patient Gating Amplitude

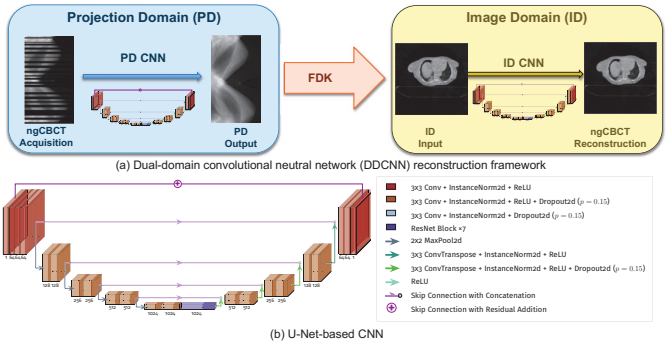
b) Gated CBCT Projections c) Emulated Nonstop Gated CBCT Projections

Fig. 2. The nonstop gated CBCT projections in this study were emulated from the clinical gated CBCT projections.

Three retrospective datasets of half-fan gCBCT scans (offset detector) were collected. The first dataset included 54 scans from 23 lung cancer patients, the second included 58 scans from 13 pancreatic cancer patients, and the third included 52 scans from 12 liver cancer patients. Pancreas and liver patients were typically implanted with fiducial markers near the tumor.

B. Dual-Domain Reconstruction Framework

The dual-domain CNN (DDCNN) framework (Fig. 3) comprises a projection-domain (PD) CNN and an image-domain (ID) CNN, linked by the fast Feldkamp–Davis–Kress (FDK) reconstruction. The end-to-end reconstruction time is within 1 minute (on a single GPU), supporting real-time ngCBCT adaptation for image-guided radiotherapy. Both the PD and ID CNNs adopt U-Net–based architectures. A global skip connection between the network input and output facilitates residual learning, while seven ResNet blocks are incorporated to enhance texture preservation. Dropout (15%) is applied to all layers except the first to mitigate overfitting while preserving essential input information.



(a) Dual-domain convolutional neutral network (DDCNN) reconstruction framework

(b) U-Net-based CNN

Fig. 3. Dual-domain convolutional neural network (DDCNN), consisting of a projection-domain CNN and an image-domain CNN connected by the fast FDK reconstruction. Both networks adopt a U-Net–based architecture.

Three site-specific DDCNN models were trained: (1) a lung model using a data split of 42 training, 6 validation, and 6 test scans; (2) a pancreas model with 47/6/5 scans for training/validation/testing; and (3) a liver model with 42/5/5 scans, respectively. In addition, a cross-site DDCNN model was trained using data from all three sites, with 45 scans (15 per site) for training and 12 scans (4 per site) for validation, while all remaining scans were reserved for testing. The size of the cross-site training dataset was matched to that of the single-site models to enable fair performance comparisons.

Data augmentation was applied for both projection-domain and image-domain training. For the PD CNNs, selecting projections from either odd or even respiratory cycles during down-sampling allowed two distinct ngCBCT datasets to be generated from each gCBCT scan, effectively doubling the training data. For the ID CNNs, horizontal and vertical image flips were applied, tripling the size of the training dataset. The PD CNNs were trained for 20 epochs, and the ID CNNs were trained for 50 epochs.

III. Results

We emulated ngCBCT acquisition for three respiratory motion affected sites: lung, pancreas, and liver. Three site-specific DDCNN models were trained and evaluated using held-out test scans from their respective sites, as well as cross-evaluated using scans from other anatomical sites (e.g., applying the lung-trained model to liver scans). In addition, a unified cross-site DDCNN model was trained and evaluated using test scans from all three sites. Comparisons between gCBCT and ngCBCT reconstructions, including both site-specific and cross-site models, are presented in Figures 3–5.

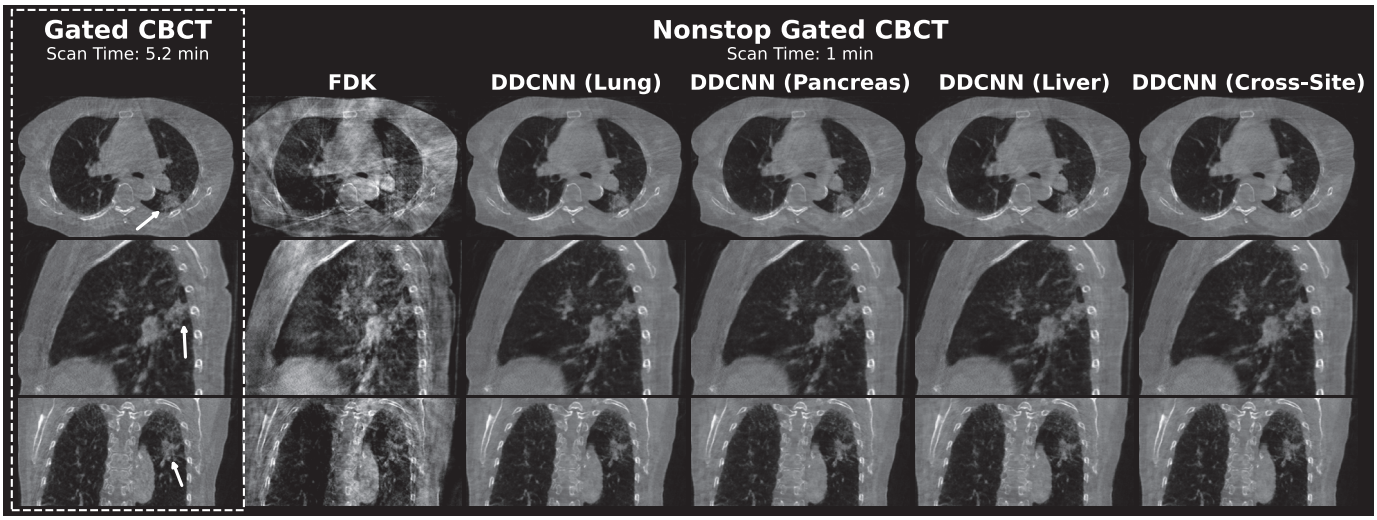


Fig. 4. For a lung cancer patient, the 5.2-minutes gated CBCT scan is compared with the much faster nonstop gated CBCT reconstructed with FDK and DDCNN models trained with data from various sites. The lung tumor in the gCBCT is pointed out by the white arrow.

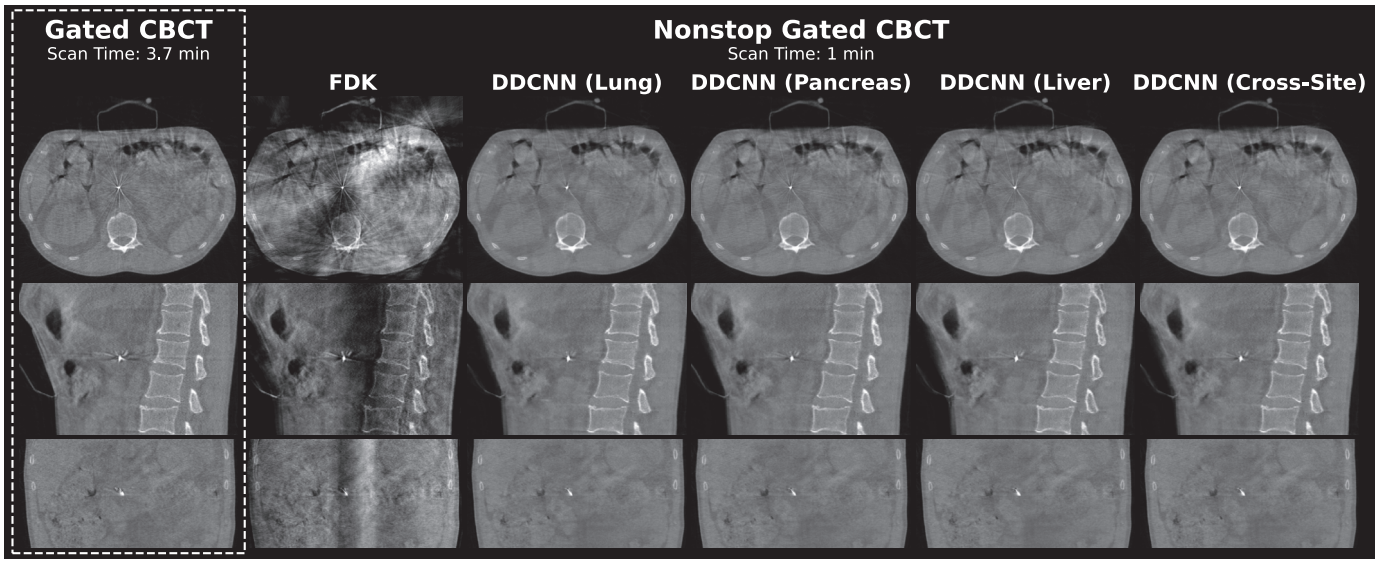


Fig. 5. For a pancreatic cancer patient, the 3.7-minutes gated CBCT scan is compared with the much faster nonstop gated CBCT reconstructed with FDK and DDCNN models trained with data from various sites.

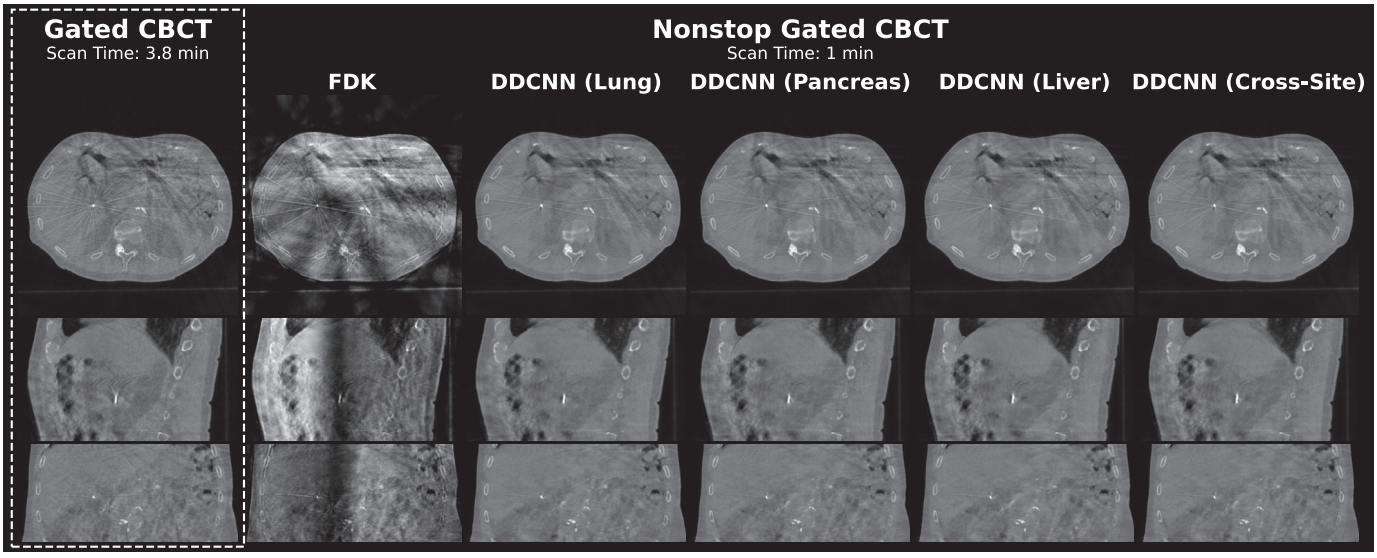


Fig. 6. For a liver cancer patient, the 3.8-minutes gated CBCT scan is compared with the much faster nonstop gated CBCT reconstructed with FDK and DDCNN models trained with data from various sites.

Quantitative image quality metrics, including PSNR and SSIM, are summarized in Tables I and II.

TABLE I
DDCNN Models Performance Comparison: PSNR(dB)

DDCNN Model Site	*Test Scan Site* Lung	Pancreas	Liver
Lung	**33.44 ± 2.07**	34.17 ± 1.52	34.72 ± 0.79
Pancreas	33.16 ± 2.22	**34.43 ± 1.54**	34.99 ± 0.85
Liver	33.10 ± 2.13	34.18 ± 1.47	34.69 ± 0.70
Cross-site	33.43 ± 2.13	34.33 ± 1.48	**35.12 ± 0.71**

TABLE II
DDCNN Models Performance Comparison: SSIM (%)

DDCNN Model Site	*Test Scan Site* Lung	Pancreas	Liver
Lung	**93.90 ± 2.33**	94.43 ± 1.67	95.70 ± 0.59
Pancreas	93.63 ± 2.54	**94.67 ± 1.68**	95.95 ± 0.59
Liver	93.60 ± 2.45	94.51 ± 1.63	95.70 ± 0.52
Cross-site	93.89 ± 2.37	94.61 ± 1.67	**96.02 ± 0.46**

IV. Discussions

One limitation of image-domain only CNNs is their strong dependence on the quality of the input images. When the input reconstructions are severely degraded—as is the case for ngCBCT due to highly nonuniform projection sampling—important image details such as texture, fine structures, and tumor appearance can be compromised. The proposed DDCNN more effectively preserves anatomical structures and soft-tissue contrast compared with image-domain–only CNNs, highlighting the advantage of jointly leveraging projection- and image-domain information.

A key finding of this study is that a DDCNN trained on a single disease site generalizes well to other anatomical sites, despite substantial differences between thoracic and abdominal projection data. This behavior suggests that the projection-domain CNN primarily learns the acquisition- and physics-driven sampling deficiencies imposed by ngCBCT—namely the nonuniform angular gaps and view correlations dictated by the gating scheme—rather than site-specific anatomical features. Because these sampling patterns are similar across disease sites, the network can reliably predict missing projections independent of anatomy. Operating in the measurement domain further reduces sensitivity to tissue-specific contrast, while the image-domain CNN mitigates residual artifacts and enforces structural consistency. Together, these factors enable robust, site-agnostic reconstruction across diverse disease sites.

Quantitative evaluations show that the unified cross-site model performs comparably to single-site models. Notably, although not fully captured by PSNR and SSIM metrics, the cross-site model often reduces artifacts present in the ground-truth gated CBCT images, suggesting improved robustness to site-dependent imaging inconsistencies. Several strategies may further enhance cross-site performance. One approach is to incorporate site-specific network heads, whereby a unified backbone is first trained using pooled multi-site data and then fine-tuned with additional site-specific training. Such an approach could allow targeted optimization of anatomical features unique to each site, such as tumor shape and intensity fidelity for lung imaging or soft-tissue contrast for liver and pancreas imaging.

V. Conclusions

We introduced ngCBCT as a scan-time- and dose-efficient replacement for the current clinical gCBCT implementation for respiratory gating radiotherapy. To address the challenging reconstruction problem posed by the highly non-uniform and under-sampled ngCBCT projection data, we developed a DDCNN that jointly leverages information in the projection and image domains and enables end-to-end reconstruction within 1 minute. The proposed framework was evaluated using both site-specific models and a unified cross-site model trained on lung, liver, and pancreas data. Our results demonstrate that a single, site-agnostic DDCNN achieves reconstruction quality comparable to disease-site–specific models across all evaluated sites. By eliminating the need for multiple site-specific models, this unified approach simplifies clinical deployment and enhances scalability. Together, these findings support a standardized ngCBCT acquisition and reconstruction paradigm for respiratory gating radiotherapy across diverse disease sites. The integration of ngCBCT with a fast, robust, and site-agnostic DDCNN reconstruction framework has the potential to significantly improve clinical workflow efficiency, patient experience, and ultimately treatment accuracy and outcomes.

Appendix

To help readers understand the operation of clinical gCBCT and the proposed ngCBCT, we have prepared short video clips to illustrate the underlying mechanisms. We also include a brief video showing ngCBCT implementation on a Varian TrueBeam linear accelerator with a CIRS dynamic thorax phantom, demonstrating technical feasibility. The full video playlist is available on YouTube:
https://www.youtube.com/@gatedCBCT/playlists.

References

[1] B. K. Chang and R. D. Timmerman, "Stereotactic body radiation therapy: A comprehensive review," Dec. 2007. doi: 10.1097/COC.0b013e3180ca7cb1.

[2] B. Bussels *et al.*, "Respiration-induced movement of the upper abdominal organs: a pitfall for the three-dimensional conformal radiation treatment of pancreatic cancer," *Radiotherapy and Oncology*, vol. 68, no. 1, pp. 69–74, Jul. 2003, doi: 10.1016/S0167-8140(03)00133-6.

[3] A. M. Berson *et al.*, "Clinical experience using respiratory gated radiation therapy: Comparison of free-breathing and breath-hold techniques," *International Journal of Radiation Oncology*Biology*Physics*, vol. 60, no. 2, pp. 419–426, Oct. 2004, doi: 10.1016/j.ijrobp.2004.03.037.

[4] B. Zhao, Y. Yang, T. Li, X. Li, D. E. Heron, and M. S. Huq, "Statistical analysis of target motion in gated lung stereotactic body radiation therapy," *Phys Med Biol*, vol. 56, no. 5, pp. 1385–1395, Mar. 2011, doi: 10.1088/0031-9155/56/5/011.

[5] R. E. Kincaid, E. D. Yorke, K. A. Goodman, A. Rimner, A. J. Wu, and G. S. Mageras, "Investigation of gated cone-beam CT to reduce respiratory motion blurring," *Med Phys*, vol. 40, no. 4, 2013, doi: 10.1118/1.4795336.

[6] X. He *et al.*, "Next-generation nonstop gated CBCT for respiratory gating lung radiotherapy: scan time and imaging dose," *Med Phys*, vol. 52, no. 10, Oct. 2025, doi: https://doi.org/10.1002/mp.70065.

[7] M. Yu *et al.*, "Nonstop gated CBCT for respiratory gating lung SBRT: A feasibility study," *Med Phys*, vol. 52, no. 9, Sep. 2025, doi: 10.1002/mp.18084.

Differentiable Forward and Back Projectors for Gradient-Based Deformable Motion Estimation

Xiao Jiang, Grace J. Gang, and J. Webster Stayman

***Abstract*—Accurate estimation and compensation of patient motion are critical for many X-ray imaging applications, including image-guided interventions, cone-beam CT (CBCT) reconstruction, and multi-phase CT imaging. While gradient-based optimization may offer an effective framework for motion estimation, its practicality has been limited by the difficulty of computing analytical gradients of forward and back-projection operators with respect to motion. To address this issue, our recent work proposed a scalable, accurate, and efficient differentiable projector with respect to rigid patient motion. In this work, we extend this method from rigid motion to deformable motion by deriving analytical gradients of forward and back-projectors with respect to a spatially varying displacement vector field (DVF) in the continuous domain. We show that, similar to the rigid case, these gradients preserve the same integration structure as the original projection operators, enabling effective gradient computation using standard projector and backprojector implementations. The proposed method is evaluated in two X-ray imaging tasks: deformable 2D/3D registration using simulated biplane radiography and motion-compensated filtered backprojection (FBP) reconstruction. Simulation results demonstrate effective deformable motion estimation for both 2D/3D registration and motion-compensated FBP tasks, where both qualitative assessment and quantitative analysis reveal substantially reduced projection mismatch and motion-induced artifacts. These results indicate that the proposed differentiable projector provides a practical solution for gradient-based deformable motion estimation and compensation.**

I. Introduction

PATIENT motion estimation is critical in a variety of clinical X-ray imaging applications, including image registration between preoperative and intraoperative volumes, motion compensation during cone-beam CT (CBCT) reconstruction from motion-corrupted projections, and 4D/5D CT reconstruction using estimated respiratory or cardiac motion phases. These tasks aim to align medical volumes at different time points or mitigate image artifacts introduced by unwanted patient motion.

A large class of motion estimation algorithms is formulated as an optimization problem. In these approaches, motion parameters are estimated by optimizing projection-domain similarity measures or image-domain sharpness metrics, followed by motion-compensated reconstruction or registration. Many existing methods rely on derivative-free optimization techniques, such as evolution strategies [1] or direction-set methods [2], due to the difficulty of computing analytical gradients of projection and backprojection operators with respect to motion. While robust, these approaches typically require a large number of forward and back-projection operations, leading to high computational cost. Differentiable solutions have also been investigated to enable gradient-based optimization. Such methods [3], [4] often adopt partial angle reconstruction (PAR) strategies, combined with PAR-based deformation models, which allow end-to-end differentiation but may sacrifice computational efficiency due to repeated volume deformation operations.

In our recent work, we introduced a general and memory-efficient framework for differentiable forward and back-projectors with respect to rigid motion.[5] By deriving analytical expressions for motion gradients in the continuous domain, we showed that the gradients of forward and back-projection share the same integration structure as the original projection operators. This insight enables efficient gradient computation using existing projector implementations without relying on full automatic differentiation. The resulting framework demonstrated improved efficiency, scalability, and broad applicability across multiple rigid-motion imaging tasks, including 2D/3D registration, motion-compensated analytical reconstruction, and online geometry calibration.

In this work, we extend the proposed differentiable projector framework to the more general and challenging case of deformable motion. Instead of parameterizing motion using a small number of rigid degrees of freedom [1], deformable motion is modeled using a spatially varying Displacement Vector Field (DVF). We derive analytical gradients of forward and back-projectors with respect to the DVF in the continuous domain and show that the same structural properties observed in the rigid case continue to hold. This formulation enables practical and efficient computation of projector and backprojector gradients with respect to the DVF. The proposed method is evaluated on deformable 2D/3D registration and motion-compensated FBP reconstruction, highlighting its potential for advanced X-ray imaging applications involving complex deformable motion.

II. Methodology

A. X-ray Projection Model with Deformable Motion

The forward- and back-projection of a cone-beam X-ray imaging system with deformable object motion are illustrated in Fig. 1. Their mathematical formulations can be written as

$$\text{Forward:}\quad l(\mathbf{d},\boldsymbol{\theta}) = \int \mu(\mathbf{r}'(\mathbf{r},\mathbf{h}(\mathbf{r},\boldsymbol{\theta})))\delta(p(\mathbf{r})-\mathbf{d})\mathrm{d}\mathbf{r}, \quad \text{(1a)}$$

$$\text{Back:}\quad \mu(\mathbf{r}',\boldsymbol{\theta}) = \int l(\mathbf{d})\delta(p(\mathbf{r}(\mathbf{r}',\mathbf{h}'(\mathbf{r}',\boldsymbol{\theta})))-\mathbf{d})\mathrm{d}\mathbf{d}, \quad \text{(1b)}$$

Xiao Jiang and J.Webster Stayman are with the Department of Biomedical Engineering. Email: {xjiang43, web.stayman}@jhu.edu

Grace Gang is with the Department of Radiology, University of Pennsylvania. Email: grace.j.gang@pennmedicine.upenn.edu

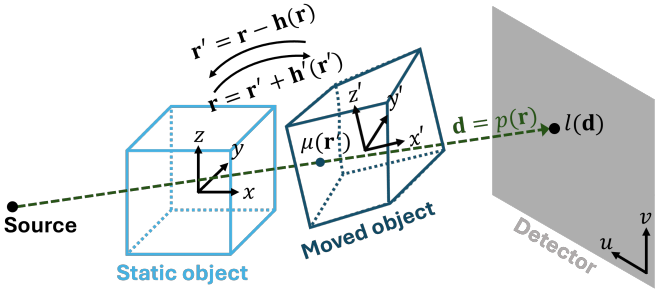


Fig. 1. Illustration of X-ray projection model with deformable object motion.

The Dirac function $\delta(\cdot)$ defines the projection ray geometry, ensuring that attenuation values are integrated along the corresponding X-ray paths in forward projection and distributed back along the same paths in backprojection. The $\mathbf{h}$ denotes a three-dimensional displacement vector that is spatially varying for deformable motion and parameterized by $\boldsymbol{\theta}$. A point $\mathbf{r}$ in the deformed volume is mapped back to the static reference volume $\mathbf{r}'$ via $\mathbf{r}' = \mathbf{r} - \mathbf{h}(\mathbf{r})$, and the collection of $\mathbf{h}$ over the image domain forms the inverse DVF. In the backprojection, motion is characterized by the forward DVF $\mathbf{h}'$, which maps points in the reference volume to the deformed volume according to $\mathbf{r} = \mathbf{r}' + \mathbf{h}'(\mathbf{r}')$. The projective geometry is described by projection matrix $\mathbf{P} \in \mathbb{R}^{3\times4}$, which maps a point $\mathbf{r}$ in 3D space to a point $\mathbf{d}$ on 2D detector:

$$\mathbf{d} = p(\mathbf{r}) : \begin{pmatrix} u' \\ v' \\ w' \end{pmatrix} = \mathbf{P}\mathbf{r}, \;\; \mathbf{d} = \begin{pmatrix} u'/w' \\ v'/w' \end{pmatrix} \tag{2}$$

B. Gradient of Projectors with Respect to Deformable Motion

Following a derivation analogous to the rigid-motion case [5], the gradients of the forward and back-projectors with respect to the motion parameters $\boldsymbol{\theta}$ can be factorized as $\nabla_{\boldsymbol{\theta}} l(\mathbf{d}, \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \mathbf{h}(\boldsymbol{\theta}) \nabla_{\mathbf{h}} l(\mathbf{d}, \mathbf{h})$ and $\nabla_{\boldsymbol{\theta}} \mu(\mathbf{r}', \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \mathbf{h}'(\boldsymbol{\theta}) \nabla_{\mathbf{h}'} \mu(\mathbf{r}', \mathbf{h}')$. The terms $\nabla_{\boldsymbol{\theta}} \mathbf{h}(\boldsymbol{\theta})$ and $\nabla_{\boldsymbol{\theta}} \mathbf{h}'(\boldsymbol{\theta})$ depend on the specific DVF parameterization. In this work, we focus on deriving the projector gradients $\nabla_{\mathbf{h}} l(\mathbf{d}, \mathbf{h})$ and $\nabla_{\mathbf{h}'} \mu(\mathbf{r}', \mathbf{h}')$.

Applying the chain rule, the gradient of the forward projector with respect to the inverse DVF $\mathbf{h}$ can be written as

$$\begin{aligned} \nabla_{\mathbf{h}} l(\mathbf{d}, \mathbf{h}) &= \nabla_{\mathbf{h}} \int \mu(\mathbf{r}'(\mathbf{r}, \mathbf{h})) \delta(p(\mathbf{r}) - \mathbf{d}) \mathrm{d}\mathbf{r} \\ &= \int \nabla_{\mathbf{r}'}^T \mu(\mathbf{r}') \nabla_{\mathbf{h}} \mathbf{r}'(\mathbf{r}, \mathbf{h}) \delta(p(\mathbf{r}) - \mathbf{d}) \mathrm{d}\mathbf{r} \\ &= \int \mathbf{G}_f(\mathbf{r}) \delta(p(\mathbf{r}) - \mathbf{d}) \mathrm{d}\mathbf{r}, \end{aligned} \tag{3}$$

where

$$\mathbf{G}_f(\mathbf{r}) = \begin{pmatrix} \partial_{x'}\mu & \partial_{y'}\mu & \partial_{z'}\mu \end{pmatrix} \begin{pmatrix} \nabla_{\mathbf{h}} x' \\ \nabla_{\mathbf{h}} y' \\ \nabla_{\mathbf{h}} z' \end{pmatrix}. \tag{4}$$

Using a similar approach, the gradient of the backprojector with respect to the forward DVF $\mathbf{h}'$ is given by

$$\begin{aligned} \nabla_{\mathbf{h}'} \mu(\mathbf{r}', \mathbf{h}') &= \nabla_{\mathbf{h}'} \int l(\mathbf{d}) \delta(p(\mathbf{r}(\mathbf{r}', \mathbf{h}')) - \mathbf{d}) \mathrm{d}\mathbf{d} \\ &= \int l(\mathbf{d}) \nabla_{\mathbf{h}'} p(\mathbf{r}(\mathbf{r}', \mathbf{h}')) \delta'(p(\mathbf{r}(\mathbf{r}', \mathbf{h}')) - \mathbf{d}) \mathrm{d}\mathbf{d} \\ &= \int \nabla_{\mathbf{d}}^T l(\mathbf{d}) \nabla_{\mathbf{r}}^T p(\mathbf{r}) \nabla_{\mathbf{h}'} \mathbf{r}(\mathbf{r}', \mathbf{h}')) \delta(p(\mathbf{r}(\mathbf{r}', \mathbf{h}')) - \mathbf{d}) \mathrm{d}\mathbf{d} \\ &= \int \mathbf{G}_b(\mathbf{d}) \delta(p(\mathbf{r}(\mathbf{r}', \mathbf{h}'))) - \mathbf{d}) \mathrm{d}\mathbf{d}, \end{aligned} \tag{5}$$

where

$$\mathbf{G}_b(\mathbf{d}) = \begin{pmatrix} \partial_u l & \partial_v l \end{pmatrix} \begin{pmatrix} \partial_x u & \partial_y u & \partial_z u \\ \partial_x v & \partial_y v & \partial_z v \end{pmatrix} \begin{pmatrix} \nabla_{\mathbf{h}'} x \\ \nabla_{\mathbf{h}'} y \\ \nabla_{\mathbf{h}'} z \end{pmatrix}. \tag{6}$$

For the inverse and forward DVFs, the gradients of spatial coordinates can be written as

$$\nabla_{\mathbf{h}} x' = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix}, \nabla_{\mathbf{h}} y' = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}, \nabla_{\mathbf{h}} z' = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}, \tag{7}$$

$$\nabla_{\mathbf{h}'} x = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \nabla_{\mathbf{h}'} y = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \nabla_{\mathbf{h}'} z = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{8}$$

Eqs. (1a) and (3), as well as Eqs. (1b) and (5), share exactly the same integration structure. As a result, the gradients of the forward and back-projectors can be computed by applying standard forward- and back-projection operators to each component of $\mathbf{G}_f$ and $\mathbf{G}_b$, respectively.

C. Evaluation

1) DVF Parameterization: DVFs are typically smooth in both space and time. In this work, we adopt a simple and efficient parameterization based on spatial downsampling. Specifically, the optimization variable $\boldsymbol{\theta}$ is defined as a downsampled version of the full-resolution DVF $\mathbf{h}$. In this work, the sampling of $\boldsymbol{\theta}$ is reduced by a factor of $8\times$ along each spatial dimension relative to $\mathbf{h}$, and the full DVF $\mathbf{h}$ is resampled from $\boldsymbol{\theta}$ using trilinear interpolation. This parameterization significantly reduces the number of unknowns while enforcing smoothness of the estimated DVF.

2) 2D/3D Registration using the Differentiable Forward Projector: We consider a biplane radiography imaging system for image-guided interventions, where two orthogonal views are simulated. As shown in Fig. 2, the XCAT phantom is used to simulate a chest volume at two different respiratory phases separated by 2.5 s. One phase, denoted by μ_s, serves as the preoperative CT volume, while the other phase is used to generate the biplane projection data $\mathbf{l}$. 2D/3D registration is performed by deforming μ_s such that its forward projections match the measured projections:

$$\boldsymbol{\theta}^* = \arg\min -\mathbf{NCC}(\mathbf{A}(\mathbf{h}(\boldsymbol{\theta}))\mu_s, \mathbf{l}) + \lambda \|\nabla \boldsymbol{\theta}\|_2^2, \tag{9}$$

where $\mathbf{NCC}$ denotes the normalized cross-correlation metric. A gradient-based regularization term is applied to enforce the spatial smoothness of the DVF. The DVF parameters $\boldsymbol{\theta}$

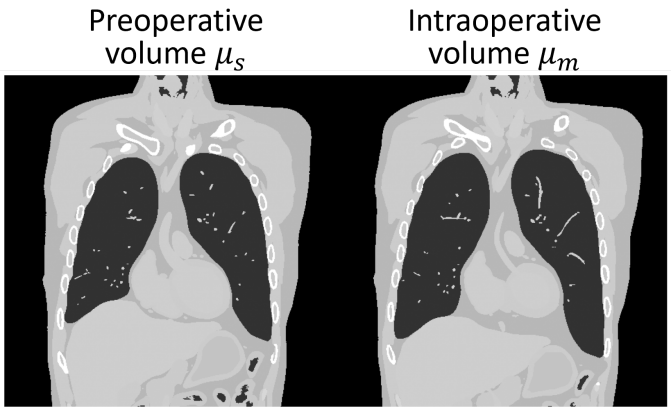


Fig. 2. Illustration of two XCAT phantoms generated with 2.5s time offset, the left and right phantom emulate preoperative volume μ_s and intraoperative volume μ_m, respectively. 2D/3D registration aims to deform μ_s to match the orthogonal projections of μ_m.

are initialized to zero. Optimization is performed using the Adam optimizer, with a step size of 2.0 mm. The optimization terminates when the NCC exceeds 0.999.

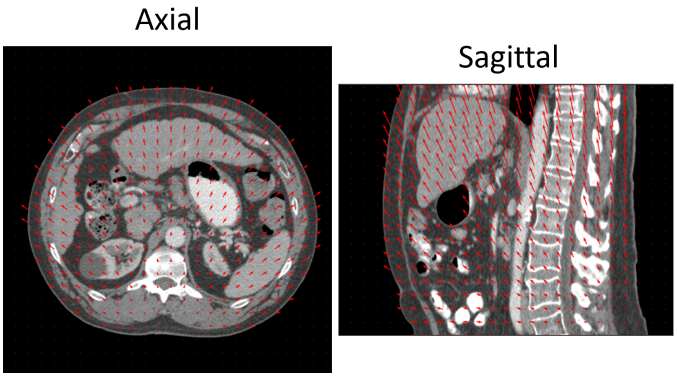


Fig. 3. Illustration of DVF used in simulating patient motion during a CBCT scan. The DVF patterns are multiplied with a time-dependent sinusoid signal to emulate periodic motion.

3) Motion Compensation using the Differentiable BackProjector: Conventional filtered backprojection (FBP) is sensitive to patient motion, often resulting in severe motion-induced artifacts. Previous work [1], [5], [6] has investigated incorporating rigid motion models directly into the backprojection for motion compensation. A similar rationale can be extended to deformable motion models. Specifically, we estimate the DVF by minimizing the gradient entropy of the motion-compensated reconstruction:

$$\boldsymbol{\theta}^* = \arg\min \mathbf{GE}(\mathbf{A}^T(\mathbf{h}'(\boldsymbol{\theta}))\mathbf{Ramp}(\mathbf{l})) + \lambda\|\nabla\boldsymbol{\theta}\|_2^2, \quad (10)$$

where $\mathbf{GE}$ and $\mathbf{Ramp}$ denote the gradient entropy and ramp filter, respectively, and a smoothness penalty is applied on the downsampled DVF. We emulate spatially varying deformable patient motion (Fig. 3). A circular CT scan with 600 projections is simulated, with the temporal motion modeled as

$$\mathbf{h}'(\mathbf{r}', t) = \mathbf{h}'(\mathbf{r}', 0)\sin\left(2\pi\frac{t}{T}\right), t = 0, 1...599 \quad (11)$$

where T denotes the motion period. The downsampled DVF parameters $\boldsymbol{\theta}$ are initialized to zero and optimized using the Adam optimizer with a step size of 1.0 mm. The optimization is terminated after 100 iterations.

III. Results

A. 2D-3D Registration

Fig. 4 presents qualitative results of deformable 2D/3D registration using the proposed differentiable projector with simulated biplane radiography data, including posterior–anterior (PA) and lateral views. The top row compares forward projections generated from the preoperative CT volume without registration, with registration, and the corresponding ground-truth projections. Without registration, noticeable anatomical misalignment is observed in both views, particularly around the diaphragm and lung boundaries. After registration, the position of the diaphragm in the projected images closely matches the ground truth. The bottom row shows the corresponding projection error maps, where blue and red indicate negative and positive differences, respectively. Without registration, pronounced structural mismatches are observed in the ribs and diaphragm, as highlighted by strong edge patterns in the error maps. In contrast, the proposed 2D/3D registration substantially reduces these errors, yielding spatially more homogeneous residuals and demonstrating effective deformable registration in both orthogonal views.

B. Motion Compensation

Fig. 5 presents representative axial slices comparing conventional FBP reconstruction without motion compensation (Mo) and the proposed motion-compensated FBP (MoCo) under different motion speeds. For slow motion ($T = 600$), conventional FBP exhibits mild but noticeable double-image artifacts on high-contrast structures such as the ribs and spleen boundaries. As motion speed increases ($T = 300$), these artifacts become more pronounced, leading to geometric distortion of air cavities and bony anatomy. With fast motion ($T = 100$), severe star-like streaking artifacts appear around small vascular and bony structures. After motion compensation, these motion-induced artifacts are substantially reduced across all motion regimes, resulting in sharper organ boundaries and improved structural fidelity. Two regions of interest (ROIs), indicated by the pink (bone) and green (soft tissue) boxes, are shown for detailed comparison. In the bone ROI, MoCo largely restores the vertebral shape and suppresses motion-induced streaking artifacts, while also recovering adjacent soft-tissue boundaries along both sides of the spine. In the soft-tissue ROI, MoCo improves the visibility of high-contrast point-like vessels and more accurately delineates contrast-enhanced hepatic veins that are difficult to resolve in the motion-corrupted reconstructions. Quantitatively, motion compensation improves image quality across all evaluated ROIs and motion speeds. Compared to conventional FBP, PSNR increases from approximately 31–35 dB to 38–43 dB, while SSIM improves from 0.83–0.92 to 0.94–0.96. The largest quantitative gains are observed in the fast-motion cases, indicating that the proposed method is particularly effective in mitigating severe motion

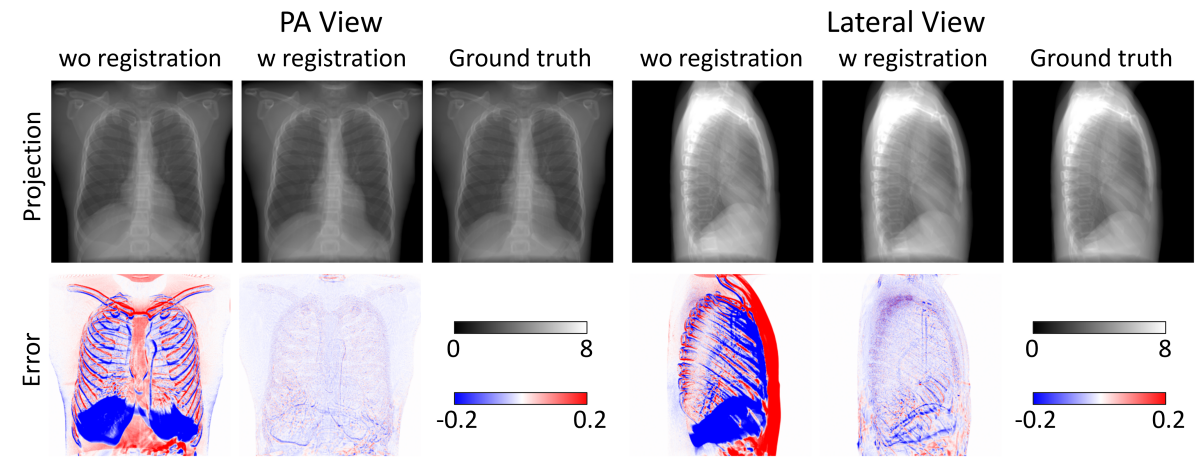


Fig. 4. Deformable 2D/3D registration using the proposed differentiable forward projector. Posterior–anterior (PA) and lateral views are shown. The top row compares forward projections generated from the preoperative CT volume without registration, with deformable registration, and ground-truth projections. The bottom row shows the projection error maps.

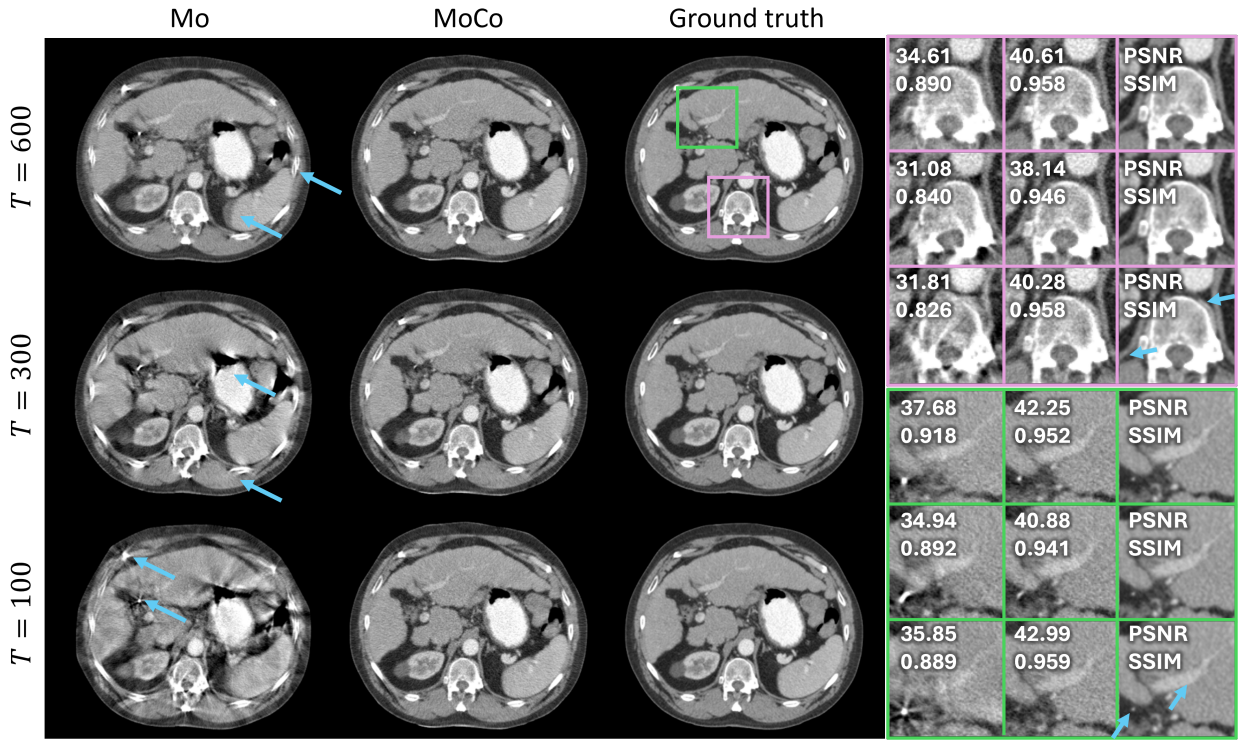


Fig. 5. Motion-compensated filtered backprojection (FBP) reconstruction using the proposed differentiable backprojector. Representative axial slices are shown in the left panel for three motion periods (top to bottom). The left, middle, and right columns show reconstructions without motion compensation (Mo), with deformable motion compensation (MoCo), and ground-truth images, respectively. Zoomed-in bone and soft tissue regions of interest (ROIs) and corresponding quantitative metrics (PSNR and SSIM) are shown on the right panel.

artifacts and preserving structural similarity under challenging motion conditions.

IV. Conclusion and Discussion

In this work, we proposed a differentiable forward and back-projection framework for gradient-based deformable motion estimation. The proposed framework was validated on two X-ray imaging tasks. In deformable 2D/3D registration, the method accurately aligned preoperative volumes to intraoperative biplane projections, substantially reducing projection mismatch in both PA and lateral views. In motion-compensated reconstruction, the proposed approach effectively mitigated motion-induced artifacts across a range of motion speeds, recovering fine structures and improving quantitative metrics. Future work will focus on evaluating the proposed framework with more sophisticated DVF parameterizations and incorporating temporal motion models to better capture complex patient motion. In addition, we plan to evaluate the proposed method on clinical patient data and investigate its integration with learning-based motion priors and auto-focus metrics.

Acknowledgments

This work is supported, in part, by NIH grants R01CA249538 and 1R01EB035908.

References

[1] S. Capostagno, A. Sisniega, J. Stayman, T. Ehtiati, C. Weiss, and J. Siewerdsen, "Deformable motion compensation for interventional cone-beam ct," *Physics in medicine & biology*, vol. 66, no. 5, p. 055010, 2021.

[2] J. Hahn, H. Bruder, C. Rohkohl, T. Allmendinger, K. Stierstorfer, T. Flohr, and M. Kachelrieß, "Motion compensation in the region of the coronary arteries based on partial angle reconstructions from short-scan ct data," *Medical physics*, vol. 44, no. 11, pp. 5795–5813, 2017.

[3] H. Huang, Y. Liu, J. H. Siewerdsen, A. Lu, Y. Hu, W. Zbijewski, M. Unberath, C. R. Weiss, and A. Sisniega, "Deformable motion compensation in interventional cone-beam ct with a context-aware learned autofocus metric," *Medical physics*, vol. 51, no. 6, pp. 4158–4180, 2024.

[4] J. Maier, S. Sawall, M. Arheit, P. Paysan, and M. Kachelrieß, "Deep learning-based cone-beam ct motion compensation with single-view temporal resolution," *Medical Physics*, 2025.

[5] X. Jiang, X. Wang, A. Uneri, W. B. Zbijewski, and J. W. Stayman, "Differentiable forward and back-projector for rigid motion estimation in x-ray imaging," *IEEE Transactions on Biomedical Engineering*, 2025.

[6] M. Thies, F. Wagner, N. Maul, H. Yu, M. Goldmann, L.-S. Schneider, M. Gu, S. Mei, L. Folle, A. Preuhs, M. Manhart, and A. Maier, "A gradient-based approach to fast and accurate head motion compensation in cone-beam ct," *IEEE Transactions on Medical Imaging*, vol. 44, no. 2, pp. 1098–1109, 2025.

Segmentation-free automatic geometry calibration method using a neural-network prediction model in CBCT

Sungho Yun, and Seungryong Cho

***Abstract*— Accurate geometry calibration is crucial for high-quality CBCT image reconstruction, yet conventional workflows remain labor-intensive and tailored to a particular phantom. In practice, many calibration pipelines rely on a specially manufactured ball phantom, ball segmentation, coordinate estimation, and subsequent projection-matrix extraction—steps that often must be repeated. We propose a segmentation-free, AI-based geometry calibration approach that directly estimates the device-specific projection matrix (P_{mat}) without use of ball phantom and segmentation process. Our formulation expresses P_{mat} as the product of intrinsic (K) and extrinsic (E) matrices. In the 6DoF setting, the model only updates the extrinsic parameters. Starting from an initial projection matrix P0, the network updates geometry parameters by minimizing the discrepancy between true projections and simulated projections generated by a differentiable forward projector. Since gantry motion is repeatable and exhibits smooth, continuous per-view variations along the trajectory, we use an MLP to jointly estimate the full set of geometry parameters across all views. We validated the method using commercial CBCT systems. A trusted reference system provided P0 and a reconstruction that served as a 3D reference object. Projections acquired on an uncalibrated device using the same reference object were then used to train the model, employing LNCC (local normalized cross-correlation) as the image-matching loss. Results demonstrate accurate and robust image reconstruction in clinical evaluation, comparable to that obtained using the vendor-calibrated projection matrix. This segmentation-free AI calibration provides a simple and practical route to P_{mat} updates, improving calibration efficiency and scalability.**

***Index Terms*—segmentation-free, geometry calibration, neural prediction model, projection matrix**

I. Introduction

Accurate geometric calibration is a fundamental prerequisite for high-quality cone-beam CT (CBCT) image reconstruction. Even small errors in the projection geometry such as detector rotations, offsets or source position inconsistencies can produce noticeable artifacts including blurring, streaking, residual misalignment, and loss of fine anatomical detail. This sensitivity becomes more pronounced in modern clinical environments where CBCT is used not only for image guidance but also for quantitative and adaptive workflows that demand consistent reconstruction quality.

Conventional CBCT geometry calibration method [1] still remains labor-intensive and tightly coupled to a particular phantom design. A typical workflow relies on a specially manufactured ball phantom, followed by ball detection and segmentation in each projection, sub-pixel coordinate estimation, and subsequent extraction or refinement of the per-view projection matrix (P). In practice, such pipelines can be difficult to scale across different commercial systems because they depend on the phantom design, scan settings, and segmentation robustness, and the whole process must be repeated whenever the system geometry is updated or revalidated.

Recent advances in differentiable network functions and learning-based optimization have opened the possibility of integrating physics-based forward models directly into data-driven calibration [2]. In this work, we propose a segmentation-free, AI-based geometry calibration framework that directly estimates device-specific projection matrices without use of a particularly designed calibration phantom and image segmentation.

To model system geometric variations in the projection matrix, we adopt 6DoF transformation model [3]. A projection matrix can be written as $P=K\cdot E$, where K collects the projection intrinsic parameters (e.g., source-to-detector distance and detector principal point/offsets), and E represents the extrinsic rigid transformation (rotation and translation) describing the acquisition pose between the object coordinate system and the source–detector geometry. Under the 6DoF formulation, the model updates only the extrinsic term using a rigid transform (3D rotation and 3D translation). We assumed the intrinsic parameters remain fixed in this work, which is legitimate when the gantry holds the source and detector firmly allowing no relative motion between the source and detector.

Starting from an initial projection matrix $P_0=K_0\cdot E_0$, our method optimizes the 6DoF parameters by minimizing the discrepancy between the measured projections and the simulated projections generated by a differentiable forward projector using the currently updated projection matrix.

Assuming that CBCT gantry motion is sufficiently repeatable and varies smoothly with the view angle, we use a multi-layer-perceptron (MLP) network to jointly predict the

This research was supported in part by Korean National Research Foundation [NRF RS-2025-25460686, Project Number: 2710003307, RS-2023-00285089]

S. Yun, and S. Cho are with the Department of Nuclear and Quantum Engineering, Korea Advanced Institution of Science and Technology, Yuseong-gu, Daejeon, 34141 Korea (e-mail: ysh3399@kaist.ac.kr). S. Cho is in correspondence of this paper. (e-mail: scho@kaist.ac.kr).

complete set of per-view geometric parameters as a continuous function across the entire trajectory. The use of MLP serves as an implicit regularizer, ensuring smooth and continuous geometric parameter updates.

We validate the proposed approach on commercial CBCT systems, using a trusted reference system to provide P_0 and a reconstruction that serves as a 3D reference object. Projections acquired on an uncalibrated device using the same reference object are then used to train the model. Our results demonstrate accurate and robust reconstruction quality comparable to vendor-provided calibration results in clinical evaluation, while substantially simplifying the overall calibration workflow. This segmentation-free AI calibration offers a practical path to "one-click" projection-matrix updates, improving calibration efficiency.

II. Method

A. Formulation of Projection matrix

Cone-beam CT (CBCT) projection geometry can be represented compactly by a 3×4 projection matrix that maps a 3D point in homogeneous coordinates to a 2D detector coordinate. Let a 3D point (e.g., a voxel center in the object/world coordinate system) be denoted by

$$\tilde{\boldsymbol{X}} = [X\ Y\ Z\ 1]^T$$

, and its corresponding detector point in homogeneous coordinates by

$$\tilde{\boldsymbol{y}} = [u\ v\ 1]^T$$

, where $(u,\ v)$ are detector coordinates. The projection matrix $P \in R^{3\times4}$ defines the mapping

$$s\,\tilde{\mathbf{y}} = \mathbf{P}\tilde{\mathbf{X}}$$

, where s is an arbitrary scale factor. Writing $\mathbf{p} = \mathbf{P}\tilde{\mathbf{X}} = [p_1\ p_2\ p_3]^{\mathrm{T}}$ the inhomogeneous detector coordinates are obtained by perspective division:

$$u = \frac{p_1}{p_3}, \qquad v = \frac{p_2}{p_3}$$

We parameterize the projection matrix using the standard factorization into intrinsic and extrinsic terms:

$$\mathbf{P} = \mathbf{K}\,\mathbf{E}, \qquad \mathbf{E} = [\mathbf{R}\ \mathbf{t}],$$

, where $\mathbf{R} \in \mathbb{R}^{3\times3}$ and $\mathbf{t} \in \mathbb{R}^3$ represent the rigid pose (rotation and translation) that relates the object/world coordinate system to the source–detector coordinate frame. The intrinsic $\mathbf{K}$ encodes the detector sampling and principal point (offset) parameters, and can be written as

$$\mathbf{K} = \begin{bmatrix} f_u & \gamma & c_u \\ 0 & f_v & c_v \\ 0 & 0 & 1 \end{bmatrix}$$

where f_u, f_v are effective focal lengths in detector pixel units (coupling source–detector distance and pixel size), $(c_u,\ c_v)$ denotes the detector principal point (detector offsets), and γ is the skew term of the detector pixel array which is typically zero.

B. 6DoF transformation for projection matrix update

We model per-view geometric deviations by updating the projection matrix through low-dimensional motion parameters. Let the initial projection matrix be

$$\mathbf{P_0} = \mathbf{K_0}\mathbf{E_0}\,, \mathbf{E_0} = [\mathbf{R_0}\ \mathbf{t_0}]$$

For each view, the goal is to estimate a motion parameter set that updates $\mathbf{P_0}$ to $\mathbf{P}$ such that the simulated projection matches the measured projection.

In the 6DoF setting, we assume the intrinsic parameters remain fixed and only update the extrinsic term by a rigid transform consisting of rotation $\mathbf{\Delta R} \in \mathbb{R}^{3\times3}$ and translation $\mathbf{\Delta t} \in \mathbb{R}^3$:

$$\mathbf{\Delta T}_{\mathrm{rig}} = \begin{bmatrix} \mathbf{\Delta R} & \mathbf{\Delta t} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}.$$

$\mathbf{\Delta T}_{\mathrm{rig}}$ denotes a per-view equivalent rigid transform that captures relative gantry geometry deviations. Using the homogeneous extrinsic matrix $\mathbf{E}_{0,h} = \begin{bmatrix} \mathbf{R_0} & \mathbf{t_0} \\ \mathbf{0}^{\mathrm{T}} & 1 \end{bmatrix}$, the updated extrinsic is

$$\mathbf{E}_h = \mathbf{E}_{0,h}\mathbf{\Delta T}_{\mathrm{rig}},\ \mathbf{E} = \mathbf{E}_h(1{:}3,1{:}4),$$

and the updated projection matrix becomes

$$\mathbf{P} = \mathbf{K_0}\mathbf{E}.$$

This parameterization captures per-view rigid perturbations of the acquisition pose while keeping the detector sampling model unchanged.

C. Neural prediction model

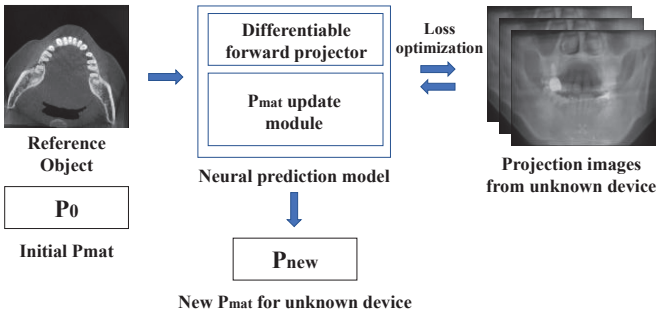


Fig. 1. Schematic illustration of the proposed neural prediction model. The model consists of a differentiable forward projector and an MLP-based projection-matrix update module.

We train a neural prediction model to estimate per-view geometric transformation parameters that update the projection matrix. Given a measured projection y_i at view i and an initial projection matrix $P_{0,i}$, the network predicts either a 6DoF rigid-motion parameter vector $\xi_i^6 = (t_{p,i}, \theta_{p,i})$, where $t_{p,i} \in \mathbb{R}^{3\times3}$, and $\theta_{p,i} \in \mathbb{R}^3$ denote the object translation and rotation. The predicted parameters are converted to an updated projection matrix P_i by applying the corresponding 6DoF transformation model.

The predictor is implemented as a multi-layer perceptron (MLP) equipped with multi-resolution hash encoding. Specifically, the view identifier (and/or continuous view angle) is mapped to a high-dimensional embedding using hash encoding, which enables compact representation of view-dependent geometric variations. The encoded features are then fed into an MLP to regress the transformation parameters ξ_i. This design allows the network to model smooth yet flexible per-view deviations with a lightweight architecture.

Training is performed in a segmentation-free and self-supervised manner. For each view, we generate a simulated

projection $\hat{y}_i$ using a differentiable forward projector $F(\cdot;P_i)$ with the current updated projection matrix:

$$\hat{y}_i = F(x; P_i)$$

where x denotes the current 3D volume estimate. The network parameters are optimized by minimizing an image-matching loss between the measured projection y_i and the simulated projection $\hat{y}_i$. We adopt the local normalized cross-correlation (LNCC) loss, which is robust to intensity scaling and bias [4]:

$$L_{LNCC} = 1 - \mathrm{LNCC}(y_i, \hat{y}_i).$$

The overall objective sums the LNCC loss over all projection views, and gradients are backpropagated through both the projection-matrix update and the differentiable forward projector, enabling end-to-end optimization of the hash-encoded MLP.

III. Materials

Experiments were conducted using commercial dental cone-beam CT (CBCT) systems, denoted as Device A and Device B. Both systems employ a full-beam, short-scan acquisition geometry with 600 projection views, a designed source-to-object distance (SOD) of 393 mm, and a source-to-detector distance (SDD) of 600 mm. The detector pixel size is 0.2 mm.

FDK Reconstructions were performed on a volume of 780×780×632 with an isotropic voxel size of 0.15 mm. Device A served as the reference device, providing a reliable initial projection matrix set $\{P_{0,i}^A\}$ derived from the vendor-calibrated geometry and a reference 3D object for forward projection. Device B was treated as the target device whose projection matrix was calibrated by the proposed segmentation-free framework.

As a reference 3D object, we used a reconstructed CBCT volume of a head phantom on the reference device (Device A). This volume, denoted as x_{ref}, was used as the fixed 3D object for generating simulated projections and for cross-device evaluation. To evaluate device-specific projection-matrix estimation under controlled conditions, we generated projection data for the target device (Device B) by forward-projecting the reference volume x_{ref} using a differentiable forward projector. Specifically, for each view i, the projection for Device B was synthesized as

$$y_i^B = F(x_{ref}; P_{i,GT}^B)$$

, where $P_{i,GT}^B$ denotes the ground-truth projection matrix of Device B (vendor-calibrated projection matrix). This setup ensures that the generated projections follow the target-device imaging geometry while using an identical underlying 3D object, enabling direct assessment of whether the proposed method can recover device-specific projection matrix from projection-domain consistency alone. In experiments, the optimization was initialized from P_0^A.

IV. Results and Discussion

In this section, we present the optimized geometry parameters under the 6DoF models and evaluate their effects on the reconstruction of the reference object using projections

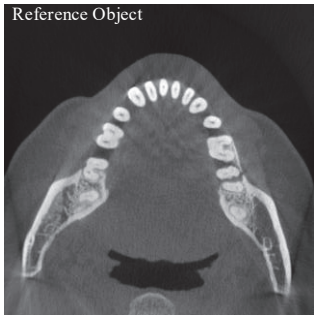


Fig. 2. FDK-reconstructed image of the reference object acquired with Device A, reconstructed using the initial projection matrix P_0^A.

from Device B. Under the 6DoF formulation, the optimization updates only the extrinsic component of the projection matrix using a rigid transformation (3D rotation and 3D translation), while keeping the intrinsic parameters fixed.

The training was run for 50 epochs until convergence. As a result, the LNCC-based loss decreased from 4.89×10^{-1} at initialization to 1.68×10^{-1} after optimization, corresponding to a 65.6% reduction in the projection-domain discrepancy. The estimated per-view transformation parameters are visualized in Fig. 3 and Fig. 4, where the optimized translation and rotation parameter sets are plotted as a function of the view index.

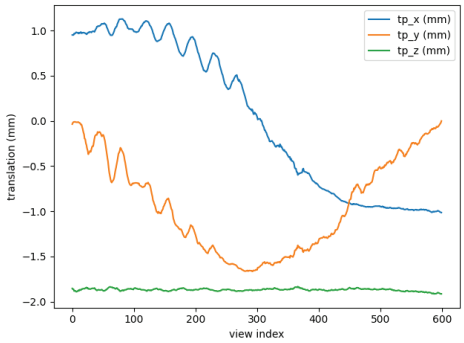


Fig. 3. Estimated translation parameters obtained from the proposed prediction model, plotted as a function of view index. The y-axis denotes the estimated translation deviation for each view.

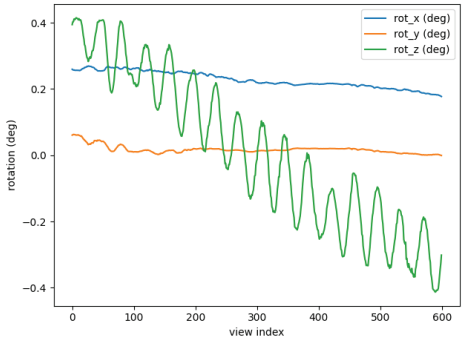


Fig. 4. Estimated rotation parameters obtained from the proposed prediction model, plotted as a function of view index. The y-axis denotes the estimated translation deviation for each view.

As shown in Fig. 3 and Fig. 4, the optimized geometry parameters exhibit a continuous drift along the gantry trajectory, and some parameters also show periodic perturbations. These trends imply the presence of substantial geometric deviations during acquisition, likely due to the

manufacturing tolerances, even among devices of the same model. This observation motivates the need for geometry calibration that directly estimates device-specific projection matrices, rather than relying solely on reconstruction using the preset system geometry.

In Fig. 5, FDK reconstructed images from Device B's projection data are presented using the initial projection matrix P_0^A and the optimized projection matrix P^B, respectively, to illustrate the improvement achieved by the proposed geometry calibration.

As shown in the figure, the reconstructed image calibrated by the proposed method yields significantly improved dental structural sharpness and geometric fidelity, with clearer delineation of tooth boundaries and reduced blurring compared to the initialization only.

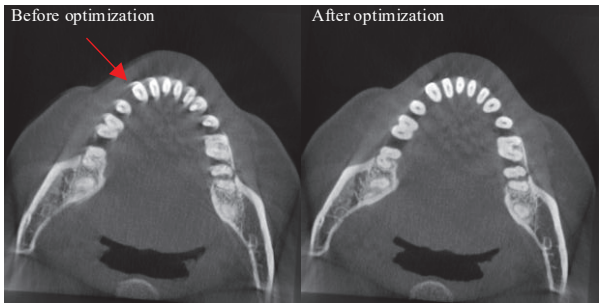


Fig. 5. The left image shows the FDK reconstructed image from Device B's projections using the initial projection matrix P_0^A, while the right image shows the reconstructed image after the proposed geometry-parameter optimization.

To further validate the estimated projection matrices P^B for practical scenarios. We conducted a clinical validation using raw projection data of a patient from Device B. As shown in Fig. 6, the reconstructed image using the proposed calibration is highly consistent with the reference reconstructed image which calibrated with the vendor-provided projection matrix, while introducing no side effects or artifacts in the reconstructed images.

In quantitative analysis, the reconstructed images generated using the vendor-calibrated (ground-truth) projection matrix were used as the reference, and SSIM, PSNR, and RMSE were compared for reconstructed images obtained with the initial projection matrix and with the proposed geometry-parameter optimization, all computed in Hounsfield unit (HU) values. The quantitative results are summarized in Table 1.

These results indicate that the proposed method achieves reconstruction performance comparable to the vendor-calibrated projection matrix in practical clinical settings. Moreover, compared to reconstruction without calibration (i.e., using the initial projection matrix), the proposed optimization yields effective improvements of 93.8% in PSNR, 31.3% in SSIM, and a 74.5% reduction in RMSE, demonstrating its practical benefit and reliability.

In future work, we will extend the framework to a 9DoF formulation that explicitly models relative source position to detector during acquisition. Following the de novo construction strategy, the updated projection matrix will be formed from intrinsic and extrinsic components for each view, enabling more flexible correction of coupled geometric variations.

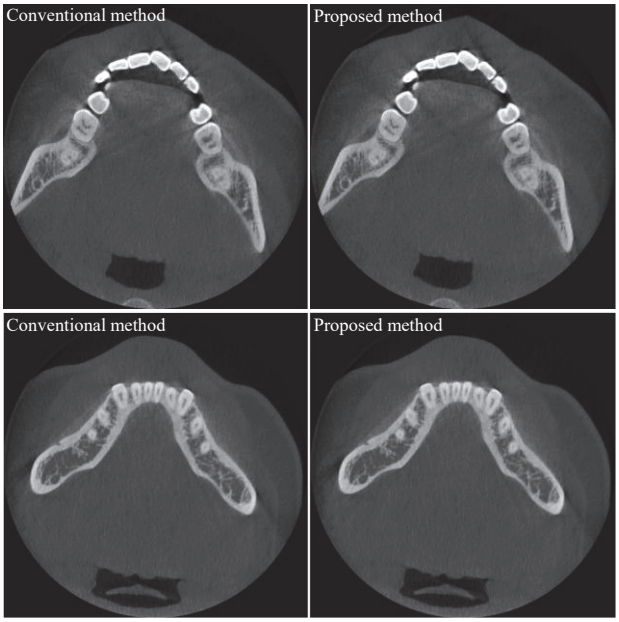


Fig. 6. The left column shows the FDK reconstructed image from Device B's clinical projections using the ground-truth projection matrix, whereas the right column shows the reconstructed image using the proposed geometry-parameter optimization.

In addition, while this study used an available P_0 initialization, we aim to generate P_0 directly from a nominal system setup (e.g., scanner design parameters) so that the proposed method can be deployed without requiring projection matrices from any reference device.

TABLE 1.
QUANTITATIVE ANALYSIS IN HU UNIT

Metric	RMSE (HU)	PSNR (dB)	SSIM
Initialization only	323.23	24.73	0.732
Proposed method	21.07	47.92	0.961

V. CONCLUSION

In this work, we present a segmentation-free, AI-based geometry calibration framework that estimates device-specific projection matrices for CBCT without requiring any segmentation. The main strength of the proposed approach lies in its simplicity and scalability: given an initial projection matrix and a 3D reference object, the framework automatically converges to a refined geometry. Overall, the framework offers a practical and robust solution for device-specific CBCT geometry calibration, and its effectiveness is demonstrated through clinical validation in real-world settings.

REFERENCES

[1] J. Hwang et al., "Geometry calibration for a dental cone-beam CT system with an offset detector," Precision Engineering, vol. 79, pp. 240–251, 2023.

[2] R. Zha, Y. Zhang, and H. Li, "NAF: Neural Attenuation Fields for Sparse-View CBCT Reconstruction," MICCAI 2022, 2022, pp. 442–452.

[3] S. Ouadah et al., "Correction of patient motion in cone-beam CT using 3D-2D registration," Physics in Medicine & Biology, vol. 62, no. 19, pp. 7613–7639, 2017.

[4] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A Learning Framework for Deformable Medical Image Registration," IEEE Transactions on Medical Imaging, vol. 38, no. 8, pp. 1788–1800, 2019

Image Reconstruction for Helical CBCT With a Laterally Shifted Flat-Panel Detector

Zhicong Yu, Amit Jain, Chuanyong Bai, Jacob Shea, and Daniel Gagnon

***Abstract*—Cone-beam computed tomography (CBCT) with a laterally shifted flat-panel detector is widely used in image-guided radiation therapy to extend the field-of-view. However, this geometry introduces significant reconstruction challenges due to combined lateral and axial data truncation, particularly in helical scans. Existing Feldkamp–Davis–Kress (FDK)-based methods with shift-invariant view completion compromise mathematical exactness for axially invariant objects. We propose a reconstruction algorithm tailored to helical CBCT with a laterally shifted flat-panel detector, incorporating normalized view completion and a 2D aperture weighting scheme to properly handle data redundancy. By back-projecting only measured data, the method preserves data fidelity and achieves exactness for axially shift-invariant objects. Numerical simulations and real-data experimental results demonstrate that the proposed approach is able to robustly produce high fidelity images.**

***Index Terms*—Helical, CBCT, Shifted Detector**

I. Introduction

Cone-beam computed tomography (CBCT) utilizing a flat-panel detector is a critical imaging modality for image-guided radiation therapy (IGRT) and interventional radiology. To accommodate a large scan field-of-view (FOV) with limited detector dimensions, systems often implement a lateral detector shift. While such offset-detector geometries are common in circular CBCT, they introduce significant challenges for image reconstruction; the resulting projection data are truncated in both the lateral and axial directions. Significant research has been dedicated to addressing these challenges, and successful reconstruction results have demonstrated the clinical viability of this configuration ([1], [2], [3], [4], [5], [6], [7]).

Recently, an imaging platform capable of providing both real-time 2D guidance [8] and 3D volumetric imaging was introduced [9] for a commercial IGRT system (Radixact, Accuray Inc., Sunnyvale, CA). The 3D imaging functionality utilizes helical acquisition protocols with a laterally shifted detector to facilitate an extended field-of-view (FOV). This geometry introduces significantly greater reconstruction complexity compared to standard circular trajectories. One existing solution for this configuration is based on the Feldkamp-Davis-Kress (FDK) framework [10], originally developed for curved detectors with marginal lateral detector shift. It employs a shift-invariant view-completion scheme, where the synthesized data are utilized for both the filtration and back-projection steps. However, this view-completion requirement breaks the FDK algorithm's exactness for axially invariant objects.

Z. Yu is with University of California, Santa Cruz. J. Shea is with Accuray Incorporated. Z. Yu, A. Jain, C. Bai, and D. Gagnon were formerly with Accuray Incorporated.

In this work, we propose a reconstruction solution for helical scans with a substantially shifted flat-panel detector [11] in the lateral direction. This solution is derived from general reconstruction theory employing view-dependent differentiation and the Hilbert transform [12], a combination well-regarded for its efficacy in handling truncated projection data. By back-projecting data strictly within the measured spatial ranges, the proposed solution preserves high data fidelity. Furthermore, the algorithm is exact for objects that are shift-invariant in the axial direction. The remainder of this paper will establish the mathematical foundations of the proposed approach and demonstrate its robustness and high fidelity image quality via both computer simulations and real data experiments.

II. Notation and Geometry

The data acquisition scheme is illustrated in Figure 1 left, employing a world coordinate system (x, y, z) defined according to the IEC 61217 standard [13]. The system consists of an X-ray source coupled to a slip ring that rotates about the y-axis, while the couch translates along the same axis (indicated by red arrows). We denote the source-to-axis distance (SAD), the source-to-detector distance (SDD), and the source view angle by R, D, and λ, respectively. We also denote the linear attenuation coefficient by $f(\underline{x})$, where $\underline{x} = (x, y, z)^\top$ represents the position vector in the world coordinate system.

The source coordinate system is centered at the X-ray source and spanned by the basis $\underline{e}_u(\lambda)$, $\underline{e}_v(\lambda)$, and $\underline{e}_w(\lambda)$ as defined in Equation (1). The detector plane is spanned by $\underline{e}_u$ (channel) and $\underline{e}_v$ (row), with its origin, O_d, defined as the piercing point of the ray from the source through the isocenter onto the detector. For simplicity, the detector is assumed to be perpendicular to the source-isocenter ray, with its channels aligned with the y-axis.

$$\begin{aligned} \underline{e}_v(\lambda) &= [0, \quad 1, \quad 0 \quad]^\top \\ \underline{e}_u(\lambda) &= [\sin(\lambda), \quad 0, \quad \cos(\lambda) \quad]^\top \\ \underline{e}_w(\lambda) &= [\cos(\lambda), \quad 0, -\sin(\lambda) \quad]^\top \end{aligned} \tag{1}$$

Let $\underline{a}(\lambda)$ be the source position and $\mathcal{L}(\lambda, u, v)$ be the X-ray pointing from $\underline{a}(\lambda)$ to the detector position (u, v). The unit vector of $\mathcal{L}(\lambda, u, v)$, denoted by $\underline{\alpha}(\lambda, u, v)$, and the corresponding line integral, denoted by $g(\lambda, u, v)$, are defined by Equations 2 and 3 as below, respectively.

$$\underline{\alpha}(\lambda, u, v) = \frac{u\,\underline{e}_u(\lambda) + v\,\underline{e}_v(\lambda) - D\,\underline{e}_w(\lambda)}{\sqrt{D^2 + u^2 + v^2}} \tag{2}$$

$$g(\lambda, u, v) = \int_0^{+\infty} f\left(\underline{a}(\lambda) + t\,\underline{\alpha}\right)\, dt \tag{3}$$

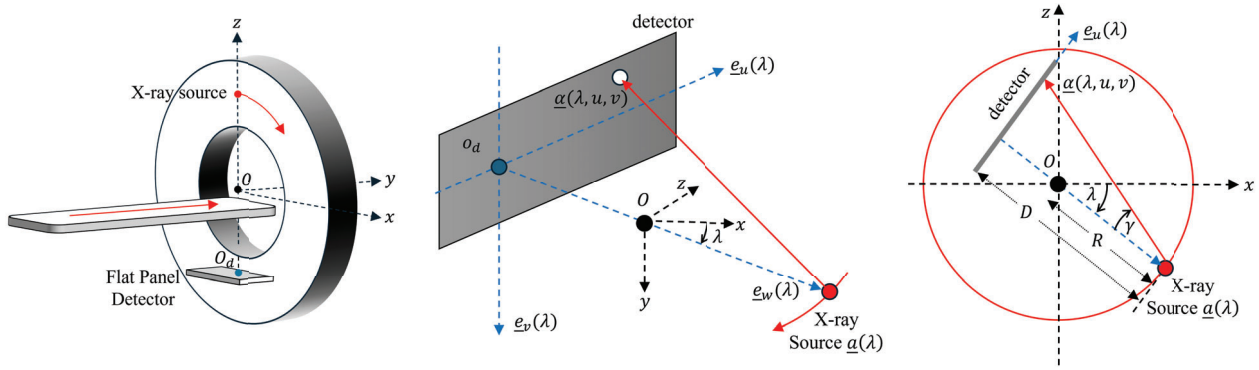


Fig. 1. Helical scan geometry and coordinate systems. Left: Helical trajectory formed by gantry rotation and couch translation, with a lateral detector shift to extend the FOV. Middle: Relationship among the world, source, and detector coordinate systems. Right: Geometry in the (x, z) trans-axial plane.

The helical source trajectory is defined over the angular range $[\lambda_s, \lambda_e]$. Let H_p denote the axial distance traveled by the source per rotation relative to the couch, where $H_p < 0$ indicates couch translation toward the positive y-axis. Given the detector axial coverage H_d at isocenter, the helical pitch is defined as $p = |H_p/H_d|$. The source axial position $h(\lambda)$ relative to the couch can be expressed as $h(\lambda) = H_p(\lambda - \lambda_s)/2\pi + h_s$, where h_s is the starting position. This leads to the helical source trajectory as defined in Equation (4) below.

$$\underline{a}(\lambda) = [R\cos\lambda, h(\lambda), -R\sin\lambda]^\top \tag{4}$$

III. Reconstruction Algorithm

We consider a point of interest, $\underline{x}$, and a source position $\underline{a}(\lambda)$. Let (u^*, v^*) be the detector coordinates of the piercing point corresponding to the line going through $\underline{x}$ and $\underline{a}(\lambda)$. By construction, the two coordinates can be obtained via the following equations.

$$\begin{aligned} u^*(\lambda,\underline{x}) &= -\frac{D(\underline{x}-\underline{a}(\lambda))\cdot\underline{e}_u}{(\underline{x}-\underline{a}(\lambda))\cdot\underline{e}_w} \\ v^*(\lambda,\underline{x}) &= -\frac{D(\underline{x}-\underline{a}(\lambda))\cdot\underline{e}_v}{(\underline{x}-\underline{a}(\lambda))\cdot\underline{e}_w} \end{aligned}. \tag{5}$$

Let $\hat{f}(\underline{x})$ be the reconstruction of $f(\underline{x})$. Let $\hat{g}(\lambda, u, v)$ be the measured line integral data. Let $\Lambda(\underline{x})$ be the set of view angles that $\underline{x}$ is visible to the detector. Our proposed reconstruction formula is expressed in Equation 6:

$$\hat{f}(\underline{x}) = -\frac{1}{2\pi}\int_{\Lambda(\underline{x})} \frac{w(\lambda,\underline{x})\, g_H(\lambda,u^*,v^*)}{(\underline{a}(\lambda)-\underline{x})\cdot\underline{e}_w}\, d\lambda \tag{6}$$

where $w(\lambda, \underline{x})$ is a 2D aperture weighting function and g_H represents the filtered projection data defined as:

$$g_H(\lambda,u,v) = h_H^{\Delta}(u) * g'_{\text{CB}}(\lambda,u,v). \tag{7}$$

Here, $h_H^{\Delta}(u)$ is the Hilbert transform kernel [12] apodized by a window function $\alpha + (1-\alpha)\cos(2\pi\sigma\Delta u)$, with $\alpha \in [0.5, 1]$.

The term g'_{CB} in Equation 7 is the pre-weighted derivative of the cone-beam (CB) projections $\bar{g}$ as defined in Equation 8 below. Here $\bar{g}$ is the view-completed projection data such that it is free of trans-axial truncation, identical to the acquired data $\hat{g}$ within the measurement range, and estimated using data from neighboring rotations outside the measurement range. We implemented the view differentiation following Noo et al. [14], using a resolution-control parameter ϵ, where smaller values correspond to higher spatial resolution.

$$g'_{\text{CB}}(\lambda,u,v) = \frac{D}{\sqrt{D^2+u^2+v^2}}\frac{\partial\bar{g}(\lambda,u,v)}{\partial\lambda}\Big|_{\underline{\alpha}\text{ fixed}}, \tag{8}$$

The complete procedure for implementing Equation 6 is detailed in Algorithm 1. Thus far, all components have been defined with the exception of the view completion step and the 2D aperture weighting, w, used for redundancy handling. The remainder of this subsection clarifies these two aspects.

Algorithm 1 CB FBP: Laterally Shifted Flat-Panel Detector

Require: $\hat{g}(\lambda, u, v)$ {Inputs}
Ensure: $\hat{f}(\underline{x})$ {Output}
1: View completion using data from conjugate views: $\hat{g} \to \bar{g}$
2: View differentiation and CB weighting: $\bar{g} \to g'_{CB}$ {Eq.8}
3: Hilbert transform for each row: $g'_{CB} \to g_H$ {Eq.7}
4: Back-projection with aperture weighting: $g_H \to \hat{f}$ {Eq.6}

The view completion process (Steps A-E as detailed below) estimates missing rays using conjugate data from adjacent rotations. The estimation strategy depends on the helical pitch: for small pitches, data from both the preceding and subsequent rotations are utilized; for large pitches, where data may only be available from a single adjacent rotation, that available data is used directly. To ensure image quality, the algorithm maintains continuity across varying pitch values and enforces smooth transitions when merging measured and estimated data along the channel direction.

Step A: Identify a reference plane Π that contains both $\underline{x}$ and the current view $\underline{a}(\lambda)$ and is oriented parallel to the rotation v-axis.

Step B: Identify conjugate rays from neighboring rotations in the reference plane Π.

Step C: Normalize measurements of the identified conjugate rays using the following equation,

$$\tilde{g}(\bar{\lambda},\bar{u},\bar{v}) = \frac{\sqrt{u^{*2}+v^{*2}+D^2}}{\sqrt{\bar{u}^2+\bar{v}^2+D^2}}\hat{g}(\bar{\lambda},\bar{u},\bar{v}),$$

where $\tilde{g}((\bar{\lambda}, \bar{u}, \bar{v}))$ and $\hat{g}(\bar{\lambda}, \bar{u}, \bar{v})$ are the normalized and measured conjugated rays.

Step D: Interpolate the value of missing rays using the normalized conjugate ray(s) $\tilde{g}$.

Step E: The interpolated data is merged with the measured data so that it remains unchanged within the measured range and blends smoothly in a buffer region along the channel direction.

The 2D aperture weighting is defined in Equation 9, where $\Phi(\lambda, \underline{x})$ represents the set of detector coordinates corresponding to all redundant rays of the current view. Unlike conventional 1D aperture weighting along the row direction (e.g., [15]), the proposed function w is defined using both channel and row positions of redundant rays. Here, w_u and w_v are 1D single- and double-sided smooth transition functions respectively, such that w smoothly transitions from unity at the center to zero at the detector row edges and the isocenter-proximal channel edge. This 2D aperture weighting effectively excludes filtered projection data outside the measurement range from the back-projection process.

$$w(\lambda, \underline{x}) = \frac{w_u(u^*)\, w_v(v^*)}{\sum_{(u,v)\in\Phi(\lambda,\underline{x})} w_u(u)\, w_v(v)}, \tag{9}$$

It shall be noted that, under the assumption of an ideal implementation of Step E of the view-completion process, the proposed reconstruction achieves exactness for axially shift-invariant objects because of the normalization in Step C in view completion and the 2D aperture weighting functions.

IV. Computer Simulations and Experiments

Computer simulations were performed using analytically defined phantoms. The SAD and SDD were 1000 mm and 1500 mm, respectively. Line integrals were computed for a point source by averaging four rays per detector pixel, corresponding to the pixel corners. The detector configuration consisted of a 960×320 (channels $\times$ rows) matrix, laterally shifted to achieve a 54 mm overlap at the isocenter. Data acquisition used 900 views per rotation and a helical pitch of 0.6. All images were reconstructed using a kernel of $(\alpha, \epsilon) = (0.5, 0.01)$ with an isotropic voxel pitch of $0.8\times0.8\times0.8$ mm^3. Here α is the coefficient of the apodization window used in the Hilbert transform kernel involved in Equation 7, and ϵ is the spatial-controlling parameter of the view-differentiation step.

To evaluate the proposed reconstruction algorithm, two analytical phantoms were employed. The first phantom, which is referred as the CNST phantom, comprising three cylinders representing water, air, and cortical bone (Figure 2, left), verifies the reconstruction algorithm's mathematical exactness for axially shift-invariant objects. The second, the FORBILD thorax phantom, is employed to demonstrate the robustness of the proposed algorithm to both noise (Poisson and electronic) as well as to complex anatomical structures.

We also acquired helical data of an anthropomorphic pelvic phantom from a prototype system equipped with a laterally shifted flat-panel detector. Projection data were acquired at 150 kVp and 200 mA with a 5 ms pulse width, with imaging chain comprising a combination of flat and bowtie filters. The SAD and SDD were 1027.4 mm and 1543.0 mm, respectively. The axial beam aperture at the isocenter was 35 mm, and the detector was shifted to the positive u-axis such that there was a 68 mm overlap at the isocenter along the channel direction. The detector matrix within the exposure area comprised 960×117 pixels with a pitch of 0.45×0.45 mm^2. We used 600 views per rotation and a helical pitch of 0.75. Pre-processing included gain, offset, bad pixel, lag, and scatter corrections. Images were reconstructed with a kernel of $(\alpha, \epsilon) = (0.5, 0.01)$ with an isotropic voxel pitch of $0.8 \times 0.8 \times 0.8$ mm^3 and a 3.2 mm slice thickness. The results were compared against reference images obtained from a clinical PET/CT scanner.

V. Results

Reconstruction results of the CNST phantom are presented in Figure 2. Using ground truth projections for missing rays confirms the algorithm's mathematical exactness; see the middle image in Figure 2 and the blue profile in Figure 3. In practice, the object is unknown and the the proposed view completion method shall be employed. The resulting image (Figure 2, right) and the corresponding profile (red in Figure 3) exhibit high fidelity, with only minor CT number deviations (up to a few HU) near the merging region. The robustness of the algorithm is further confirmed by its ability to maintain smooth, artifact-free reconstructions under challenging conditions, including noisy projections of the FORBILD thorax phantom (Figure 4) and real-data reconstructions of an anthropomorphic pelvic phantom from the prototype system (Figure 5).

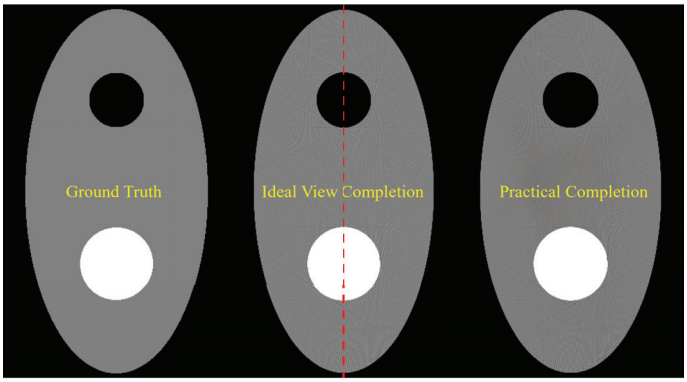


Fig. 2. Image of the CNST phantom comprising cylinders of water, air, and bone. Left: Ground truth. Middle: Reconstruction using ideal view completion. Right: Reconstruction using practical view completion. A profile comparison along the red dashed line is shown in Figure 3. Display: L/W = 0/100 HU.

VI. Discussion

The proposed reconstruction algorithm is dose efficient because it utilizes all detected X-rays for back-projection across all image pixels. The algorithm addresses substantial data truncation caused by the limited detector size and one-sided lateral shift. The high-fidelity image quality is attributed to the novel view completion and 2D aperture weighting steps; specifically, view completion improves the accuracy of the filtering process, while the weighting function successfully excludes filtered data outside the measurement area.

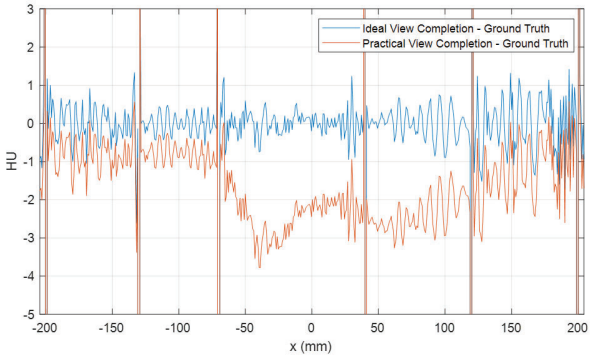


Fig. 3. The blue difference profile along the red dashed line in Figure 2 demonstrates the reconstruction algorithm's mathematical exactness under ideal view completion. In practical applications where the object is unknown, achieving perfect view completion from conjugate views is nearly impossible; consequently, the red difference profile exhibits minor HU discrepancies, typically within a few HU, near the isocenter-proximal detector edge channels.

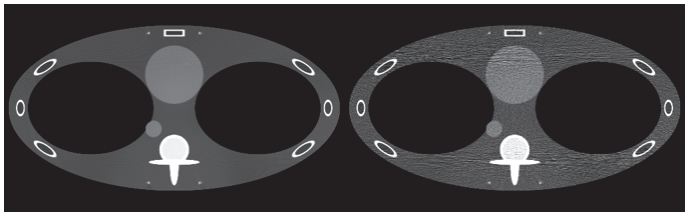


Fig. 4. Image reconstructions of the FORBILD thorax phantom under noise-free (left) and noisy (right) conditions. Poisson noise was simulated using an exposure of 100k counts per view, supplemented with additive white Gaussian noise with a mean of 10. Display window: L/W = 60/300 HU.

Conversely, our investigation reveals that the lateral displacement of the flat-panel detector restricts the permissible helical pitch to approximately half that of a standard helical CT acquisition. It is also important to understand the impact of the detector overlap around the rotation axis on image quality. A comprehensive evaluation of these parameters will be reported in our future work.

VII. Summary

We have developed an analytical reconstruction algorithm for helical CT scans utilizing a laterally shifted flat-panel detector. The algorithm utilizes all visible X-rays and thus is dose efficient. It achieves mathematical exactness for axially shift-invariant objects under ideal view completion conditions. Validations through both numerical simulations and experimental prototype data demonstrate that the proposed reconstruction algorithm yields high-fidelity reconstructions with robust artifact suppression. This approach effectively addresses the substantial data truncation in both axial and trans-axial directions, offering a computationally efficient solution for expanding the effective helical scan field-of-view in IGRT.

Acknowledgment

This work is based on preliminary research; the algorithms and results described herein may differ from current or future commercial implementations. The first author gratefully acknowledges data and tool support from Accuray.

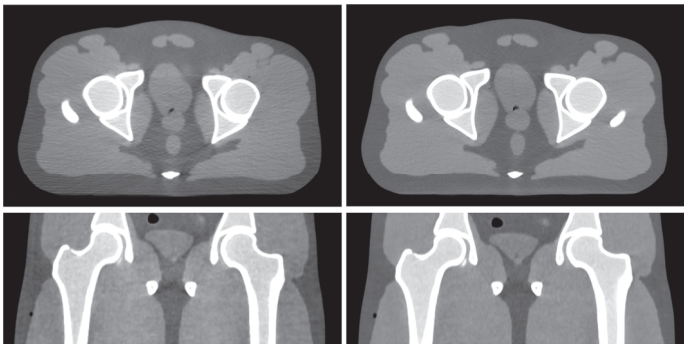


Fig. 5. Comparison of the pelvic phantom reconstructions: (left) Image reconstructed using the proposed algorithm from prototype system data with a laterally shifted detector (68 mm overlap, 0.75 pitch); (right) clinical PET/CT reference scan. The prototype reconstruction employed basic data corrections and a smooth kernel. This comparison is intended to provide a general visual assessment rather than a rigorous, dose-matched evaluation. Display window: L/W = 0/800 HU.

References

[1] P. S. Cho, R. H. Johnson, and T. W. Griffin, "Cone-beam CT for radiotherapy applications," *Phys. Med. Biol.*, vol. 40, no. 11, pp. 1863–1883, 1995.

[2] P. S. Cho, A. D. Rudd, and R. H. Johnson, "Cone-beam CT from width-truncated projections," *Comput. Med. Imag. Grap.*, vol. 20, no. 1, pp. 49–57, 1996.

[3] G. Wang, "X-ray micro-CT with a displaced detector array," *Med. Phys.*, vol. 29, no. 7, pp. 1634–1636, 2002.

[4] L. Yu, C. Pelizzari, X. Pan, H. Riem, P. Munro, and W. Kaissl, "Application of asymmetric cone-beam CT in radiotherapy," in *Nucl. Sci. Symp. Conf. Rec., IEEE*, vol. 5. IEEE, 2004, pp. 3249–3252.

[5] H. Kunze and F. Dennerlein, "Cone beam reconstruction with displaced flat panel detector," in *Proceeding of the 10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Beijing, PR China*, 2009, pp. 138–141.

[6] D. Schäfer and M. Grass, "Cone-beam filtered back-projection for circular x-ray tomography with off-center detector," in *Proceeding of the 10th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Beijing, PR China*, 2009, pp. 86–89.

[7] D. Schäfer, M. Grass, and P. van de Haar, "FBP and BPF reconstruction methods for circular x-ray tomography with off-center detector," *Med. Phys.*, vol. 38, no. S1, 2011.

[8] Accuray Incorporated, *Synchrony on the Radixact System*, v1.0 ed., Accuray Incorporated, Sunnyvale, CA, USA, 2017, white Paper.

[9] ——, *ClearRT Helical kVCT Fan-Beam Imaging: Benefits Guide*, v3.0 ed., Accuray Incorporated, Sunnyvale, CA, USA, 2021, white Paper.

[10] A. A. Zamyatin, K. Taguchi, and M. D. Silver, "Helical cone beam CT with an asymmetrical detector," *Med. Phys.*, vol. 32, no. 10, pp. 3117–3127, 2005.

[11] Z. Yu and D. Gagnon, "Helical cone-beam computed tomography imaging with an off-centered detector," U.S. Patent 11,179,132, Nov. 23, 2021.

[12] F. Noo, J. Pack, and D. Heuscher, "Exact helical reconstruction using native cone-beam geometries," *Phys. Med. Biol.*, vol. 48, no. 23, p. 3787, 2003.

[13] International Electrotechnical Commission, "IEC 61217:2011 Radiotherapy equipment - Coordinates, movements and scales," International Electrotechnical Commission, Geneva, Switzerland, Standard IEC 61217:2011, 2011. [Online]. Available: webstore.iec.ch

[14] F. Noo, S. Hoppe, F. Dennerlein, G. Lauritsch, and J. Hornegger, "A new scheme for view-dependent data differentiation in fan-beam and cone-beam computed tomography," *Phys. Med. Biol.*, vol. 52, no. 17, pp. 5393–5414, 2007.

[15] K. Stierstorfer, A. Rauscher, J. Boese, H. Bruder, S. Schaller, and T. Flohr, "Weighted FBP—a simple approximate 3D FBP algorithm for multislice spiral ct with good dose usage for arbitrary pitch," *Phys. Med. Biol.*, vol. 49, no. 11, p. 2209, 2004.

Normalized Metal Artifact Reduction using Physics-Inspired Deep Prior Images

Anton Kabelac, Andreas Specovius, and Marc Kachelrieß

***Abstract*—Metal inserts within a patient typically induce image artifacts in the CT images. Artifacts manifest as bright and dark streaks.**

In this work we present a deep learning-based physics-inspired network that builds on top of the established classical normalized metal artifact reduction (NMAR) algorithm. The metal artifact afflicted image is preprocessed to reduce metal-induced image artifacts by a U-Net. The U-Net receives additional spatial information about metal artifacts, inferred from beam hardening considerations. This artifact-reduced image is then used for normalization before linear interpolation to create a final artifact reduced image. On simulated data, the proposed method outperforms the classical iMAR algorithm.

***Index Terms*—Metal artifacts, artifact reduction, deep learning based artifact reduction.**

I. Introduction

COMPUTED tomography (CT) is an established quantitative diagnostic technique that produces precise volumes of material attenuation in seconds at sub millimeter spatial resolution. However, in cases of metal objects being present in the scan, artifacts in the reconstructed images may make a medical diagnosis based on CT images difficult and sometimes impossible. Metal artifacts are caused by dense, high Z materials in the field of measurement that lead to a much higher X-ray attenuation than soft tissue or bones. This high attenuation results in scatter artifacs, beam hardening, photon starvation, and edge gradient effects.

Most classical and deep learning-based metal artifact reduction (MAR) methods proposed over the last decades roughly fall into one of the following categories: iterative [1], image domain [2], sinogram domain or dual domain methods [3]. In particular, sinogram domain inpainting methods have been widely used and form the basis of many vendor implementations.

Sinogram domain-based inpainting techniques commonly employ dirrect inpainting of the metal trace in the sinogram. However, such approaches tend to produce nonsmooth interpolation boundaries and are physically inconsistent, which, in turn, can cause notable secondary artifacts in the metal artifact-reduced CT images. To mitigate this problem, the normalized metal artifact reduction (NMAR) [4] method based on linear interpolation constrained by a prior image was proposed and is today's gold standard of metal artifact reduction. Further, the frequency split MAR (FSMAR) method [5] introduced an image-based frequency-splitting approach to recover high frequency features that might have been lost during the normalized interpolation process.

However, NMAR relies on the quality of the prior image to normalize the sinogram before interpolation. If there are artifacts in the prior image, they cannot be corrected for and will be reintroduced during denormalization. To mitigate this effect and improve the quality of the prior image, several methods have been introduced to enhance the quality of the prior image. These range from deep learning methods like leveraging a neural attenuation field as prior image [3] to iterative update schemes that optimize the prior image calculation [6].

In this work, we propose to deploy a physics-driven U-Net for improved prior image generation that is aware of, e.g., the beam hardening artifacts.

II. Materials and Method

Metal artifacts appear as bright and dark bands that can span large distances across the reconstructed image. Correcting such artifacts with a U-Net can be challenging because of the limitied size of the receptive field. To mitigate this, we provide additional information about regions most affected by beam hardening through an additional beam hardening map as input to the network and an additional beam-hardening loss term that emphasizes voxels and regions affected by metal artifacts.

A. Beam-Hardening Awareness

Following reference [7], we decompose the energy-dependent spatial distribution of the linear attenuation coefficient $\mu(E, \boldsymbol{r})$ into contributions $\mu_1(E, \boldsymbol{r})$ from soft tissue (and bone) and $\mu_2(E, \boldsymbol{r})$ from metal:

$$\mu(E, \boldsymbol{r}) = \mu_1(E, \boldsymbol{r}) + \mu_2(E, \boldsymbol{r}). \tag{1}$$

E is the photon energy and $\boldsymbol{r} = (x, y, z)$ the voxel position. We further use decomposition $\mu_i(E, \boldsymbol{r}) = \psi_i(E) f_i(\boldsymbol{r})$ of the linear attenuation coefficient into an energy dependency

A. Kabelac, and Prof. Dr. M. Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls-University, Heidelberg, Germany. Dr. A. Specovius and A. Kabelac are with Siemens Healthineers, Forchheim, Germany.
Corresponding author: anton.kabelac@dkfz.de

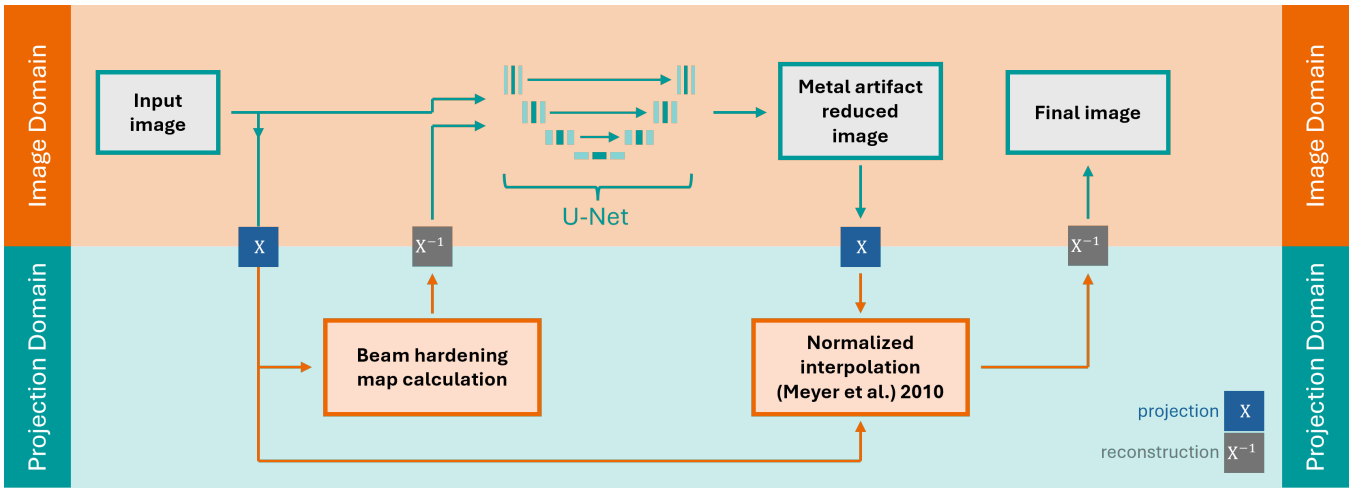


Fig. 1. Flow chart of the proposed algorithm. The U-Net receives the metal-afflicted image as well as spatial information about the metal artifacts, inferred from beam hardening considerations (beam hardening map), resulting in two input channels.

$\psi_i(E)$ and a spatial dependency $f_i(\boldsymbol{r})$. Index $i = 1$ corresponds to water (and bone) and $i = 2$ corresponds to metal. We obtain:

$$\begin{aligned}\mu(E,\boldsymbol{r}) &= f_1(\boldsymbol{r})\psi_1(E) + f_2(\boldsymbol{r})\psi_2(E)\\ &= (f_1(\boldsymbol{r}) + f_2(\boldsymbol{r}))\,\psi_1(E) + f_2(\boldsymbol{r})\,(\psi_2(E) - \psi_1(E))\\ &= \hat{f}_1(\boldsymbol{r})\psi_1(E) + f_2(\boldsymbol{r})\hat{\psi}_2(E). \qquad (2)\end{aligned}$$

Let X be the X-ray transform, i.e. the forward projection operation, and let $\hat{p}_1 = \mathsf{X}\hat{f}_1(\boldsymbol{r})$ and $p_2 = \mathsf{X}f_2(\boldsymbol{r})$. Let us further assume, we start from the water-precorrected initial image $f_0(\boldsymbol{r})$, whose X-ray transform would be p_0.

Using the standard polychromatic forward model with detected X-ray spectrum $w(E)$, that is normalized to unit area, we find

$$\int w(E)\, e^{-p_0\psi_1(E)}\, dE = \int w(E)\, e^{-\hat{p}_1\,\psi_1(E) - p_2\,\hat{\psi}_2(E)}\, dE.$$

This defines $\hat{p}_1$ as a (local, analytic) function of (p_0, p_2). Because the data are water-precorrected, the boundary condition for rays that does not intersect metal is $\hat{p}_1(p_0, 0) = p_0$. A Taylor expansion of $\hat{p}_1(p_0, p_2)$ around $(p_0, p_2) = (0, 0)$ yields

$$\hat{p}_1(p_0,p_2) = p_0 + c_{01}\,p_2 + c_{11}\,p_0p_2 + c_{02}\,p_2^2 + \mathcal{O}\big(\|(p_0,p_2)\|^3\big)\ .$$

The non-linear terms represent the beam hardening artifacts and correspond to the images

$$f_{11}(\boldsymbol{r}) = \mathsf{X}^{-1}p_0p_2 \quad f_{02}(\boldsymbol{r}) = \mathsf{X}^{-1}p_2^2,$$

with X^{-1} being the inverse X-ray transform, i.e. the reconstruction operator.

The idea of utilizing this additional beam hardening information is twofold: the importance map f_{02} is provided to the U-Net as an additional input channel. In addition, we employ a beam-hardening–inspired loss by rescaling the importance maps f_{11} and f_{02} to the range $[0, 1]$ and using them as spatial weights in a pixel-wise loss. An example of such a weighting map is shown in Fig. 2.

B. Deep-Normalized MAR Method

The general scheme of the proposed method can be seen in Fig. 1. We first train a U-Net on simulated metal artifacts to estimate a metal artifact-reduced output. To avoid hallucinations introduced by the U-Net, the metal artifact-reduced image is not used directly as the final reconstruction; instead, it is forward projected and used to normalize the input sinogram in a subsequent NMAR step. Potential hallucinations in the NMAR prior image are less critical, since the prior is used only to compute the sinogram domain normalization compared to being taken as the final output.

C. Datasets and Metal Artifact Simulation

In order to enable a supervised training scheme for the U-Net, we simulated metal artifacts. For the simulation, the abdomen atlas was used [8]. It comprises 8,448 CT volumes with a total of 3.2 million CT slices. In addition, we used 100 binary implant masks (90 training, 10 test),

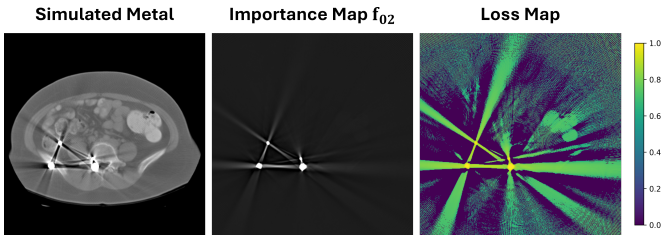


Fig. 2. Simulated metal image, beam hardening importance map and loss map. Image centered with $C = 50$ HU, $W = 1200$ HU.

whose shape resembles real metal implants such as hip replacements, rods or screws. For each simulation, a mask and one of four materials, gold (Au), titanium (Ti), cobalt (Co), or a stainless-steel alloy with nominal composition $(0.68\,\mathrm{Fe}, 0.18\,\mathrm{Cr}, 0.10\,\mathrm{Ni}, 0.02\,\mathrm{Mo}, 0.01\,\mathrm{C}, 0.01\,\mathrm{Mn})$ were sampled at random.

Our simulation takes into account all relevant physical effects such as beam hardening, photon starvation, scatter radiation, and quantum noise. First, the input image is segmented into water- and bone-equivalent components by thresholding. We then determine a position of the metal mask by maximizing its overlap with the bone-equivalent component, using affine transformations (translation, rotation). In the next step, polychromatic forward projections of the three components (water, bone, metal) are performed with a spectrum from 50 kV to 125 kV. Let p_{W}, p_{B}, and p_{M} be the intersection lengths (forward projection) of the water (soft tissue) background, bone structures and metal objects, then

$$q = -\ln \int w(E)\, e^{-p_{\mathrm{W}}\psi_{\mathrm{W}}(E) - p_{\mathrm{B}}\psi_{\mathrm{B}}(E) - p_{\mathrm{M}}\psi_{\mathrm{M}}(E)}\, dE$$

is the sinogram that is simulated. It is then used for scatter simulation using the pep model [9]. Finally, Poisson noise is added (in intensity domain) and then the sinogram is normalized and log-transformed.

D. Network Architecture and Training

We employ a 2D U-Net operating on axial slices with five resolution levels and $[16, 32, 64, 128, 256]$ feature channels. The network takes as input a two-channel image consisting of the metal artifact afflicted slice and the normalized importance map $f_{02}(\boldsymbol{r})$ and predicts a single-channel metal artifact reduced image. Each level uses two convolutional layers with leaky ReLU activations and instance normalization, followed by 2×2 max-pooling in the encoder and bilinear upsampling with skip connections in the decoder; a final 1×1 convolution yields a linear output. Training is performed with Adam (starting learning rate 10^{-4}, batch size 48), using mixed precision on an NVIDIA RTX 4090 GPU and a `ReduceLROnPlateau` scheduler.

The loss combines a standard mean-squared error with a beam hardening–aware term that emphasizes voxels close to metal via the normalized maps $\tilde{f}_{11}$ and $\tilde{f}_{02}$:

$$w(\boldsymbol{r}) = 1 + \alpha\, \tilde{f}_{11}(\boldsymbol{r}) + \beta\, \tilde{f}_{02}(\boldsymbol{r}), \tag{3}$$

$$\mathcal{L} = \frac{1}{|\Omega|} \sum_{\boldsymbol{r}\in\Omega} \big(\hat{f}(\boldsymbol{r}) - f_{\mathrm{ref}}(\boldsymbol{r})\big)^2 \Big(1 + \lambda_{\mathrm{EBHC}}\, w(\boldsymbol{r})\Big), \tag{4}$$

where $\hat{f}$ is the network prediction, f_{ref} the metal-free ground truth image, and Ω the image domain.

III. Results

Qualitative results for three test cases are shown in Fig. 3. Strong streak artifacts are visible in the simulated metal reconstructions. The proposed beam hardening–aware deepMAR pipeline more consistently suppresses both bright and dark streaks around the implants and better preserves soft-tissue contrast across the different anatomies, compared to classical iMAR [6]. Voxel-wise errors were computed only inside a body mask, excluding air and metal. Table I reports the mean masked MAE for each method (averaged over 29 test patients). In addition, we calculated a slice-wise win rate comparing the proposed method against the native iMAR reconstruction: for each axial slice, we compute the masked MAE and count a win if the proposed method achieves a lower masked MAE than iMAR on that slice. The win rate is the fraction of wins among all valid slices (i.e., slices with at least one voxel inside the body mask). Overall, the proposed method achieved a slice-wise win rate of 99.15 %.

IV. Conclusion

We proposed a novel metal artifact reduction method that combines deep learning–based metal artifact reduction techniques with established classical algorithms. We employed a U-Net that additionally receives beam hardening information related to the metal artifacts. This allows the U-Net to generate a precise initial metal artifact–reduced prediction. Since this prediction could, in theory, lead to hallucinations, it is not used directly. Instead, it is used for normalization within the established normalized metal artifact reduction (NMAR) algorithm. The network was trained on simulated data. On simulated test data, the proposed method was able to outperform iMAR, a classical iterative implementation of the NMAR algorithm. Future work will include testing the proposed approach on real data rather than simulated data.

V. Acknowledgements

The projection simulation software RayConStruct-PS and the image reconstruction software RayConStruct-IR were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] R. Pua et al. “A Pseudo-Discrete Algebraic Reconstruction Technique (PDART) Prior Image-Based Suppression of High Density Artifacts in Computed Tomography”. In: *Nucl. Instrum. Methods Phys. Res. Sect. A* 840 (2016), pp. 42–50.

TABLE I
Masked MAE and standard deviation of the absolute error within the patient body excluding the metal part itself.

Method	MAE (masked)
U-Net Corrected	12.05 HU
Proposed Method	10.50 HU
iMAR [6]	22.35 HU

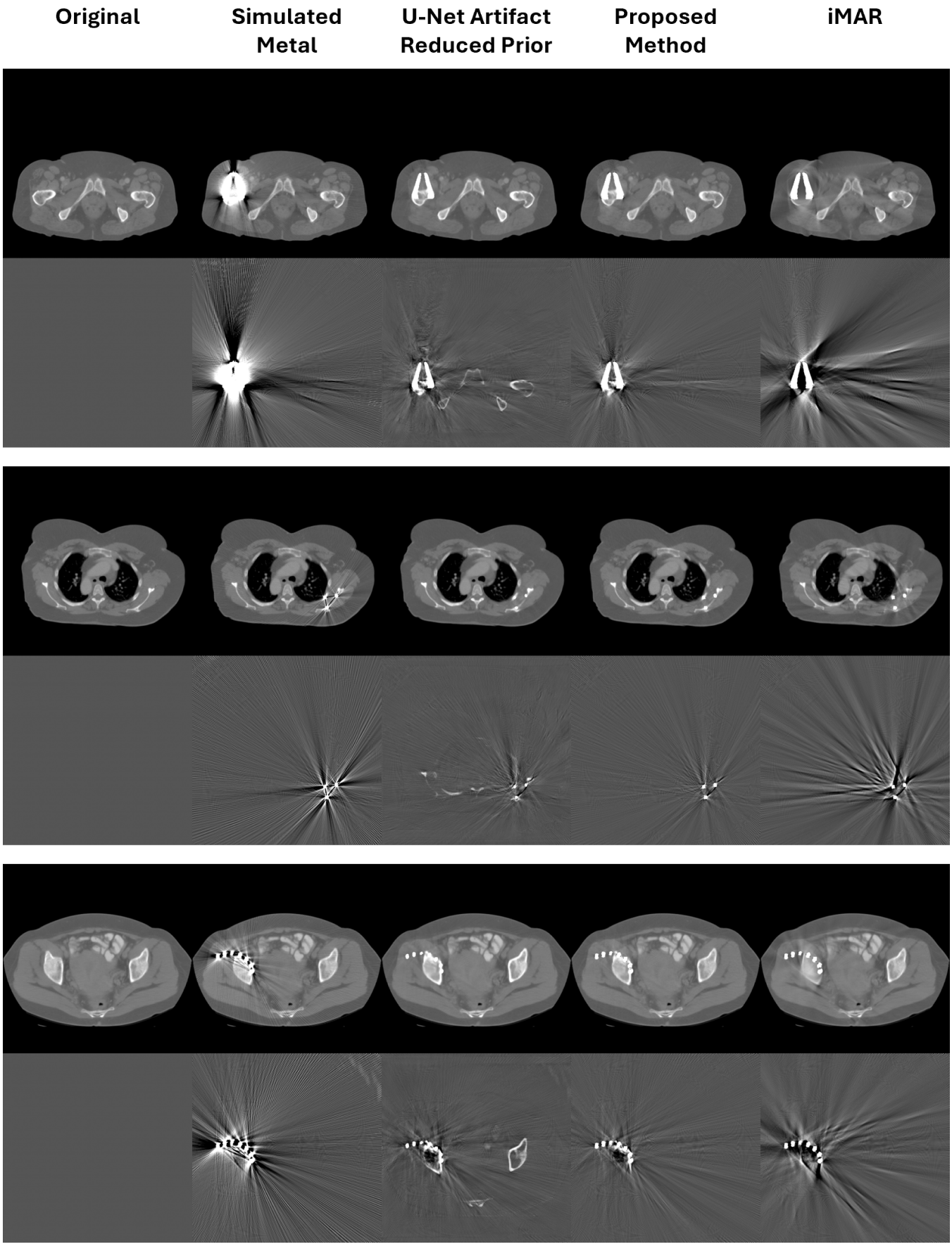


Fig. 3. Results of metal artifact reduction. The top row always displays the images, and the bottom row shows the difference to the ground truth artifact-free images. Left to right: ground truth artifact-free images, simulated metal artifact images, metal artifact–reduced images (U-Net output), results of the proposed algorithm using the U-Net–corrected image for normalization, iMAR [6] for comparison. Image centered with $C = 50$ HU, $W = 1200$ HU and difference images with $C = 50$ HU, $W = 300$ HU.

[2] Y. Zhang and H. Yu. “Convolutional Neural Network Based Metal Artifact Reduction in X-Ray Computed Tomography”. In: *IEEE Trans Med Imaging* 37.6 (June 2018), pp. 1370–1381.

[3] J. Lee et al. “X-Ray CT Metal Artifact Reduction Using Neural Attenuation Field Prior”. In: *Med. Phys.* 52.7 (July 2025). Epub 2025 Apr 30, e17859.

[4] E. Meyer et al. “Normalized Metal Artifact Reduction (NMAR) in Computed Tomography”. In: *Med. Phys.* 37.10 (Oct. 2010), pp. 5482–5493.

[5] E. Meyer et al. “Frequency Split Metal Artifact Reduction (FSMAR) in Computed Tomography”. In: *Med. Phys.* 39.4 (Apr. 2012), pp. 1904–1916.

[6] M. Kachelrieß and A. Krauss. *Iterative Metal Artifact Reduction (iMAR): Technical Principles and Clinical Results in Radiation Therapy*. White Paper A91RO-07052-09C1-7600. Erlangen, Germany: Siemens Healthcare GmbH, 2016.

[7] Y. Kyriakou et al. “Empirical Beam Hardening Correction (EBHC) for CT”. In: *Med. Phys.* 37.10 (Oct. 2010), pp. 5179–5187.

[8] W. Li et al. “AbdomenAtlas: A Large-Scale, Detailed-Annotated, & Multi-Center Dataset for Efficient Transfer Learning and Open Algorithmic Benchmarking”. In: *Med Image Anal* 97 (Oct. 2024), p. 103285.

[9] B. Ohnesorge, T. Flohr, and K. Klingenbeck–Regn. “Efficient Object Scatter Correction Algorithm for Third and Fourth Generation CT Scanners”. In: *European Radiology* 9 (1999), pp. 563–569.

Self-Operating Geometry Calibration for Cone-Beam Computed Tomography Using Bayesian Optimization

Jiyong Shim, and Hyosung Cho*

***Abstract*—Accurate scan geometry is crucial for achieving high-quality three-dimensional (3D) reconstructions in cone-beam computed tomography (CBCT). This study introduces a practical, self-operating geometric calibration method that estimates key scan parameters—transverse shift (δu) and longitudinal shift (δv)—without requiring dedicated calibration phantoms. The proposed method employs Bayesian Optimization (BO) to determine optimal detector offsets by minimizing a hybrid objective function that combines reprojection error and image-roughness penalty. By modeling the objective function as a Gaussian Process (GP), the algorithm efficiently explores the parameter space to find a near-optimal solution with a limited number of evaluations, thereby significantly reducing computational overhead. The method was validated through numerical simulations using a complex foam structure phantom and experimental tests on an industrial micro-CT system with a tissue-equivalent phantom. Results showed that the proposed framework accurately recovers the scan geometry, achieving a 99.4% reduction in RMSE in simulations and effectively suppressing edge blurring and double-contour artifacts in experimental reconstructions. In conclusion, this phantom-free approach offers a reliable and efficient solution for practical CBCT imaging applications where precise mechanical alignment or external calibration objects are unavailable.**

***Index Terms*— Bayesian optimization, CBCT, Gaussian process, geometry calibration, phantom-free calibration**

I. Introduction

THREE-dimensional (3D) X-ray imaging, especially cone-beam computed tomography (CBCT), has become a fundamental tool in medical diagnostics and high-resolution industrial inspection [1,2]. The primary aim of CBCT is to reconstruct an object's internal structure with high spatial accuracy. However, the integrity of the reconstructed volume heavily relies on precise knowledge of the system's scan geometry. In practice, even after meticulous mechanical alignment, residual geometric misalignments, such as detector shifts, often remain. These inaccuracies significantly degrade image quality, resulting in artifacts like edge blurring and double contours. To mitigate these artifacts and improve overall image quality, precise geometric calibration is essential. Such distortions are not merely visual flaws; they undermine the quantitative accuracy and clinical utility of imaging results, particularly in micro-CT applications where sub-pixel precision is crucial [3–5]. Accurate identification of the orientation of system components is essential, as geometric errors can significantly compromise the efficacy of reconstruction algorithms. Conventional methods often rely on specialized phantoms with precisely positioned markers, but these approaches can be impractical in dynamic environments where system components may shift or vary between scans [6–8]. Therefore, there is a pressing need for a self-operating calibration mechanism capable of recovering scan geometry directly from the acquired projection data of the object being scanned [9]. The challenge in this approach lies in the high computational cost and the non-linear, often non-convex relationship between geometric parameters and resulting image quality.

This study proposes a robust self-calibration framework that leverages Bayesian Optimization (BO) to estimate the detector's transverse (δu) and longitudinal (δv) shifts. The proposed method is grounded in the principle of reprojection consistency, identifying the optimal geometry by minimizing a hybrid objective function that simultaneously assesses data fidelity and image structural characteristics. Unlike standard optimization techniques that require derivative information or extensive sampling, the Bayesian approach employs a probabilistic strategy [10]. By modeling the unknown error surface as a Gaussian Process (GP), the algorithm can predict the behavior of the objective function and its associated uncertainty across the parameter space.

The efficiency of this approach lies in its ability to balance the exploration of unknown parameter regions with the exploitation of promising regions through an Expected Improvement (EI) acquisition function. This ensures that a near-optimal solution for multiple geometric parameters can be found within a limited number of evaluations, significantly reducing the computational overhead inherent in complex CT reconstruction tasks. To validate the proposed method, numerical simulations using a foam structure phantom and experimental tests on an industrial micro-CT system were conducted. Image quality was quantitatively assessed using the Root Mean Square Error (RMSE) for simulation results and

Jiyong Shim is with the Department of Radiation Convergence Engineering and iTOMO Research Group, Yonsei University, 1 Yonseidae-gil, Wonju, Gangwon-do 26493, Republic of Korea (telephone: +82-33-761-9660, e-mail: jiyong2155@yonsei.ac.kr).

Hyosung Cho is with the Department of Radiation Convergence Engineering and iTOMO Research Group, Yonsei University, 1 Yonseidae-gil, Wonju, Gangwon-do 26493, Republic of Korea (*corresponding author, telephone: +82-33-761-9660, e-mail: hscho1@yonsei.ac.kr).

intensity profile analysis for experimental data to evaluate the structural accuracy of the reconstruction. Both simulation and experimental results confirm that the proposed method can accurately recover the scan geometry and yield high-quality reconstructed images without the need for any external calibration objects. By framing the calibration problem as a probabilistic search, this research offers a reliable and highly efficient solution for enhancing 3D image quality in practical CBCT systems.

II. Materials and Methods

A. Proposed Geometric Calibration Method

Figure 1 shows the configuration of a standard industrial micro-CT system. In this arrangement, the X-ray source and flat-panel detector are fixed, while the object rotates on a stage around a central axis to obtain projection images. In contrast to the rotating-gantry systems found in medical CT, rotating-object systems deliver enhanced mechanical stability and uniform motion. However, no system is entirely devoid of geometric misalignments, which can affect the quality of reconstruction. The main goal of the proposed self-operating calibration framework is to precisely determine the detector's transverse (δu) and longitudinal (δv) shifts using solely the acquired projection data. This approach is fundamentally rooted in the principle of reprojection consistency, which asserts that a volume reconstructed with accurate geometric parameters will

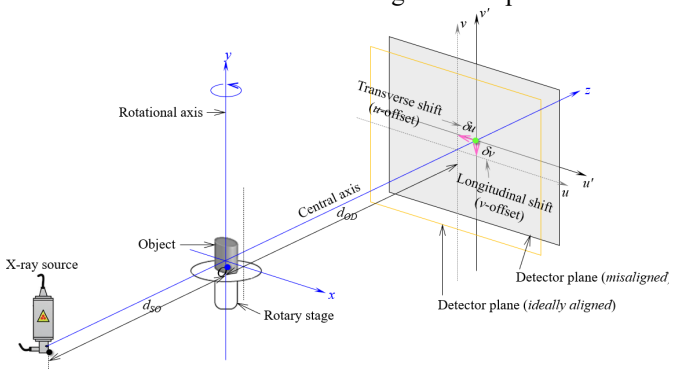


Fig. 1. Schematic of a micro-CT geometry. The X-ray source and the flat-panel detector remain stationary, while the object rotates around a fixed axis on a rotary stage during projection data acquisition.

generate a calculated projection that closely matches the originally measured data.

The optimization process begins by defining an objective function, $f(\theta)$, where $\theta = (\delta_u, \delta_v)$ represents the vector of unknown shift parameters. For each candidate θ evaluated at the k-th step of the optimization process, we reconstruct a volume $x^{(k)}(\theta)$ and assess its quality using the following objective function:

$$f(\theta) = \frac{1}{2}\|p^c - p^m\|_2^2 + \alpha\|Dx^{(k)}(\theta)\|_2 \quad (1)$$

where p^c is the calculated projection obtained by forward-projecting the volume $x^{(k)}(\theta)$, and p^m denotes the experimentally measured projection data , and α is a regularization weight set to unity in this study. The first term, the data fidelity component, minimizes the L_2-norm discrepancy between the measured and calculated projections. The second term penalizes image roughness via a first-order finite difference operator D, applied

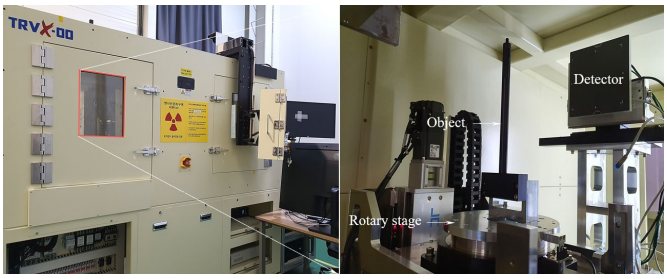


Fig. 2. Industrial micro-CT system used in the experiment, comprising a microfocus X-ray tube with a 10-µm focal spot, a precision rotary stage for object mounting, and a CMOS flat-panel detector with a pixel resolution of 119 µm.

element-wise to the reconstructed volume $x^{(k)}(\theta)$. Geometric misalignment often induces high-frequency artifacts, such as streaks or double-edge patterns, which increase $\|Dx^{(k)}(\theta)\|_2$ and destabilize the reprojection error surface. By minimizing this hybrid function, the algorithm encourages both data consistency and structural clarity.

To navigate the complex and computationally expensive search space without relying on derivative information, we employ Bayesian Optimization (BO) [11,12]. BO treats the objective function as a black-box and approximates its behavior using a Gaussian Process (GP) surrogate model. Although GP fitting scales as $O(N^3)$ with the number of observations, the total number of evaluations here is strictly limited to 40; the dominant computational cost lies in each FDK reconstruction, not in GP inference. By providing both a predicted value and an associated uncertainty estimate, the GP surrogate strategically selects evaluation points, requiring far fewer reconstructions than grid or random search to converge to a near-optimal solution:

$$f(\theta) \sim GP(m(\theta), k(\theta, \theta')) \quad (2)$$

where $m(\theta)$ is the mean function and $k(\theta, \theta')$ is the covariance kernel. This probabilistic approach offers not only an estimate of the objective function's value but also a measure of uncertainty across the parameter space. To determine the most promising sampling points, the Expected Improvement (EI) acquisition function is utilized, specifically defined for minimization as:

$$EI[\theta] = \mathrm{E}[max(0, f_{min} - f(\theta))] \quad (3)$$

where $f_{\min} = \min_i f(\theta_i)$ represents the minimum function value observed across previous evaluations. The EI function balances the exploration of unknown regions with the exploitation of known local minima. In this study, the maximum number of objective evaluations was set to 40 to identify an optimal solution within the bounded search space of -5 mm to 5 mm for both shifts. This configuration allows the algorithm to recover the scan geometry efficiently while maintaining the fidelity required for micro-CT imaging. Final estimated shift values are then integrated into the reconstruction process to produce a geometrically corrected 3D volume.

B. Simulation and Experimental Setup

The proposed geometry calibration algorithm was implemented in MATLAB (R2021a; MathWorks, Natick, MA, USA). All computational procedures were performed on a

personal computer equipped with an AMD Ryzen 7 5800X CPU and 64 GB of RAM operating on Windows 10. The experimental validation was conducted using an industrial micro-CT system, as illustrated in Fig. 2. The imaging chain consists of a microfocus X-ray tube with a 10 μm focal spot, a high-precision rotary stage for object mounting, and a complementary metal-oxide semiconductor (CMOS) flat-panel detector with a pixel resolution of 119 μm. During the experiment, the X-ray tube operated at 100 kVp and 80 μA.

Figure 3 displays the two phantoms utilized to evaluate the performance of the Bayesian optimization-based calibration. For the numerical simulation, a 3D foam structure phantom was employed to provide complex structural details suitable for assessing high-frequency artifact reduction. The numerical phantom occupies a volume of 400 × 400 × 400 voxels with a voxel size of 0.25 mm, and its center is precisely aligned with the isocenter. For experimental validation, a tissue-equivalent phantom containing various density inserts was used to verify the algorithm's effectiveness on medically relevant contrast

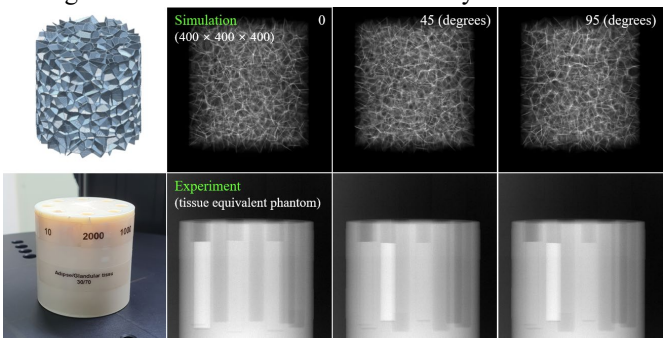


Fig. 3. Two phantoms used for this study: the 3D foam structure phantom (400 × 400 × 400 voxels) for simulation (top) and the tissue equivalent phantom for experimental validation (bottom). Representative projection images (3 out of 400) are shown on the right.

levels. Both phantoms were scanned using a full-rotation CT protocol consisting of 400 projection views. The detailed imaging conditions and geometry parameters for both settings are summarized in Table I.

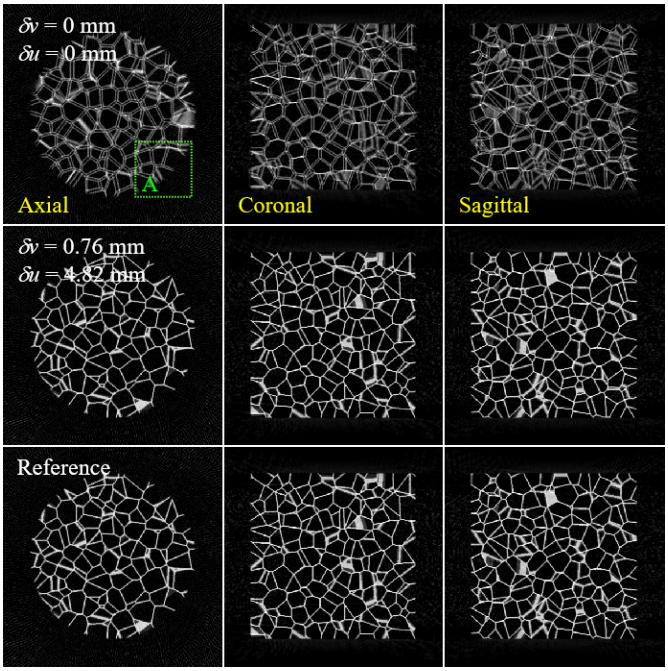
TABLE I
PARAMETERS USED IN THIS STUDY

Parameter	Simulation	Experiment
Source-to-object distance	500 mm	200 mm
Object-to-detector distance	1000 mm	400 mm
Number of projections	400	400
Detector pixel size	0.4 mm	0.119 mm
Focal spot size	–	10 μm
X-ray tube voltage and current	–	100 kVp, 80 μA
Total exposure time	–	3.3 sec
Reconstruction algorithm	FBP-based	FBP-based
Voxel size	0.25 mm	0.05 mm
Test phantom	Foam structure	Tissue equivalent

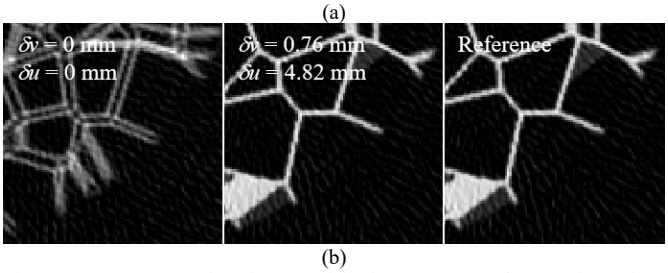
III. RESULTS AND DISCUSSION

The effectiveness of the proposed Bayesian optimization-based calibration was assessed both quantitatively and qualitatively through numerical simulations and experimental data. Figure 4(a) displays the reconstructed 3D axial, coronal, and sagittal slice images of the foam structure phantom under various calibration conditions. The first row illustrates the misaligned state, where both vertical and horizontal offsets were set to zero (δv = 0 mm, δu = 0 mm), revealing significant geometric artifacts such as double contours and structural blurring. These distortions are further emphasized in the enlarged view of region A, where the fine skeletal structures of the foam appear duplicated and poorly defined. Quantitatively, this uncorrected state resulted in a high root mean square error (RMSE) of 0.1753.

Conversely, the second row of Fig. 4(a) illustrates the reconstruction using the optimized parameters obtained through Bayesian optimization (δv = 0.76 mm, δu = 4.82 mm). These results exhibit an almost perfect recovery of the phantom geometry, closely aligning visually with the reference slices in the third row, which employed the designed ground-truth parameters (δv = 0.76 mm, δu = 4.81 mm). The optimized reconstruction achieved a significantly lower RMSE of 0.001, marking a 99.4% reduction in error compared to the misaligned state. The enlarged views in Fig. 4(b) confirm that the Bayesian framework effectively eliminated double-edge artifacts, restoring the sharp structural integrity essential for high-resolution imaging.



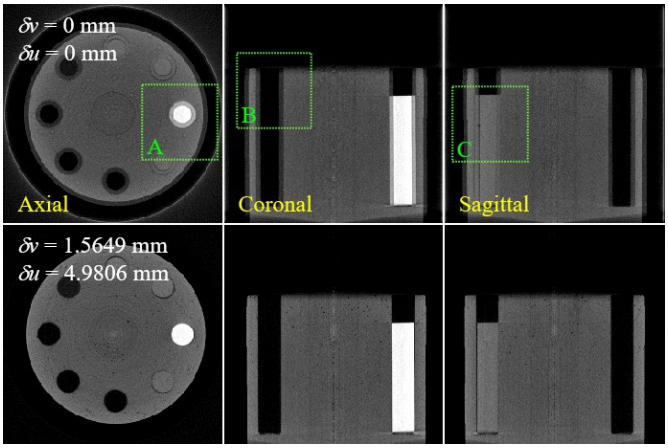
(a)



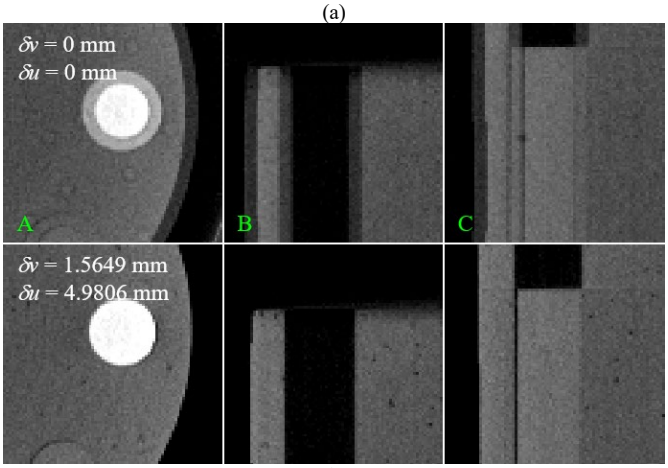
(b)

Fig. 4. (a) Reconstructed 3D images of the foam structure phantom for various geometric calibration conditions: (first row) misaligned both the vertical and horizontal offsets, and (second row) using both the vertical and horizontal offsets (δv = 0.76 mm, δu = 4.82 mm).The corresponding reference slices using designed parameters (δv = 0.76 mm, δu = 4.81 mm) are shown in the third row. (b) Enlarged views of the regions marked as A are shown. All slices were displayed using a WL of 0.5 and a WW of 1.

The experimental performance was validated using a tissue-equivalent phantom, as depicted in Fig. 5. Consistent with the simulation trends, the uncorrected images in the first row (δv = 0 mm, δu = 0 mm) display noticeable blurring and positional shifts of the density inserts. However, upon applying the extracted parameters (δv = 1.5649 mm, δu = 4.9806 mm), the structural clarity of the adipose and glandular tissue-equivalent inserts was significantly enhanced. The magnified views of regions A–C in Fig. 5(b) clearly demonstrate the suppression of geometric artifacts, confirming that the proposed phantom-free method facilitates reliable and accurate 3D reconstruction in practical micro-CT settings.



(a)



(b)

Fig. 5. (a) Reconstructed 3D images of the tissue-equivalent phantom using the extracted scan parameters of δv = 1.5649 mm, and δu = 4.9806 mm (second row). For comparison, 3D images reconstructed with uncorrected detector offsets (first row) are also shown. (b) Enlarged views of the regions marked as A–C are shown. All slices were displayed using a WL of 0.5 and a WW of 1.

The accuracy of the proposed method is primarily governed by the bounded search space and the deterministic nature of the FDK reconstruction, which produces a smooth and stable optimization landscape. In this study, the method successfully recovered shifts as large as 4.98 mm with sub-pixel accuracy, and the use of the reprojection-based objective may reduce sensitivity to local projection noise by aggregating reprojection errors across all views, though a quantitative evaluation under varying noise levels remains future work. Furthermore, because the estimated transverse shift in the experimental case (δu = 4.9806 mm) was near the upper bound of the predefined search range (−5 to 5 mm), the search bounds should be set according to the expected mechanical tolerance of the imaging system; in practice, a wider preliminary range or a coarse-to-fine BO strategy may be employed when larger misalignments are anticipated.

IV. Conclusion

In this study, we proposed and validated a self-operating geometric calibration method for CBCT systems based on Bayesian optimization. Unlike traditional approaches that require specialized phantoms, the proposed framework identifies optimal detector transverse and longitudinal shifts by minimizing a hybrid objective function based on reprojection consistency and structural smoothness, directly from the object's projection data. By utilizing a Gaussian process surrogate model, the algorithm efficiently navigates the complex parameter space, identifying near-optimal solutions within a strictly limited number of objective evaluations.

Numerical simulations with a foam structure phantom demonstrated the method's high precision, achieving a 99.4% reduction in RMSE compared to the misaligned state. Experimental results using a tissue-equivalent phantom further confirmed the robustness of the approach, showing a marked enhancement in structural clarity and the effective suppression of geometric artifacts such as double contours and blurring. These results indicate that the proposed phantom-free calibration is both reliable and practical for high-resolution imaging applications. Overall, this research provides a powerful tool for improving 3D reconstruction quality in CBCT systems, particularly in scenarios where conventional calibration setups are unavailable or mechanical stability is compromised. The current framework addresses translational detector offsets (δu, δv); extension to rotational misalignment parameters represents a natural direction for future work, as the BO framework is dimensionally flexible.

References

[1] T. Kiljunen, T. Kaasalainen, A. Suomalainen, and M. Kortesniemi, “Dental cone beam CT: A review,” Physica Medica, vol. 31, no. 8, pp. 844–860, Oct. 2015, doi: 10.1016/j.ejmp.2015.09.004.

[2] M. C. Felix et al., “Image-Guided Radiotherapy Using a Modified Industrial Micro-CT for Preclinical Applications.,” PLOS ONE, vol. 10, no. 5, p. e0126246, May 2015, doi: 10.1371/journal.pone.0126246.

[3] S. M. Johnston, C. T. Badea, and G. A. Johnson, “Geometric calibration for a dual tube/detector micro-CT system.,” Medical physics, vol. 35, no. 5, pp. 1820–1829, Apr. 2008, doi: 10.1118/1.2900000.

[4] K. Yang, J. M. Boone, A. L. C. Kwan, and D. F. Miller, “A geometric calibration method for cone beam CT systems.,” Medical physics, vol. 33, no. 6, pp. 1695–1706, May 2006, doi: 10.1118/1.2198187.

[5] J. Zhao, X. Hu, J. Zou, and X. Hu, “Geometric Parameters Estimation and Calibration in Cone-Beam Micro-CT.,” Sensors, vol. 15, no. 9, pp. 22811–22825, Sep. 2015, doi: 10.3390/s150922811.

[6] G. Li, S. Luo, C. You, M. Getzin, L. Zheng, G. Wang, and N. Gu, “A novel calibration method incorporating nonlinear optimization and ball-bearing markers for cone-beam CT with a parameterized trajectory,” Med. Phys., vol. 46, no. 1, pp. 152–164, Jan. 2019, doi: 10.1002/mp.13278.

[7] V. Nguyen et al., “A low-cost geometry calibration procedure for a modular cone-beam X-ray CT system,” Nondestructive Testing and Evaluation, vol. 35, no. 3, pp. 252–265, Jun. 2020, doi: 10.1080/10589759.2020.1774580.

[8] J. C. Ford, D. Zheng, and J. F. Williamson, “Estimation of CT cone-beam geometry using a novel method insensitive to phantom fabrication

inaccuracy: Implications for isocenter localization accuracy," Med. Phys., vol. 38, no. 6, pp. 2829–2840, Jun. 2011, doi: 10.1118/1.3589130.

[9] V. Patel, D. R. Bednarek, C. N. Ionita, R. N. Chityala, S. Rudin, and K. R. Hoffmann, "Self-calibration of a cone-beam micro-CT system.," Medical physics, vol. 36, no. 1, pp. 48–58, Dec. 2008, doi: 10.1118/1.3026615.

[10] X. Jiang, X. Wang, A. Uneri, W. B. Zbijewski, and W. W. Stayman, "Differentiable forward and back-projector for rigid motion estimation in X-ray imaging," IEEE Trans. Biomed. Eng., early access, 2025, doi: 10.1109/TBME.2025.3643742.

[11] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the Human Out of the Loop: A Review of Bayesian Optimization," Proc. IEEE, vol. 104, no. 1, pp. 148–175, Jan. 2016, doi: 10.1109/JPROC.2015.2494218.

[12] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient Global Optimization of Expensive Black-Box Functions," J. Global Optim., vol. 13, pp. 455–492, Dec. 1998, doi: 10.1023/A:1008306431147.

Auxiliary Orbit Compensation for Cone-Beam Artifacts in CBCT

Zeyu Zhou[1], Andriy Andreyev[1], and Faguo Yang[1]

***Abstract*—Multi-orbit cone-beam CT systems acquire projection data from multiple circular trajectories to extend the axial field of view, but combining reconstructions from different orbits is challenging due to cone-beam artifacts in the overlap regions. We propose a region-selective reconstruction framework with two key contributions. First, we introduce Sequential FDK (Seq-FDK) with auxiliary orbit trajectory design that minimizes the cone angle in artifact-prone regions by optimizing the source-to-isocenter distance, effectively compensating cone-beam artifacts from the primary orbit. Second, we develop ROI-SIRT, an iterative method that refines only the overlap region while keeping non-overlap regions fixed, substantially reducing computation by iterating only over the small overlap volume. When initialized with trajectory-designed Seq-FDK, ROI-SIRT achieves the best overall reconstruction quality. Experimental results on phantom data demonstrate that Seq-FDK with trajectory design outperforms standard Seq-FDK, and ROI-SIRT with the designed trajectory provides the optimal trade-off between image quality and computational efficiency.**

***Index Terms*—Cone-beam CT, multi-orbit reconstruction, FDK, iterative reconstruction, region of interest.**

I. Introduction

THE Feldkamp-Davis-Kress (FDK) algorithm [1] is a widely used approximate reconstruction method for cone-beam CT. Its main advantages lie in its computational efficiency and simplicity. However, as an approximate method, FDK does not satisfy Tuy's data sufficiency condition, leading to severe cone-beam artifacts, particularly for voxels far from the orbital plane.

Helical trajectories address this fundamental limitation by satisfying Tuy's condition and providing complete data coverage, thereby suppressing cone-beam artifacts. Despite these advantages, helical scanning requires continuous axial motion, which is impractical for C-arm systems and many industrial inspection setups.

Multi-orbit acquisition offers an intuitive and practical alternative by extending axial coverage through multiple circular scans at different axial positions, while preserving the mechanical simplicity of circular trajectories. However, multi-orbit reconstruction introduces a new challenge in overlap regions, where projection data from adjacent orbits contribute unevenly due to orbit-dependent data completeness and reconstruction quality. Simple strategies such as independent reconstruction followed by averaging or stitching [5], [6] often result in visible artifacts or discontinuities at the boundaries.

[1]All authors are with Carl Zeiss X-ray Microscopy, Inc. Dublin, CA, USA. Corresponding author: Zeyu Zhou, zeyu.zhou@zeiss.com

In this work, we propose a region-selective reconstruction framework for multi-orbit cone-beam CT. The key idea is to designate one orbit as the primary scan, while auxiliary orbits are designed to compensate cone-beam artifacts in regions where the primary orbit produces degraded reconstruction quality. The reconstruction volume is divided into non-overlap regions, reconstructed from individual orbits, and overlap regions, where information from multiple orbits is selectively fused. We implement this framework with two fusion strategies: (1) Sequential FDK with distance-based weighting for fast reconstruction, and (2) ROI-SIRT [7], [8], which iteratively refines only the overlap region for higher reconstruction quality.

II. Materials and Methods

A. The Missing Cone Problem

The fundamental cause of cone-beam artifacts in circular trajectories is the violation of Tuy's data sufficiency condition [13]. Tuy's condition requires that every plane intersecting the region of interest (ROI) must also intersect the source trajectory. For a single circular orbit, planes parallel to the source trajectory (i.e., horizontal planes at $z \neq 0$) do not intersect the orbit. In the frequency domain, this corresponds to a "missing cone" of unsampled frequencies around the k_z axis [14].

The extent of this missing data increases with the distance from the central orbital plane. Consequently, reconstruction quality degrades as the cone angle increases, manifesting as characteristic shading and streak artifacts in the off-center regions. By utilizing a multi-orbit configuration, we "fill" this missing frequency cone in the overlap region using data from the auxiliary orbit, thereby restoring data sufficiency.

B. Multi-Orbit CT Geometry

Consider a multi-orbit cone-beam CT system with circular trajectories at different axial positions. We denote the primary orbit at z_1 and auxiliary orbits at $z_2, z_3, \ldots$ Each orbit acquires N_{proj} projections over 360. The source-to-isocenter distance D_{SI} and source-to-detector distance D_{SD} can vary between orbits. The detector has dimensions $N_u \times N_v$ pixels. For simplicity, we present the dual-orbit case ($z_1 = z_{\text{primary}}$, $z_2 = z_{\text{aux}}$) but the framework extends to more orbits.

Let Ω denote the reconstruction volume. For the dual-orbit case, we partition Ω into three disjoint regions based on axial position: the primary-only region Ω_{primary} where only the primary orbit provides good coverage, the auxiliary-only

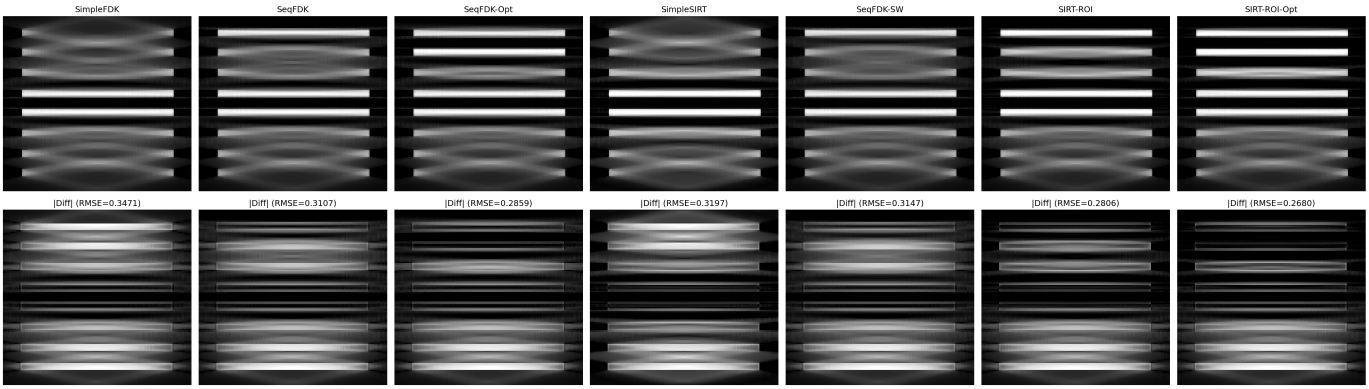


Fig. 1. Reconstruction comparison (XZ slices) for 30° cone angle. Top row: grayscale display. Bottom row: absolute difference image for artifact visualization. From left to right: Simple FDK (primary orbit), Seq-FDK (Gated fusion), SeqFDK-Opt (with trajectory design), SeqFDK-SW (Soft-weighted fusion), SIRT (primary orbit), ROI-SIRT, and ROI-SIRT-Opt. The trajectory-optimized methods (SeqFDK-Opt, ROI-SIRT-Opt) show reduced cone-beam artifacts in the overlap region compared to standard geometry methods.

TABLE I
QUANTITATIVE EVALUATION (RMSE / PSNR / SSIM) ACROSS CONE ANGLES 10° TO 60°

Method	**Angle 10°** (RMSE/PSNR/SSIM)	**Angle 20°** (RMSE/PSNR/SSIM)	**Angle 30°** (RMSE/PSNR/SSIM)	**Angle 40°** (RMSE/PSNR/SSIM)	**Angle 50°** (RMSE/PSNR/SSIM)	**Angle 60°** (RMSE/PSNR/SSIM)
Simple FDK	0.057 / 24.31 / 0.887	0.117 / 18.06 / 0.771	0.174 / 14.61 / 0.675	0.223 / 12.50 / 0.605	0.255 / 11.26 / 0.587	0.273 / 11.08 / 0.562
Seq-FDK	0.052 / 25.12 / 0.901	0.103 / 19.14 / 0.806	0.155 / 15.61 / 0.730	0.199 / 13.43 / 0.672	0.225 / 11.85 / 0.652	0.250 / 11.58 / 0.624
SeqFDK-Opt	0.047 / 26.04 / 0.916	0.091 / 20.25 / 0.829	0.137 / 16.72 / 0.760	0.175 / 14.58 / 0.713	0.192 / 12.47 / 0.708	0.225 / 12.14 / 0.679
SIRT	0.055 / 24.64 / 0.892	0.114 / 18.28 / 0.783	0.170 / 14.80 / 0.695	0.219 / 12.50 / 0.700	0.232 / 11.94 / 0.648	0.247 / 11.72 / 0.625
ROI-SIRT	0.047 / 26.08 / 0.920	0.083 / 21.05 / 0.865	0.119 / 17.89 / 0.827	0.152 / 13.67 / 0.787	0.177 / 12.89 / 0.736	0.204 / 12.55 / 0.707
ROI-SIRT-Opt	**0.041 / 27.28 / 0.932**	**0.071 / 22.38 / 0.889**	**0.104 / 19.14 / 0.849**	**0.136 / 13.85 / 0.802**	**0.151 / 13.24 / 0.765**	**0.189 / 12.95 / 0.743**

region $\Omega_{\text{aux}} \backslash \Omega_{\text{overlap}}$ covered mainly by the auxiliary orbit, and the overlap region Ω_{overlap} where both orbits contribute. The partition boundaries are determined by the viewing geometry and axial extent of each orbit.

C. Sequential FDK Fusion

For fast analytical reconstruction, we employ FDK independently for each orbit and fuse the results. Let f_1 and f_2 denote the FDK reconstructions from the primary and auxiliary orbits, respectively. We consider two fusion strategies:

Gated Fusion: Each voxel selects the reconstruction from the orbit that offers the minimum cone angle. The cone angle for orbit k at position z is $\theta_k = \arctan(|z - z_k|/D_{SI}^{(k)})$. The fusion rule is:

$$f(x, y, z) = \begin{cases} f_1 & \text{if } \theta_1 \leq \theta_2 \\ f_2 & \text{otherwise} \end{cases} \tag{1}$$

When both orbits share the same D_{SI}, this reduces to selecting the nearest orbit. This gated selection provides a simple baseline that avoids noise amplification from combining data.

Soft-Weighted Fusion: For smoother transitions, we use distance-based weights:

$$f(x, y, z) = w_1(z) \cdot f_1 + w_2(z) \cdot f_2 \tag{2}$$

where $w_1(z) = d_2/(d_1 + d_2)$, $w_2(z) = d_1/(d_1 + d_2)$, and $d_k = |z - z_k|$. This ensures $w_1 + w_2 = 1$ with each orbit dominating in regions closer to its focal plane.

D. ROI-SIRT for Overlap Stitching

While Seq-FDK provides fast reconstruction, the overlap region may still exhibit artifacts due to the approximate nature of FDK. We propose ROI-SIRT to iteratively refine only the overlap region using projection data from both orbits, while keeping the non-overlap regions fixed from the Seq-FDK initialization.

The key observation is that non-overlap regions Ω_{primary} and Ω_{aux} are well-reconstructed by single-orbit FDK, so iterative refinement is only needed for Ω_{overlap}, leveraging concepts from interior tomography [9], [10]. To accelerate convergence and ensure stability, we implement the ROI-SIRT algorithm with two key optimizations: (1) **Difference Projection** to handle the non-updated regions efficiently, and (2) **Sigmoid Weighting** for seamless orbital fusion. The algorithm proceeds as follows:

This formulation ensures that the iterative updates are consistent with the global projection data while computationally restricted to the small overlap volume, substantially reducing the cost compared to full-volume SIRT.

E. Auxiliary Orbit Trajectory Design

The primary orbit is kept fixed, while auxiliary orbit trajectories are designed to minimize cone-beam artifacts in regions where the primary orbit produces poor quality. For FDK reconstruction, artifacts increase with the cone angle θ between the ray and the orbital plane. By adjusting the auxiliary orbit's

Algorithm 1 ROI-SIRT for Overlap Stitching

Require: Sequence of N projections $\mathbf{p}_{\text{measured}}$, Initial volume $\mathbf{f}^{(0)}$ (from Seq-FDK), fixed non-overlap volume $\mathbf{f}_{\text{static}}$.

Pre-compute background:
$\mathbf{p}_{\text{bg}} \leftarrow \mathbf{A}(\mathbf{f}_{\text{static}})$
for $k = 1$ **to** N_{iter} **do**
  $\mathbf{p}_{\text{roi}} \leftarrow \mathbf{A}_{\text{roi}}(\mathbf{f}_{\text{roi}}^{(k-1)})$
  $\mathbf{r} \leftarrow \mathbf{p}_{\text{measured}} - (\mathbf{p}_{\text{bg}} + \mathbf{p}_{\text{roi}})$
  $\mathbf{b} \leftarrow \mathbf{A}_{\text{roi}}^{T}\mathbf{r}$
  Update with Sigmoid Weighting:
  $\sigma(z) \leftarrow (1 + e^{-\alpha(d_1-d_2)/(d_1+d_2)})^{-1}$
  $\mathbf{f}^{(k)} \leftarrow \mathbf{f}^{(k-1)} + \lambda \left(\sigma(z)\frac{\mathbf{b}_1}{\mathbf{C}_1} + (1-\sigma(z))\frac{\mathbf{b}_2}{\mathbf{C}_2}\right)$
end for
return $\mathbf{f}^{(N_{\text{iter}})}$

source-to-isocenter distance D_{SI}^{aux}, we can minimize the cone angle for voxels in artifact-prone regions.

For a voxel at axial distance Δz from the auxiliary orbital plane, the cone angle is $\theta_{\text{aux}} = \arctan(\Delta z / D_{SI}^{\text{aux}})$. To minimize θ_{aux}, we want to maximize D_{SI}^{aux}. However, increasing D_{SI}^{aux} reduces the magnification factor, which imposes a minimum constraint to ensure detector coverage of the reconstruction volume:

$$D_{SI}^{\text{aux}} \geq D_{SI,\min}^{\text{aux}} = \frac{D_{SD} \cdot r_{\max}}{w_{\text{det}}/2} \tag{3}$$

where $r_{\max}$ is the maximum reconstruction radius in the XY plane and w_{det} is the detector width. We set $D_{SI}^{\text{aux}} = D_{SI,\min}^{\text{aux}}$ to maximize the source distance while satisfying the FOV constraint.

In addition to optimizing D_{SI}^{aux}, we also optimize the auxiliary orbit position z_{aux} using a minimax criterion. Under gated fusion, each voxel selects the orbit with smaller cone angle. With both orbits sharing the same D_{SI}, the switching boundary occurs at $z^* = z_{\text{aux}}/2$. The maximum cone angle in the fused reconstruction is:

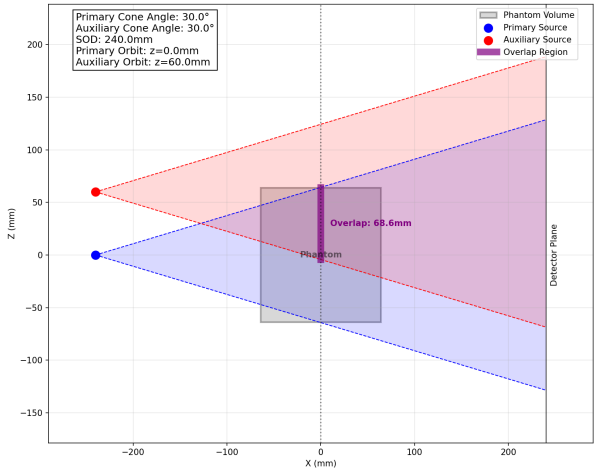
$$\theta_{\max} = \max\left(\frac{z_{\text{aux}}/2}{D_{SI}}, \frac{z_{\max} - z_{\text{aux}}}{D_{SI}}\right) \tag{4}$$

This trajectory design principle can be applied to any region with cone-beam artifacts. In our dual-orbit example, by increasing D_{SI}^{aux} to the maximum allowable value and positioning the auxiliary orbit at the minimax-optimal location, the effective cone angle in the overlap region is reduced from approximately 30° to 10.2° (designed auxiliary orbit), a 66% reduction.

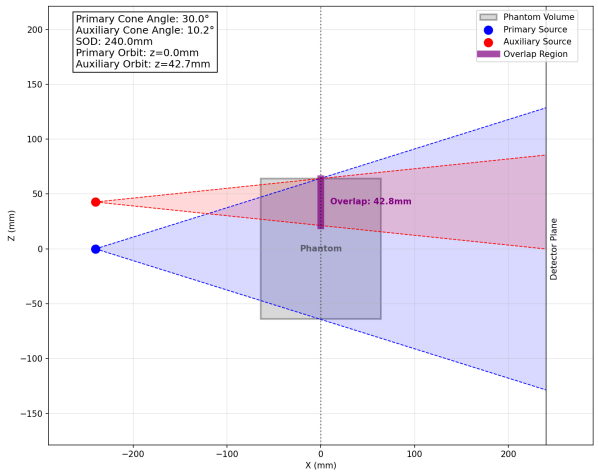
III. Results

A. Experimental Setup

We evaluated the proposed methods using simulated projection data from a 3D disk phantom ($128 \times 128 \times 128$ voxels, 1 mm voxel size, $z \in [-64, 64]$ mm). The primary orbit was centered at $z = 0$ mm, and the auxiliary orbit was initially positioned at $z = 60$ mm. To systematically study the effect of cone angle on reconstruction quality, we tested



(a) Primary Cone Angle 30° (Standard Geometry)



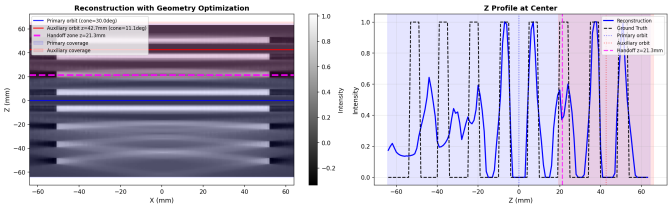
(b) Primary Cone Angle 30° (Optimized Geometry)

Fig. 2. Visualization of overlap regions. (a) Standard configuration with 30° cone angle showing large overlap. (b) Optimized configuration ($z = 42.7$ mm) with reduced auxiliary cone angle ($\sim$ 10.2°). The optimized geometry minimizes the cone angle in the overlap region while maintaining sufficient coverage for the handoff.

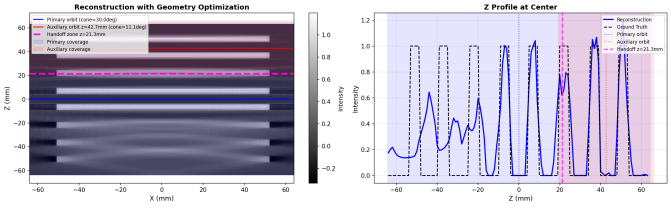
configurations from 10° to 60° in 10° increments. For each cone angle θ, the source-to-isocenter distance was computed as $D_{SI} = 64/\tan(\theta/2)$ to ensure cone coverage of the phantom half-height, ranging from $D_{SI} = 1500$ mm (5°) to $D_{SI} = 150$ mm (60°).

For trajectory-optimized methods, the auxiliary orbit position was adjusted to $z_{\text{aux}} \approx 42.7$ mm to minimize the maximum ray angle in the overlap region, with the cone angle reduced accordingly. Each orbit acquired 360 projections over 360 with detector pixel size 1 mm.

To simulate realistic imaging conditions, we applied both photon count noise and electronic noise to the projection data. Poisson noise was modeled assuming an incident photon intensity of $I_0 = 5 \times 10^4$ and an effective attenuation scaling $\mu_{\text{eff}} = 0.02$. Gaussian noise with zero mean and standard deviation $\sigma = 0.01$ was subsequently added to the line integral



(a) SeqFDK-Opt



(b) ROI-SIRT-Opt

Fig. 3. Comparison of geometry-optimized methods. (a) SeqFDK-Opt with optimized auxiliary orbit. (b) ROI-SIRT-Opt. Both panels show the reconstruction with orbit positions (left) and the Z-profile comparison against ground truth (right). ROI-SIRT-Opt provides further image quality improvement in the overlap region through iterative refinement.

data to simulate electronic noise.

We compared seven methods: Simple FDK (primary orbit only), Seq-FDK (dual-orbit gated fusion), Seq-FDK-Soft (dual-orbit soft-weighted fusion), SeqFDK-Opt (gated fusion with trajectory design), SIRT (primary orbit only), ROI-SIRT (overlap refinement with standard geometry), and ROI-SIRT-Opt (overlap refinement with trajectory design). Due to the symmetric geometry of the dual-orbit configuration, we focused our evaluation on the upper half of the reconstruction volume ($z > 0$), where the auxiliary orbit contributes; results for the lower half ($z < 0$) are expected to be similar by symmetry. Image quality was assessed using RMSE, PSNR, and SSIM computed against ground truth in both the overlap region and the mid-plane slice ($z = 0$).

B. Quantitative Results

Table I summarizes the quantitative comparison against the ground truth. The single-orbit methods (Simple FDK, SIRT) serve as baselines showing limited coverage in the overlap region. The dual-orbit Seq-FDK significantly improves over single-orbit FDK by fusing both datasets. Geometry optimization (SeqFDK-Opt) further reduces cone-beam artifacts. For iterative methods, ROI-SIRT achieves high-quality fusion in the overlap region while substantially reducing computation by iterating only over the small overlap volume. The combination of ROI acceleration and geometry optimization (ROI-SIRT-Opt) provides the best overall performance.

IV. Discussion

Our results show region-selective processing addresses the challenges of dual-orbit CT reconstruction. The key insight is that different regions of the reconstruction volume have different optimal processing strategies: non-overlap regions benefit from single-orbit reconstruction (avoiding noise amplification from combining data), while the overlap region requires careful fusion to minimize artifacts.

The ROI-SIRT variant achieves most of the quality improvement at significantly reduced computational cost. This is particularly important for industrial applications, where reconstruction throughput and computational cost are critical considerations. The approach can be further accelerated through GPU implementation and optimized projection operators.

Limitations of the current study include the use of simulated data only. Future work will validate the methods on physical phantom, and investigate extension to more than two orbits.

V. Conclusion

We presented a region-selective reconstruction framework that uses auxiliary orbit scanning to compensate cone-beam artifacts in CBCT. The framework includes two key contributions: (1) trajectory design that optimizes both the source-to-isocenter distance and orbit position using a minimax criterion to minimize the effective cone angle in artifact-prone regions, and (2) ROI-SIRT that iteratively refines only the overlap region while keeping non-overlap regions fixed. Experimental results demonstrate that the combination of trajectory-optimized Seq-FDK with ROI-SIRT achieves the best reconstruction quality, offering a practical solution for reducing cone-beam artifacts in multi-orbit CBCT.

References

[1] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, 1984.
[2] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*. New York: IEEE Press, 1988.
[3] W. van Aarle *et al.*, "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.
[4] L. Zhu, J. Wang, and L. Xing, "Noise suppression in scatter correction for cone-beam CT," *Med. Phys.*, vol. 36, no. 3, pp. 741–752, 2009.
[5] K. Zeng and Z. Chen, "Review of recent developments in cone-beam CT reconstruction algorithms for long-object problem," *Image Anal. Stereol.*, vol. 23, no. 2, pp. 83–87, 2004.
[6] S. Preibisch, S. Saalfeld, and P. Tomancak, "Globally optimal stitching of tiled 3D microscopic image acquisitions," *Bioinformatics*, vol. 25, no. 11, pp. 1463–1465, 2009.
[7] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm," *Ultrasonic imaging*, vol. 6, no. 1, pp. 81–94, 1984.
[8] J. Gregor and T. Benson, "Computational analysis and improvement of SIRT," *IEEE Trans. Med. Imag.*, vol. 27, no. 7, pp. 918–924, 2008.
[9] G. Wang and Y. Ye, "Meaning of interior tomography," *Phys. Med. Biol.*, vol. 58, no. 16, p. R161, 2013.
[10] H. Kudo, F. Noo, and M. Defrise, "Tiny a priori knowledge solves the interior problem in computed tomography," *Phys. Med. Biol.*, vol. 47, no. 15, p. 2651, 2002.
[11] Y. Sun, J. Liu, M. Sun, B. Wohlberg, and U. S. Kamilov, "CoIL: Coordinate-based internal learning for tomographic imaging," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1400–1412, 2021.
[12] L. Shen, J. Pauly, and L. Xing, "NeRP: Implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 4, pp. 1822–1834, 2022.
[13] H. K. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM J. Appl. Math.*, vol. 43, no. 3, pp. 546–552, 1983.
[14] M. Grass *et al.*, "Three-dimensional reconstruction of high contrast objects using long-range C-arm tomography," *IEEE Trans. Med. Imag.*, vol. 18, no. 10, pp. 906–919, 1999.

Comparison of Local Regression-Based Method for Removing Ring Artifacts in CT Images

Xiangyuan Liu, Xudong, Ru, Shusen Zhao*

***Abstract*—X-ray computed tomography (CT) imaging is inherently susceptible to ring artifacts due to the imperfections of the system's components, such as inconsistent detector response, fluctuations in X-ray source intensity, and variations in detector pixel energy thresholds. These artifacts degrade CT images in complex ways, thereby having negative impacts on subsequent applications. In this paper, we propose effective approaches for suppressing ring artifacts by employing local regression strategies to reconstruct the mean projection (MP) of sinogram data. Specifically, we investigate polynomial and multi-Gaussian regression models for locally restoring the ideal MP from degraded data. An intrinsic weighting scheme is incorporated to adaptively measure the reliability of the obtained MP data. Moreover, to eliminate non-stationary artifacts effectively, the sinogram data is partitioned into segments based on the physical properties of scanning angles. Then, the corresponding local MPs are restored separately, ensuring the proposed method has great robustness in eliminating diverse types of ring artifacts. Finally, we evaluate these regression methods and analyze their advantages. Experimental results confirm that the proposed method, especially the polynomial regression method, can effectively suppress ring artifacts while preserving essential image structures.**

***Index Terms*—Ring artifact, computer tomography images, sinogram data, local regression, mean projection.**

I. Introduction

X-RAY computed tomography (CT) can non-invasively visualize internal structures of scanned objects without breaking them, thereby is widely used in medical and industrial fields. However, ring artifacts can inevitably degrade CT images and then lead to misguided subsequent image analysis and diagnosis. The causes of ring artifacts are diverse, primarily due to imperfect performances of CT system components, such as non-uniform response of detector pixels, the temporal variations in X-ray beam hardness [1], energy threshold variation among PC detector elements [2], etc. Consequently, ring artifacts are more complex than imagined. It is of great significance to robustly remove or suppress diverse ring artifacts in CT images.

A CT image can be reconstructed by a back-projection algorithm applied to the sinogram data collected by detectors. We introduce some representative methods in the following.

1) Early standard calibration. Standard calibration can avoid some ring artifacts before scanning an object, mainly including flat-field correction of detectors and hardware system adjustment [3]. An ideal flat-field correcting method can remove sensitivity variations of all detector elements [4]. Several advanced flat-filed correction approaches are developed [5]. However, these methods have high hardware operating costs and are difficult to implement.

2) Pre-processing in the sinogram domain. Ring artifacts manifest as undesired stripes in sinogram data. Simple low-pass filters [6] were originally suggested to filter stripe artifacts in the sinogram, but high-frequency image details inevitably deteriorated. A Fourier low-pass filter combined with wavelet decomposition (WFT) is proposed to remove stripe artifacts[7]. A measurement-driven method [5] is developed by subtracting the mean projection (MP) of dark current acquired projection images. A fast ring artifacts removal method [8] combines the total variation with L1-norm constraint with Gaussian filtering (TV-L^1aG) to achieve accurate MP estimation.

3) Post-processing in the CT image domain. Ring artifacts form concentric circles in reconstructed CT images. The CT domain techniques can transform CT images to polar coordinates. A median filter [9] removes the transformed stripe artifacts according to the assumption that the intensities of artifacts are local extrema. The relative total variance method and an iterative image domain method are proposed in [10]. According to the geometric properties of ring artifacts, an optimization model that uses the ring total variation (RTV) regularization is developed [11] to suppress ring artifacts.

4) Reconstruction processing methods. Ring artifacts can also be removed during image reconstruction by employing specific regularization techniques. Two reconstruction frameworks based on TV regularization and dictionary learning are proposed for ring artifacts removal [12]. A ring total variation regularization in the CT image domain is introduced to suppress artifacts [13]. Sparsity-based norms of imperfect component and smoothness penalties for the reconstructed image are introduced as data fidelity terms in [14].

In general, ring artifacts have more significant properties in sinogram data. Most image processing-based methods default that ring artifacts have constant intensities. However, in practice, ring artifacts are more complex due to the instability

This work was supported by the National Key R&D Program of China (2023YFA1011400,2023YFA1011402), National Natural Science Foundation of China (NSFC) (12426301), Shenzhen Science and Technology Program (Grant No. JSGGZD20220822095600001, RCBS20221008093316046), the Longhua District Science and Innovation Commission Project Grants of Shenzhen (Grant No.20250113G43468522). (Corresponding author: Shusen Zhao.)

Xiangyuan Liu is with the National Center for Applied Mathematics Shenzhen (NCAMS), Southern University of Science and Technology, Shenzhen, 518055, China (e-mail: liuxy7@mail.sustech.edu.cn).

Xudong Ru is with School of Artificial Intelligence, Beijing Normal University, Beijing, 100875, China, (e-mail: ruxudong@126.com).

Shusen Zhao is with the National Center for Applied Mathematics Shenzhen (NCAMS), Southern University of Science and Technology, Shenzhen 518055, China, and with the Detection Institute for Advanced Technology Longhua-Shenzhen(DIATLHSZ), Shenzhen, 518000, China. (e-mail: zhaoss@sustech.edu.cn).

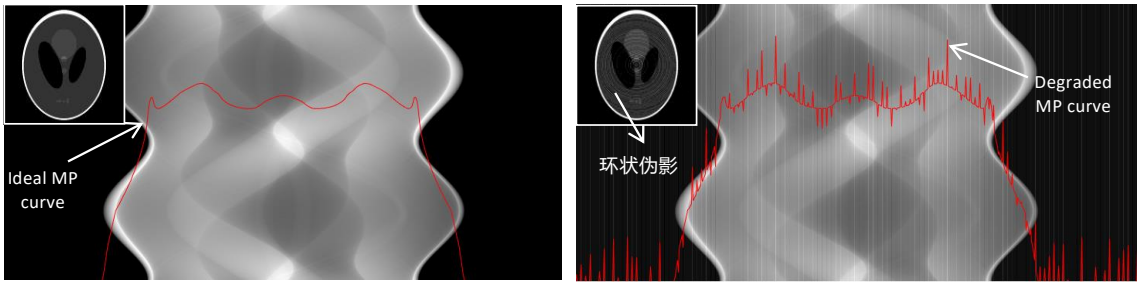


Fig. 1. Ring artifacts in various domains and MP curves. Upper-right corner: CT images of sinogram data (large images).

and uncertainties of CT systems. In this paper, we propose effective ring artifact removal methods by restoring MP using local regression techniques. The main contributions are summarized as follows: 1) effective local regression techniques are employed to restore MP values from degraded data, with appropriate boundary conditions applied to preserve the natural characteristics of ideal MP curves; 2) an intrinsic weighting scheme is introduced to assess the reliability of measurement data, thereby ensuring the accuracy of restored MP values; 3) sinogram data is segmented, and their local MPs are separately restored to effectively eliminate non-stationary ring artifacts.

II. Methodology

A CT image is reconstructed from the projected/sinogram data collected by detectors from various scanning angles. Due to imperfect performances of X-ray sources and detector pixels, etc., undesired elements are initially acquired in sinogram data, resulting in ring artifacts in CT images. Let $F = \{F(i,j)\}, i = 0, \cdots, N_r, j = 0, \cdots, N_d$ be acquired sinogram data, with N_r denotes scanning angles (360°), N_d denotes the number of detector pixels, then, the defective detector pixel j_d would collect undesired components, and degraded sinogram data can be formulated as

$$F(i, j_d) = \tilde{F}(i, j_d) + \varepsilon(i, j_d), \quad i = 0, \cdots, N_r \tag{1}$$

where $\tilde{F}(i, j_d)$ denotes ideal data, $\varepsilon(i, j_d)$ denotes undesired components.

In stable CT systems, the inconsistent response of detector elements is the primary cause of ring artifacts. In this case, undesired components $\varepsilon(i, j_d)$ are line-shaped in all scanning angles and generate stationary ring artifacts that have constant intensities. In practice, non-stationary artifacts may be introduced in unstable CT systems. The intensities of such artifacts are discontinuous globally but constant at certain scanning angles, which means $\varepsilon(i, j_d)$ can be divided into several categories, and each category has constant values.

The aforementioned properties indicate that accurately eliminating stationary artifacts is the key technology for handling non-stationary ones. Fig.1 shows that, in sinogram data, stationary stripe artifacts have significant geometric commonalities in scanning angle direction. Mean projection (MP) can extract the prominent properties of stripe artifacts in low (1D) dimensional space, where the MP value of detector pixel j is defined as

$$MP(j) = \Sigma_{i=0}^{N_r} F(i,j)/N_r, \quad j \in 0, \cdots, N_d \tag{2}$$

The red curves in Fig.1 are MP curves of sinogram data. It shows that the ideal MP curve has good smoothness, but the degraded MP curve contains undesired peaks.

In this paper, we segment the sinogram data to extract stationary subtypes from non-stationary stripe artifacts. To eliminate stationary stripes, we investigate effective local regression technologies to restore the ideal MP values from degraded data, including polynomial and multi-Gaussian regression models. Furthermore, an effective weighting mechanism is suggested to measure the reliability of degraded points, thereby ensuring the accuracy of restored results. regression

A. Local intrinsic weighted regression models

Regression technologies aim to model the relationship between variables x and y using sample data. A new target value can be predicted by the regressed function. More specifically, let $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ be sample data, then the regression model can be obtained by minimizing the residual sum of squares, i.e.,

$$h^*(x) = arg\min_{h(x)} \frac{1}{n} \sum_{i=1}^{n} (y_i - h(x_i))^2, \tag{3}$$

where $h^*(x)$ is the regression function. The target value at x_0 is predicted by $y_0 = h^*(x_0)$. The visual results of observed sample data can roughly determine the regression function. However, in practice, the relationship between variables is mostly complex (e.g., MP curve in Fig. 1). It is difficult to determine an appropriate regression function globally since the relationships differ across local sample data. Therefore, using local properties of the sample data for piecewise regression is a more reasonable approach. Moreover, there may be unreasonable values in sampled data, as peaks in Fig.1. Effectively measuring the rationality of sampled data can further ensure the accuracy of regression results.

Motivated by the aforementioned properties, we construct the local intrinsic weighted regression models. Divide the sample data into m segments, the segment points are noted as

$\{p_0 = x_1, p_1, \cdots, p_m, p_m = x_n\}$, then the regression model for the tth segment is constructed as

$$\min_{h_t(x)} \frac{1}{n_t} \sum_{i=1}^{n_t} w_i (y_i - h_t(x_i))^2, \tag{4}$$

$$s.t. \tag{5}$$

$$x_i \in \{p_{t-1}, \cdots, p_t\}, \tag{6}$$

$$h'_t(x) = h'_{t+1}(x), x = p_t, \tag{7}$$

$$h''_t(x) = h''_{t+1}(x), x = p_t, \tag{8}$$

$$t \in \{1, 2, \cdots, m\} \tag{9}$$

where n_t denotes the number of points in the tth category, and satisfy $n_t = Card([p_{t-1}, \cdots, p_t])$, Eq.7 and Eq.8 are respectively conditions for continuity and smoothness. Our weighting mechanism is defined by w_i for measuring the reliability of sample points, which is defined as

$$w_i = w_i^s * w_i^{is} \tag{10}$$

$$w_i^s = e^{-\|x_i - x_0\|/(2\sigma_s^2)}, \quad w_i^{is} = e^{-Idis(x_i)/(2\sigma_{id}^2)}, \tag{11}$$

$$Idis(x_i) = \frac{1}{n_{x_i}} \sum_{j=1}^{n_{x_i}} |y_j - y_i|^2, \tag{12}$$

where $(x_j, y_j) \in (N_{x_i}, Y(N_{x_i}))$ is the sample data in the neighbour region N_{x_i} of x_i. w_i^s denotes spatial similarity weight, and closer points deserve a larger w_i^s since neighboring data are likely to be the same. w_i^{is} denotes intrinsic similarity weight, which indicates the rationality of sample (x_i, y_i). $idis(x_i)$ is the intrinsic local distance and is defined by the difference between y_i and its neighboring data. A larger $idis(x_i)$ indicates (x_i, y_i) potentially be undesirable data. Moreover, we investigate and compare the following regression models h_t,

1) Polynomial function:

$$h_t(x) = \mu_t^1 x + \mu_t^2 x^2 + \cdots + \mu_t^n x^n + b_t, \tag{13}$$

where $\mu_t^i, i \in \{1, 2, \cdots, n\}$, and b_t denote regression coefficients of the n-th order polynomial function. $h_t(x)$ is linear function when $n = 1$.

2) Multi-Gaussian function:

$$h_t(x) = \sum_{i=1}^{n} a_i * e^{-(x-\mu_i)^2/2\sigma_i^2}, \tag{14}$$

where n denotes the number of Gaussian functions, μ_i and σ_i are parameters of the i function. In the experiment section, we evaluate these regression models with various n and recommend the appropriate n for real applications.

B. Ring artifacts removal

The geometric properties of stationary ring artifacts indicate that they can be removed using the residual MP components. Non-stationary artifacts are more complex and are piecewise continuous. Global MP values (Eq.2) can not capture their temporal variation. To handle such artifacts, we segment the sinogram data to extract stationary subtypes, then separately eliminate undesired elements using the restored local MP values. The key steps of handling non-stationary artifacts are provided below.

1) Categorization of degraded sinogram data: The segmentation of sinogram data should meet two conditions: 1) intensities of stripe artifacts in the same category are approximately constant; 2) pixels within the same category are spatially continuous. A complex categorization method may cause high time consumption or other additional parameters, bringing unfavorable factors to practical applications. Compromising practical application and segmentation conditions, we employ a simple but effective method for categorizing the sinogram data.

Scanning angles are spatially continuous. Benefiting from this property, we perform the hard categorization of pixels along the scanning angle direction. More specifically, when stripes are classified into K categories, the categorized sinogram data and their local MP values satisfy

$$\mathrm{MP}_k(j) = \sum_{i=(k-1)*N}^{k*N} F(i,j)/N, \tag{15}$$

$$k = 1, \cdots, K-1, \quad j = 1, \cdots, N_d$$

$$\mathrm{MP}_K(j) = \sum_{i=(K-1)*N}^{N_r} F(i,j)/(N_r - (K-1)*N), \tag{16}$$

where MP_k is the MP values of kth-class sinogram data; N is the number of pixles in base categories and satisfies $N = \lfloor N_r/K \rfloor$.

2) Removal of stripe artifacts: Stripes in categorized sinogram data (Eq.16) have approximately constant values internally. Based on the geometric commonalities of stripes, non-stationary stripe artifacts can be eliminated using local residual MP components. When eliminating stripe artifacts in k-th, $k \in \{1, 2, \cdots, K\}$, category, we treat local MP values $\{(j, \mathrm{MP}_k(j))\}, j = 1, 2, \cdots, N_d$ as sample data, then restored values MP'_k can be obtained by local regression models (Eq.4). Then, stripe artifacts can be removed by

$$\hat{F}_k(i,j) = F_k(i,j) - (\mathrm{MP}'_k(j) - \mathrm{MP}_k(j)), \tag{17}$$

with $i = 1, \cdots, N_r$, F_k and $\hat{F}_k$ respectively denote degraded and restored sinogram data of the k-th category, $\mathrm{MP}'_k(j) - \mathrm{MP}_k(j)$ denotes the eliminated components. Categorized stripe has approximately constant intensities. Using the categorization strategy, non-stationary artifacts can be removed.

III. Experiments

The proposed method is evaluated by comparison with the classical WFT cite M2009Stripe. The results are provided in Fig.2, where the test data is degraded by stationary (the first and third rows) and non-stationary (the second row) ring artifacts. It can be found that the proposed MP-regression-based methods (Fig.2c and Fig.2d) perform well in correcting ring artifacts and accurately restoring image structures. WFT (Fig.2b) can obtain acceptable results when eliminating weak artifacts (the first row), but it remains undesired components when handling sharp and non-stationary artifacts (see the colored arrows in Fig.2b). Our methods perform better in eliminating artifacts, and the polynomial regression method obtains the best performance, see the red arrows.

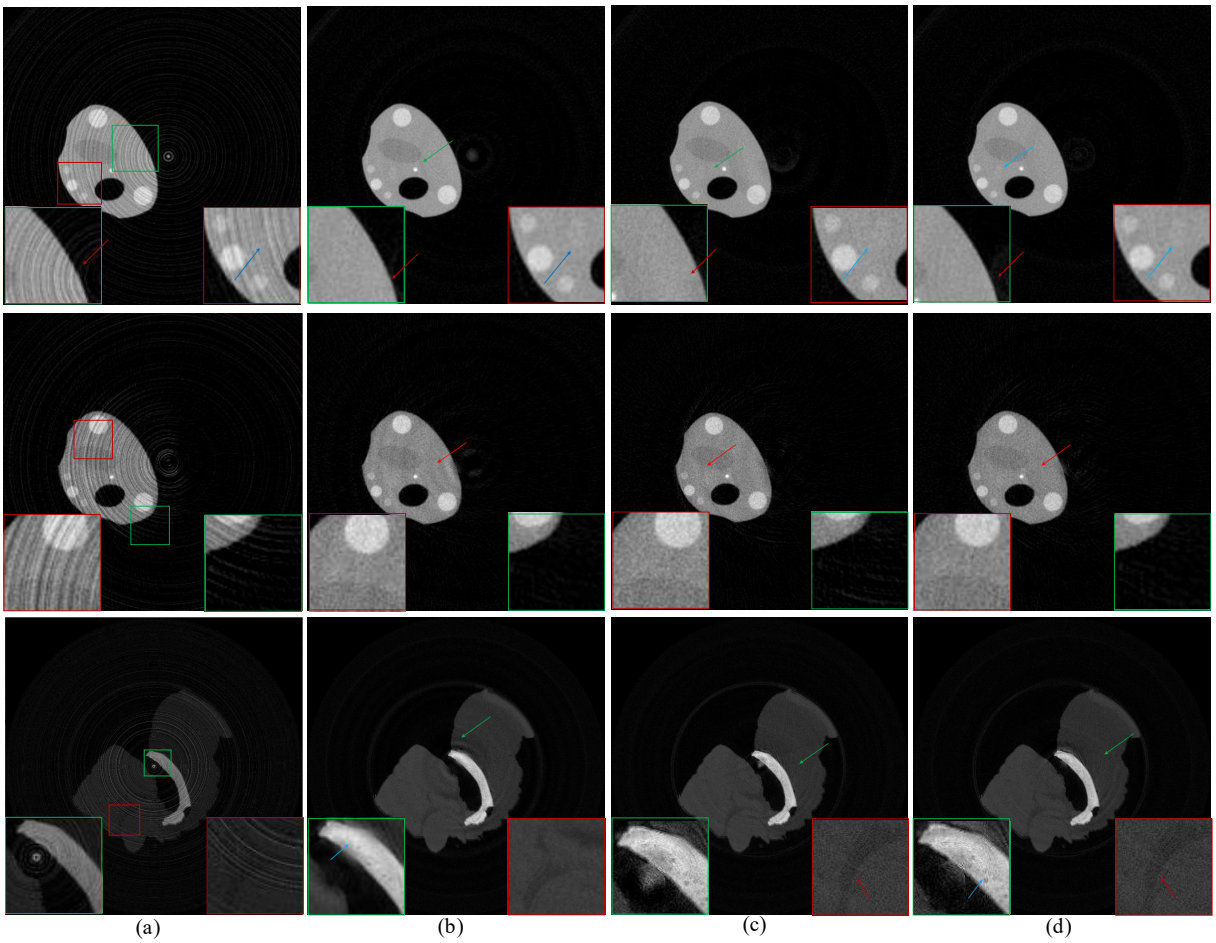


Fig. 2. Ring artifacts removal, with (a) degraded image, (b) WFT[7], (c) Multi-Gaussian regression, (d) Polynomial regression. First and third rows: stationary artifacts; second row: non-stationary artifacts.

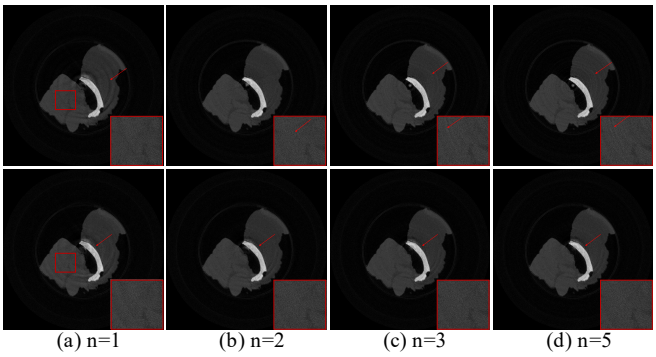


Fig. 3. Results of various n in Eq.13 and Eq.14. First row: Multi-Gaussian regression; second row: Polynomial regression.

Fig.3 evaluates the impacts of the function's order n in Eq.13 and Eq.14. It indicates that the Multi-Gaussian regression method can not effectively remove artifacts when $n = 1$ (Fig.3a), but can achieve acceptable results when $n = 2$(Fig.3b). The results of Multi-Gaussian models (the first row) indicate that this model is more susceptible to the number (n) of base Gaussian functions. As n increases, more undesired elements potentially are introduced, see the red arrows in the first row. In contrast, polynomial regression models exhibit better robustness. The function's order (n) has a smaller effect on such models. Fig.3b, Fig.3b, and Fig.3b in the second row indicate that the results from different n are not significantly different.

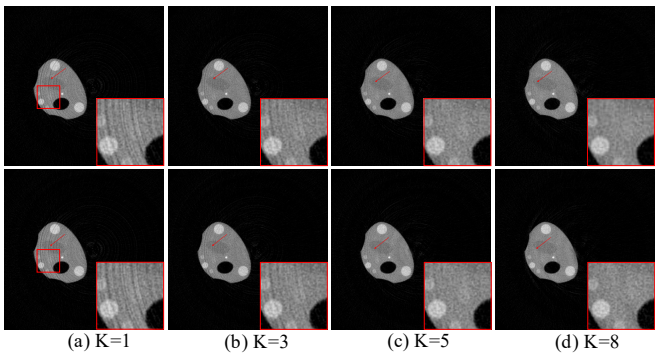


Fig. 4. Results of various numbers K of categories. First row: Polynomial regression; second row: Multi-Gaussian regression.

The number (K) of the categorised sinogram data is evaluated in Fig.4. The test data is degraded by non-stationary ring artifacts. It indicates that with $K = 1$, the proposed methods

cannot effectively remove the non-stationary ring artifacts, but they can achieve considerable results with an appropriate $K > 1$. Specifically, their performance can gradually improve as K increases (see the red arrows). A large K would lead to the disruption of image structures and potentially introduce other unreasonable elements, especially the Multi-Gaussian regression method. Consequently, we recommend $K = [5, 8]$ for correcting non-stationary ring artifacts.

IV. Conclusion

This work proposes an effective technique for ring artifact removal. The suggested method exhibits distinct robustness in eliminating diverse ring artifacts while restoring important image structures. Motivated by the natural properties of ideal MP curves, we utilize the local fitting technologies to restore MP, thereby removing stripe and ring artifacts. A reasonable weighting strategy is suggested to measure the reliability of degraded values, which can confirm the accuracy of restored results. To handle non-stationary artifacts, we separately process the segments of the sinogram data using the proposed local fitting technologies. The experimental results demonstrate that the proposed methods have significant advantages in removing ring artifacts.

References

[1] M. Persson and H. Bornefalk, "A framework for evaluating threshold variation compensation methods in photon counting spectral ct," *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1861–1874, 2012.

[2] H. O. Aggrawal, M. S. Andersen, S. Rose, and E. Y. Sidky, "A convex reconstruction model for X-ray tomographic imaging with uncertain flat-fields," *IEEE Trans. Computat. Imag.*, vol. PP, no. 99, pp. 1–15, 2018.

[3] J. Lifton and T. Liu, "Ring artefact reduction via multi-point piecewise linear flat field correction for X-ray computed tomography," *Opt. Express*, vol. 27, pp. 3217–3228, 2019.

[4] W. Vågber, J. C. Larsson, and H. M. Hertz, "Removal of ring artifacts in microtomography by characterization of scintillator variations," *Opt. Express*, vol. 25, p. 23191–23198, 2017.

[5] L. C. P. Croton, G. Ruben, K. S. Morgan, D. M. Paganin, and M. J. Kitchen, "A simple, pixel-wise response correction for ring artifact removal in both absorption and phase contrast X-ray computed tomography," *Opt. Express*, vol. 27, no. 10, pp. 14 231–14 245, 2019.

[6] G. Kowalski, "Suppression of ring artefacts in CT fan-beam scanners," *IEEE Trans. Nucl. Sci.*, vol. 25, no. 5, pp. 1111–1116, 1978.

[7] B. Münch, P. Trtik, F. Marone, and M. Stampanoni, "Stripe and ring artifact removal with combined wavelet—fourier filtering," *Opt. Express*, vol. 17, no. 10, pp. 8567–8591, 2009.

[8] K. An, J. Wang, R. Zhou, F. Liu, and W. Wu, "Ring-artifacts removal for photon-counting CT," *Opt. Express*, vol. 28, no. 17, pp. 25 180–25 193, 2020.

[9] D. Prell, Y. Kyriakou, and W. A. Kalender, "Comparison of ring artifact correction methods for flat-detector CT," *Phys. Med. Biol.*, vol. 54, no. 12, pp. 3881 – 3895, 2009.

[10] X. Liang, Z. Zhang, T. Niu, S. Yu, and S. Wu, "Iterative image-domain ring artifact removal in cone-beam CT." *Phys. Med. Biol.*, vol. 62, no. 13, p. 5276–5292, 2017.

[11] M. Salehjahromi, Q. Wang, Y. Zhang, L. A. Gjesteby, D. Harrison, G. Wang, P. M. Edic, and H. Yu, "A new iterative algorithm for ring artifact reduction in CT using ring total variation," *Med. Phys.*, vol. 46, no. 11, pp. 4803–4815, 2019.

[12] P. Paleo and A. Mirone, "Ring artifacts correction in compressed sensing tomographic reconstruction," *J. Synchrotron Radiat*, vol. 22, no. 5, p. 1268–1278, 2015.

[13] M. Salehjahromi, Q. Wang, Y. Zhang, L. A. Gjesteby, D. Harrison, G. Wang, P. M. Edic, and H., "A new iterative algorithm for ring artifact reduction in CT using ring total variation," *Med. Phys.*, vol. 46, no. 11, p. 4803–4815, 2019.

[14] M. Selim, E. A. Rashed, M. A. Atiea, and H. Kudo, "Sparsity-based method for ring artifact elimination in computed tomography," *PLoS One*, vol. 17, no. 6, p. e0268410, 2022.

DOCTOR: Physics-Constrained Flow Matching Method for Cone-Beam CT Reconstruction Under a Variety of Non-Ideal Imaging Conditions

Qiang Tang, Li Chen, Haoxiang Hu, Jun Xiang, Dong Liang, Yinsheng Li

***Abstract*—Three-dimensional cone-beam computed tomography (CBCT) is an essential imaging modality in modern interventional environments; however, achieving high-quality reconstruction remains challenging under the combined presence of scatter contamination, spectral nonlinearity, and sparse view sampling. In this work, we propose a learnable physics-informed flow framework, called DOCTOR, which enables accurate and robust CBCT reconstruction under a variety of non-ideal imaging conditions. DOCTOR first learns a fully differentiable and accurate forward model that characterizes system-specific non-ideal physical effects, compensating for scatter contamination and spectral nonlinearity, through extensive Monte-Carlo simulations in the pre-training phase and paired experimental data acquired by a specific spectral CBCT in the fine-tuning phase. DOCTOR integrates the learned physical forward model into a flow-based generative model that represents the distribution-level prior of ideal image features, eliminating image artifacts caused by sparse view sampling. By directly leveraging subject-specific measured data, DOCTOR ensures high reconstruction fidelity and image quality. Extensive numerical simulation studies (1,000 samples for testing), physical phantom studies, and in vivo human subject studies have been conducted to validate the feasibility and demonstrate the effectiveness of DOCTOR. DOCTOR provides a principled solution to long-standing CBCT challenges, substantially narrows the performance gap with diagnostic multi-detector-row CT, and represents a significant step toward personalized intraoperative imaging.**

***Index Terms*—CBCT; Learnable Physical Forward Model; Flow Model**

I. Introduction

THREE-dimensional cone-beam computed tomography (CBCT) has become an indispensable imaging modality in modern interventional suites. High-quality CBCT imaging has transformed intraoperative structural imaging by delivering MDCT-like image quality and diagnostic information in the interventional suite, eliminating the need to transfer patients from or to the diagnostic radiology department, thereby reducing workflow disruption and potentially improving procedural efficacy, treatment accuracy, safety, and clinical outcomes. Compared to MDCT, the image quality of flat-panel detector-based CBCT is fundamentally limited by the combined effects of scatter contamination, spectral nonlinearity, and sparse view sampling. It is a significant technical challenge to achieving MDCT-like imaging performance in the interventional suite.

The indirect two-step method sequentially corrects physical artifacts such as spectral nonlinearity and scatter contamination using hardware-based or algorithmic approaches [1], [2], followed by image reconstruction with Model-Based Iterative Reconstruction (MBIR) algorithms [3], thereby decoupling artifact correction from reconstruction to improve accuracy and robustness. In contrast, the supervised one-step method employs an end-to-end deep learning framework to directly map low-quality to high-quality images, implicitly learning to compensate for complex nonlinear artifacts without explicit physical modeling, resulting in a more integrated and potentially faster reconstruction pipeline [4].

In this work, we propose a learnable physics-informed flow framework, called DOCTOR, which enables accurate and robust CBCT reconstruction under a variety of non-ideal imaging conditions. The innovations in the proposed DOCTOR is summarized as follows:

- It enables accurate and high-quality CBCT image reconstruction under a variety of non-ideal imaging conditions, including scatter contamination, spectral nonlinearity, and sparse view sampling;
- It learns a fully differentiable and accurate forward model representing non-ideal physical effects of a specific CBCT system to compensate for scatter contamination, spectral nonlinearity using the actually acquired data of the system;
- It incorporates the learned physical forward model into a flow model representing the distribution-level regularization of ideal MDCT image features to address image artifacts caused by sparse view sampling;
- It explicitly leverages the measured data of each individual subject to ensure the accuracy and image quality for each individual subject.

II. Methed

A. Learnable Forward Modeling of Cone-Beam CT Data Acquisition Process

In ideal CBCT imaging, the object remains stationary and monochromatic X-ray spectra are used. After proper digitization, the forward model is a linear system. In practice, the imaging physics are more complex. spectral nonlinearity, scatter contamination, and electronic noise cause projection data to differ from the ideal. To better describe these effects, this paper introduces a learnable forward modeling framework. This framework maps ideal linear projection data into the nonlinear space of real physical measurements:

$$\bar{\mathbf{y}} = \mathcal{S}_{\Theta^*}(\mathbf{A}\mathbf{x}), \tag{1}$$

where, $\bar{\mathbf{y}}$ denotes the modeled non-ideal line-integral projection data, $\mathbf{A}$ denotes the system matrix, $\mathbf{x}$ denotes the image to be reconstructed, and $\mathcal{S}_{\Theta^*}$ denotes learnable forward modeling operator.

B. Image Reconstruction Using Maximum A Posterior Estimation

In CBCT reconstruction, image reconstruction can essentially be formulated as a linear inverse problem. To address this issue, maximum a posteriori (MAP) estimation provides a unified Bayesian inference framework:

$$\begin{aligned}\mathbf{x}^* &= \arg\min_{\mathbf{x}} \{-\log P(\mathbf{y} \mid \mathbf{x}) - \log P(\mathbf{x})\} \\ &= \arg\min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathcal{S}_{\Theta^*}(\mathbf{A}\mathbf{x}) - \mathbf{y}\|_{\mathbf{w}}^2 - \log P(\mathbf{x}) \right\}. \end{aligned} \tag{2}$$

Here, the data fidelity term $\|\mathcal{S}_{\Theta^*}(\mathbf{A}\mathbf{x}) - \mathbf{y}\|_{\mathbf{w}}^2$ enforces consistency between the reconstructed sinogram and the measured sinogram, while the regularization term constrains the solution space by incorporating prior information about the image. The image prior information is provided by the pre-trained flow matching model.

C. Learning $\mathcal{S}_\Theta(\mathbf{x})$ via Training a Imaging-Physics-Informed Deep Neural Network

To tackle the intellectual challenge of accurately characterizing complex physical effects, such as spectral nonlinearity and scatter contamination , this work proposes a data-driven neural network modeling strategy. Specifically, it directly learns the nonlinear mapping from an ideal linear projection space to a non-ideal physical measurement space. To characterize this complex physical degradation process, the observed non-ideal sinogram $\mathbf{y}_{\text{Non-Ideal}}$ is modeled as the result of applying a nonlinear degradation operator $\mathcal{S}_{\Theta^*}(\cdot)$ to the ideal sinogram $\mathbf{y}_{\text{Ideal}}$, namely,

$$\mathbf{y}_{\text{Non-Ideal}} = \mathcal{S}_{\Theta^*}(\mathbf{y}_{\text{Ideal}}). \tag{3}$$

To effectively learn this mapping while alleviating the limited interpretability of conventional deep learning models, we design an imaging-physics-informed U-Net architecture. However, sinograms exhibit pronounced physical properties in their structure.

By training this network, the proposed method approximates the spectral nonlinearity and scatter contamination associated with complex physical degradation processes. The network parameters are optimized by minimizing the L_2 norm between the predicted non-ideal sinograms and those non-ideal sinograms generated by simulations:

$$\Theta^* = \arg\min_{\Theta} \|\mathcal{S}_\Theta(\mathbf{y}_{\text{Ideal}}) - \mathbf{y}_{\text{Non-Ideal}}\|_2^2. \tag{4}$$

After training, $\mathcal{S}_{\Theta^*}(\cdot)$ serves as a differentiable physical degradation operator that can be embedded into subsequent iterative reconstruction frameworks as part of the data fidelity modeling.

D. Learning $\mathcal{R}_\Phi(\mathbf{x})$ via Training a Rectified Flow Matching Model

The core assumption of Rectified Flow (RF) is that the optimal transport path connecting two distributions should be the geometrically shortest path [5]. Accordingly, RF defines the forward process as a straight-line path between the two distributions. For any data pair $(\mathbf{x}_0, \mathbf{x}_1) \sim (\pi_0, \pi_1)$, the intermediate state $\mathbf{x}_t$ at time $t \in [0, 1]$ is given by the linear interpolation.

$$\mathbf{x}_t = t\,\mathbf{x}_1 + (1-t)\,\mathbf{x}_0, \qquad \mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{5}$$

To learn the evolution of the data distribution, it is necessary to model the instantaneous velocity field of $\mathbf{x}_t$ with respect to time. Taking the derivative of Eq. (6) with respect to t, the velocity field $\mathbf{v}_t$ is given by

$$\frac{\mathrm{d}\mathbf{x}_t}{\mathrm{d}t} = \mathbf{v}_t = \mathbf{x}_1 - \mathbf{x}_0. \tag{6}$$

Eq. (7) indicates that the transport velocity along the straight-line path is constant and equals the difference between the noise sample and the target image. Based on this observation, $\mathcal{R}_{\Phi^*}(\cdot)$ is constructed to approximate this target velocity field, with the training objective defined as:

$$\Phi^* = \arg\min_{\Phi} \|\mathcal{R}_\Phi(\mathbf{x}_t, t) - (\mathbf{x}_1 - \mathbf{x}_0)\|_2^2. \tag{7}$$

After training, the neural network $\mathcal{R}_{\Phi^*}(\cdot)$ learns the continuous evolution that maps the noise manifold to the image manifold. Within the MAP reconstruction framework, this model naturally serves as a generative regularization term. As shown in Fig. (1), the Imaging-Physics-Informed Deep Neural Network $\mathcal{S}_{\Theta^*}(\cdot)$ is designed to simulate the physical degradation process from an ideal sinogram to a non-ideal sinogram, while the Rectified Flow Matching Model $\mathcal{R}_{\Phi^*}(\cdot)$ is designed to learn high-quality CT image priors.

E. A Two-Step Alternating Minimization Strategy to Solve the Optimization Problem in DOCTOR

To address Eq. (2), we adopt a two-step alternating minimization strategy, which decouples the original problem into two interrelated subproblems that are solved alternately during the iterative process.

At the t-th iteration, the prior term is temporarily omitted and the current estimate $\mathbf{x}_t$ is updated solely based on the physical data fidelity. Owing to the differentiability of the physics-informed forward model $\mathcal{S}_{\Theta^*}(\cdot)$, a single gradient descent step is applied to obtain an intermediate estimate $\mathbf{x}_{t+\frac{\Delta t}{2}}$:

$$\mathbf{x}_{t+\frac{\Delta t}{2}} = \mathbf{x}_t + \lambda \cdot \nabla_{\mathbf{x}} \left(\frac{1}{2} \|\mathbf{y} - \mathcal{S}_{\Theta^*}(\mathbf{A}\mathbf{x})\|_2^2 \right) \Big|_{\mathbf{x}=\mathbf{x}_t} \tag{8}$$

where λ denotes the step size. This update propagates the projection-domain residual back to the image domain via the chain rule, enforcing physics-consistent data correction and guiding the reconstruction toward agreement with the measured sinogram.

Subsequently, the intermediate estimate $\mathbf{x}_{k+\frac{\Delta t}{2}}$ is refined using the RF based image prior:

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_{t+\frac{\Delta t}{2}} + \mathcal{R}_{\Phi^*}\left(\mathbf{x}_{t+\frac{\Delta t}{2}},\, t\right) \Delta t, \tag{9}$$

Where $\Delta t = 1/\text{step}$ denotes the time step used to discretize the rectified flow ODE, and step denotes the number of discrete steps used to evolve along the flow manifold. And the whole iterative reconstruction process is shown in Algorithm 1.

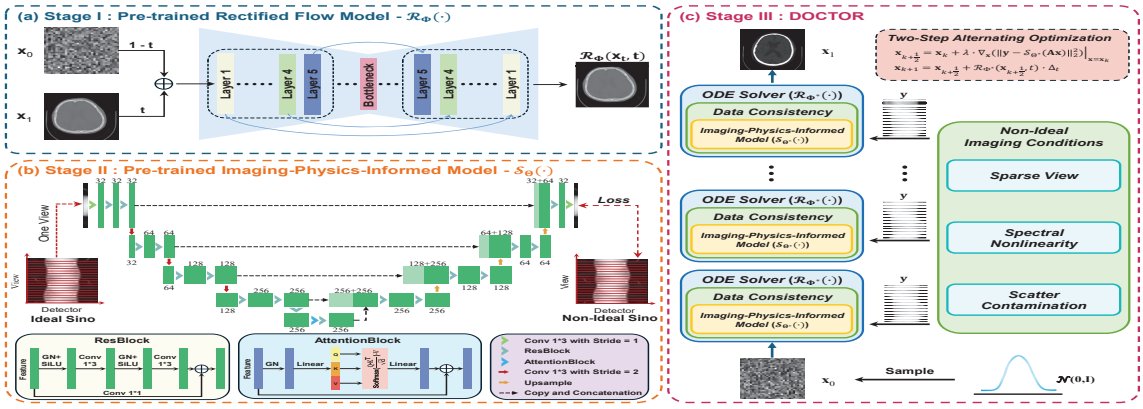


Fig. 1. **The detailed architecture of DOCTOR**. Stage I: Pre-trained Rectified Flow Model. It learn high-quality CT image priors. Stage II: Pre-trained Imaging-Physics-Informed Model. It captures the mapping between ideal sinograms to non-ideal sinograms, including noise, scatter contamination and spectral nonlinearity. Stage III: DOCTOR. During inference, data fidelity terms are integrated with physical modeling terms, while a flow model is introduced as a regularization prior.

Algorithm 1 DOCTOR

Require: Pre-trained Rectified Flow Model $\mathcal{R}_{\Phi^*}(\cdot)$, Pre-trained Imaging-Physics-Informed Model $\mathcal{S}_{\Theta^*}(\cdot)$, Measured Projection $\mathbf{y}$, Forward Operator $\mathbf{A}$ and its adjoint $\mathbf{A}^\top$, Number of Flow Steps N, Data Consistency Iterations K, Data Consistency Weight λ
Initialize variable: $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
Calculate Discretization Step Size: $\Delta t = 1/N$
for $t = 0$ **to** 1 **do**
  for $k = 0$ **to** $K - 1$ **do**
    Physics-Informed Consistency: $\bar{\mathbf{y}} \leftarrow \mathcal{S}_{\Theta^*}(\mathbf{A}\mathbf{x}_t)$
    Data Residual Computation: $\mathbf{r} \leftarrow \mathbf{y} - \bar{\mathbf{y}}$
    Back-projection Correction: $\mathbf{x}_t \leftarrow \mathbf{x}_t + \lambda \cdot \mathbf{A}^\top \mathbf{r}$
  end for
  Physics-Data Consistency Result: $\mathbf{x}_{t+\frac{\Delta t}{2}} \leftarrow \mathbf{x}_t$
  Rectified Flow Update:
  $\mathbf{x}_{t+\Delta t} \leftarrow \mathbf{x}_{t+\frac{\Delta t}{2}} + \mathcal{R}_{\Phi^*}(\mathbf{x}_{t+\frac{\Delta t}{2}}, t)\Delta t$
end for
Return: Final Reconstruction $\mathbf{x}_1$

F. Monte-Carlo Simulations Details

For the Imaging-Physics-Informed Deep Neural Network, simulated training data are generated by combining high-quality head CT volume data with Monte-Carlo based scattering simulations. Specifically, high-quality CT images are first reconstructed to obtain volumetric data [6]. The reconstructed volume, together with imaging parameters such as the X-ray energy spectrum, is then fed into the MC-GPU simulation platform to model scattering signals generated during the imaging process. When only the energy spectrum is provided as input, the corresponding air signal can be obtained from the simulation.

$$\bar{y} = \mathcal{G}(air_{\mathrm{MC}}) \cdot \exp\left(-\mathcal{P}\left(-\ln\left(\frac{\bar{y}_p}{\int \omega(E)\,\mathrm{d}E}\right)\right)\right) + \mathcal{G}(s_{\mathrm{MC}}). \tag{10}$$

Where air_{MC} and s_{MC} denote the air signal and scattering signal obtained from Monte Carlo simulation, respectively, and $\mathcal{P}(\cdot)$ and $\mathcal{G}(\cdot)$ represent the addition of Poisson noise and Gaussian smoothing. Since both air signals and scattering components are low-frequency and spatially smooth, Monte Carlo simulations with a relatively low photon count are employed, followed by Gaussian smoothing, to efficiently generate low-noise, low-frequency physical signals that more closely resemble those of real imaging systems while reducing computational cost.

After obtaining paired ideal and non-ideal projection data, image reconstruction is performed using the SART algorithm from the LEAP-CT toolkit, yielding paired ideal and non-ideal reconstructed images. Subsequently, each reconstructed slice is forward-projected to generate corresponding pairs of sinograms, which are used as training samples for the network.

III. Results

A. Qualitative and Quantitative Image Quality Assessments and Comparisons

As shown in Fig. (2), reconstruction results of different methods are compared under sparse-view acquisition conditions. And Table I summarizes the quantitative metrics of the numerical simulation experiments, reported as median ± interquartile range (IQR / 2) over 1000 slice samples.

TABLE I
QUANTITATIVE COMPARISON OF RECONSTRUCTION PERFORMANCE UNDER DIFFERENT SPARSE-VIEW CONDITIONS USING 1000 SAMPLES FOR TESTING.

Views	Metric	UNet Image Dom.	UNet Sino Dom.	DOCTOR
360	SSIM ↑	0.851 ± 0.023	0.884 ± 0.019	**0.902 ± 0.024**
	PSNR ↑	32.734 ± 0.325	35.450 ± 0.627	**39.028 ± 0.874**
120	SSIM ↑	0.671 ± 0.062	0.606 ± 0.067	**0.876 ± 0.043**
	PSNR ↑	29.502 ± 0.617	27.459 ± 1.341	**38.280 ± 0.787**

B. Validation of Learned Forward Modeling of Imaging Physics with Physical Phantom Studies

In practical clinical settings, the learned forward model $\mathcal{S}_{\Theta^*}(\mathbf{x})$ exhibits strong protocol adaptability and can be

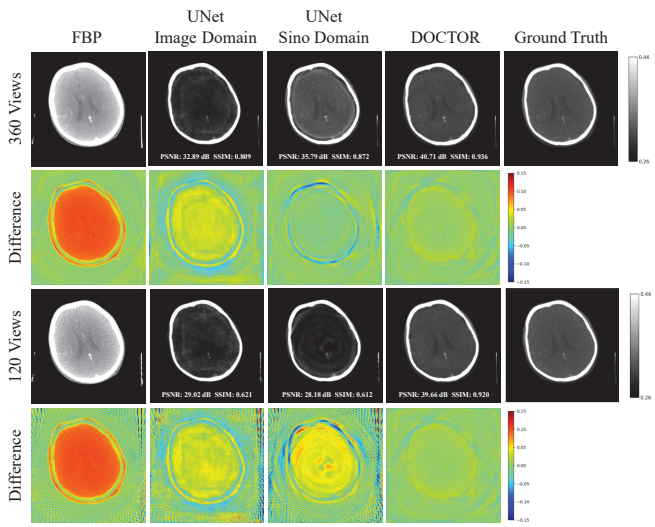


Fig. 2. **Reconstruction results on a numerical simulation under different sparse-view sampling conditions**. From left to right: ground truth, UNet Image Domain, UNet Sino Domain, and DOCTOR. From an overall image quality perspective, both U-Net baseline methods are only able to recover coarse anatomical structures under high-view settings and still suffer from noticeable detail loss and artifacts. As the number of projection views decreases, they exhibits pronounced over-smoothing, resulting in severe suppression of soft-tissue textures. In contrast, DOCTOR consistently achieves superior structural fidelity and clearer boundary delineation across all view conditions. All images are displayed with W/L: 0.26/0.44.

effectively transferred to different acquisition configurations through limited target-domain fine-tuning. As shown in Fig. (3), experiments on physical phantoms demonstrate that the fine-tuned model closely matches real non-ideal measurements and yields consistent reconstruction results.

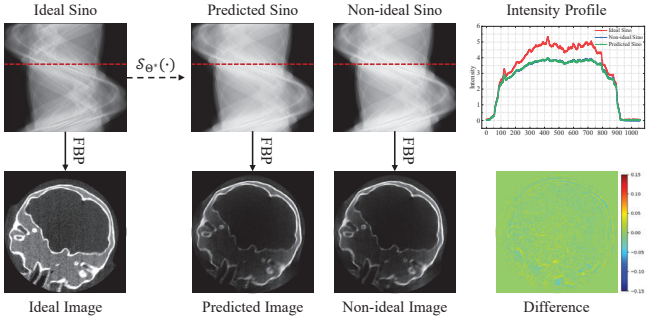


Fig. 3. **Fine-tuning of the learned forward model $\mathcal{S}_{\Theta^*}(\cdot)$ using experimental physical phantom stuies**. The top row compares the ideal, predicted, and measured non-ideal sinograms, along with a representative line profile, demonstrating that the proposed model accurately reproduces both global attenuation and local energy redistribution effects. The bottom row shows the corresponding FBP reconstructions and the residual error map, where only minor discrepancies are observed between reconstructions from predicted and real non-ideal sinograms. These results confirm that $\mathcal{S}_{\Theta^*}(\cdot)$ effectively adapts to realistic, system-specific imaging physics after fine-tuning. Images are displayed with W/L: 0.2/0.5.

C. Experimental In Vivo Human Subject Studies

Results generated from human subject studies were used to acess the overall qualitative image quality of different methods under complex anatomical conditions are shown as Fig. (4).

To comprehensively assess the adaptability of the proposed DOCTOR under diverse real-world imaging degradations, multiple acquisition scenarios are considered, including full-dose, half-dose, quarter-dose.

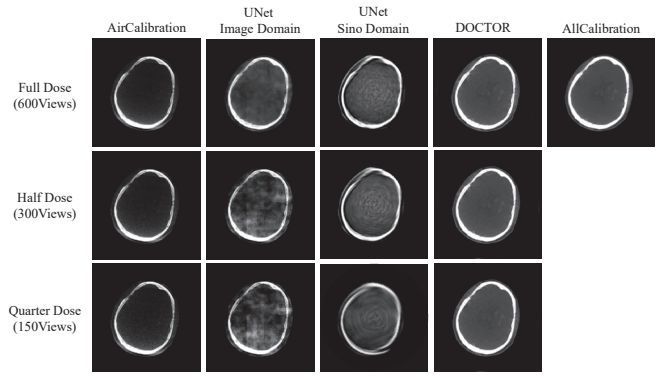


Fig. 4. **In Vivo Human Subject Experiments**. Qualitative comparison of in vivo human subject CBCT reconstructions obtained using AirCalibration, UNet Image Domain, UNet Sino Domain, and DOCTOR under different dose levels, including full dose (600 views), half dose (300 views), and quarter dose (150 views). The data were acquired using a C-arm CBCT system (uAngio 960, United Imaging Healthcare, Shanghai, China). The results illustrate the robustness of the proposed DOCTOR in suppressing noise and artifacts while preserving anatomical structures across varying sampling conditions. Images are displayed with W/L: 0.2/0.5

IV. Conclusion

This work demonstrates that high-quality CBCT reconstruction can be achieved using DOCTOR, a learnable physics-informed flow model constrained by subject-specific measured data. The proposed framework simultaneously addresses key CBCT challenges, including scatter contamination, spectral nonlinearity, and sparse-view sampling. These results indicate that DOCTOR offers a promising pathway for overcoming long-standing CBCT image quality limitations and substantially narrowing the performance gap with diagnostic MDCT in interventional imaging and therapeutic applications.

References

[1] G. Vogtmeier, R. Dorscheid, K. J. Engel, R. Luhta, R. Mattson, B. Harwood, M. Appleby, B. Randolph, and J. Klinger, "Two-dimensional anti-scatter grids for computed tomography detectors," in *Medical Imaging 2008: Physics of Medical Imaging*, vol. 6913. SPIE, 2008, pp. 1659–1669.

[2] G. Poludniowski, P. Evans, V. Hansen, and S. Webb, "An efficient monte carlo-based algorithm for scatter correction in kev cone-beam ct," *Physics in Medicine & Biology*, vol. 54, no. 12, p. 3847, 2009.

[3] D. Matenine, M. Schmittbuhl, S. Bedwani, P. Després, and J. A. de Guise, "Iterative reconstruction for image enhancement and dose reduction in diagnostic cone beam ct imaging," *Journal of X-ray Science and Technology*, vol. 27, no. 5, pp. 805–819, 2019.

[4] Y. Pang, Y. Liu, X. Chen, P.-T. Yap, and J. Lian, "Sinosynth: A physics-based domain randomization approach for generalizable cbct image enhancement," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 646–656.

[5] X. Liu, C. Gong, and Q. Liu, "Flow straight and fast: Learning to generate and transfer data with rectified flow," *arXiv preprint arXiv:2209.03003*, 2022.

[6] A. Badal and A. Badano, "Accelerating monte carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit," *Medical physics*, vol. 36, no. 11, pp. 4878–4880, 2009.

LAB-PhD: A Learnable Physics-Informed Diffusion Model for Spectral Cone Beam CT Imaging

Li Chen, Ruifeng Chen, Qiang Tang, Juan Feng, Jun Xiang, Xu Ji, Yang Chen, Dong Liang, Yinsheng Li

***Abstract*—Spectral cone-beam computed tomography (CBCT) offers material discrimination and quantitative imaging directly in the interventional suite. Although substantial efforts have been devoted to improving material basis image quality, no prior work has comprehensively addressed the combined challenges of scatter contamination, spectral distortion, sparse-view sampling, and reduced radiation dose. In this work, we develop a learnable physics-informed diffusion model, termed LAB-PhD, for accurate and high-quality material basis reconstruction under non-ideal imaging conditions. LAB-PhD first learns a fully differentiable and accurate forward model that represents non-ideal physical effects, compensating for scatter contamination and spectral distortion through extensive Monte Carlo simulations during pre-training and paired experimental data from a specific spectral CBCT system during fine-tuning. The learned physical forward model is then embedded into a diffusion model that captures the distribution-level prior of ideal material basis images, suppressing artifacts caused by sparse-view sampling and improving image quality at reduced radiation dose levels. LAB-PhD further leverages the measured data of each individual subject to ensure subject-specific accuracy and image fidelity. Extensive numerical simulations (500 samples for testing) and physical phantom studies validate the feasibility and effectiveness of LAB-PhD. Overall, LAB-PhD demonstrates a promising approach to overcoming long-standing challenges in spectral CBCT, narrowing the performance gap with diagnostic spectral multi-detector-row CT, and advancing personalized intraoperative imaging.**

***Index Terms*—Dual-Energy CBCT; Spectral CBCT; Learnable Physical Forward Model; Diffusion Model.**

I. Introduction

Spectral CBCT with conventional flat-panel detectors suffers from limited material discrimination and quantitative accuracy due to severe scatter, spectral distortion, and sparse angular sampling. Large cone angles and wide fields of view in C-arm CBCT lead to high scatter-to-primary ratios, introducing nonlinear errors in projections and energy spectra that violate material-basis reconstruction assumptions. These issues are further aggravated by polyenergetic effects from subject-dependent attenuation, contrast dynamics, and metal objects. Although rapid kVp switching enables quasi-simultaneous dual-energy acquisition, limited detector frame rates result in sparse and nonuniform angular sampling per energy, amplifying streak artifacts, noise, and bias in energy-resolved and material-basis images.

There exist indirect two-step and direct one-step approaches to address the technical challenges. Indirect two-step approaches, consisting of data correction followed by material basis image reconstruction, are the dominant paradigm for addressing the technical challenges in flat-panel spectral CBCT. More recently, direct one-step approaches have been developed by integrating scatter modeling [2] or spectrum optimization [5] into the MBIR framework for superior material basis reconstruction. Scatter-aware methods embed analytic or Monte-Carlo models into the forward operator to compensate for scatter without pre-correction. While analytic models are computationally tractable, they rely on restrictive assumptions and fail to capture subject-specific characteristics. Monte-Carlo-based models can, in principle, provide subject-specific scatter estimation, but their non-differentiable nature complicates gradient-based optimization and may degrade reconstruction accuracy. Moreover, Monte-Carlo simulations require comprehensive and accurate knowledge of the imaging system and object properties. Spectrum-optimized one-step approaches further suffer from ill-conditioned inversion and nonlinear polyenergetic measurements. In addition, incorporating appropriate priors remains challenging, as handcrafted priors (e.g., smoothness or sparsity) may introduce bias in material quantification.

Although substantial progress has been made in spectral CBCT material basis reconstruction, existing methods fail to jointly address multiple non-ideal imaging conditions, including scatter contamination, spectral distortion, sparse-view sampling, and reduced radiation dose. To this end, we propose a learnable physics-informed diffusion model for spectral CBCT, termed LAB-PhD, which integrates a fully differentiable, system-specific physical forward model into a diffusion framework that captures the distribution-level prior of ideal material basis images. The main contributions of this work are summarized as follows:

- LAB-PhD achieves accurate and high-quality material basis reconstruction under diverse non-ideal imaging conditions, including scatter, spectral distortion, sparse views, and low dose;
- A fully differentiable forward model is learned to explicitly model non-ideal physical effects, leveraging large-scale Monte Carlo simulations for pre-training and paired experimental data for system-specific fine-tuning;
- The learned forward model is seamlessly embedded into a pre-trained diffusion prior, effectively suppressing sparse-view artifacts and improving reconstruction quality at reduced dose levels;
- Subject-specific measured data are explicitly incorporated to ensure reconstruction accuracy and robustness for individual subjects.

II. LAB-PhD: A Learnable Physics-Informed Diffusion Model for Spectral Cone Beam CT

A. Learnable Forward Modeling of Spectral Cone-Beam CT

To enable model-based iterative reconstruction of material basis images, we develop a forward model for dual-energy CBCT based on the polychromatic Beer–Lambert law, explicitly incorporating scatter effects and energy-dependent detector

response. Accordingly, the projection signal under the e-th energy spectrum can be expressed as

$$\bar{\mathbf{y}}_e = \int_{\omega_e} \omega_e(E)\, \exp\left(-\sum_{k=1}^{K} \mu_k(E)\, \mathbf{A}\, \mathbf{c}_k\right) R(E)\, dE + \mathbf{s}_e, \tag{1}$$

where ω_e denotes the e-th incident X-ray source spectrum, $\mu_k(E)$ the linear attenuation coefficient of the k-th basis material, $\mathbf{A}$ the system matrix, $\mathbf{c}_k$ the voxel-wise coefficients of the k-th basis material, $R(E)$ the energy-weighting response function, and $\mathbf{s}_e$ the scatter signal corresponding to the e-th energy spectrum.

To enable backpropagation for material-basis MBIR, the entire forward model must be differentiable end-to-end. While the differentiable forward projector is straightforward, the main challenges arise from scatter modeling and source–detector coupling. Scatter is modeled using a pre-trained CNN $\mathcal{S}_\theta(\cdot)$, providing accurate, efficient, and differentiable sample-specific prediction. The source–detector coupling is represented by a learnable effective spectrum $\boldsymbol{\omega}_e(E) = \omega_e(E) \cdot R(E)$, which jointly accounts for the source spectrum and detector response and is self-calibrated during optimization. These components together form a fully learnable forward model for spectral CBCT imaging, expressed as:

$$\bar{\mathbf{y}}_e(\mathbf{c}_k, \boldsymbol{\omega}_e) = \bar{\mathbf{y}}_{e,\mathrm{p}}(\mathbf{c}_k, \boldsymbol{\omega}_e) + \mathcal{S}_\theta(\bar{\mathbf{y}}_{e,\mathrm{p}}(\mathbf{c}_k, \boldsymbol{\omega}_e)), \tag{2}$$

$$\bar{\mathbf{y}}_{e,\mathrm{p}}(\mathbf{c}_k, \boldsymbol{\omega}_e) = \int \boldsymbol{\omega}_e(E)\, \exp\left(-\sum_{k=1}^{K} \mu_k(E)\, \mathbf{A}\, \mathbf{c}_k\right) dE, \tag{3}$$

where $\bar{\mathbf{y}}_{e,\mathrm{p}}(\mathbf{c}_k, \boldsymbol{\omega}_e)$ denotes the signal before scatter contamination, i.e., the primary signal.

B. Objective Function Based on Maximum A Posteriori Estimation

The optimization objective of the MBIR method can be derived based on the maximum a posteriori (MAP) estimation principle. Let $\mathbf{C}$ denote the material basis images, $\boldsymbol{\Omega}$ denote the spectra, $\mathbf{Y}$ the observed data, and $\bar{\mathbf{Y}}$ the data predicted by the forward model. Bold uppercase letters denote collections of the corresponding lowercase variables across different material types or energy channels. Then the optimization objective can be expressed as follows:

$$\begin{aligned} \mathbf{C}^*, \boldsymbol{\Omega}^* &= \arg\min_{\mathbf{C},\boldsymbol{\Omega}} \left[-\log P(\mathbf{Y}|\bar{\mathbf{Y}}) - \log P(\bar{\mathbf{Y}})\right] \\ &= \arg\min_{\mathbf{C},\boldsymbol{\Omega}} \left[L(\mathbf{C},\boldsymbol{\Omega}) + R(\mathbf{C},\boldsymbol{\Omega})\right], \end{aligned} \tag{4}$$

where $L(\mathbf{C},\boldsymbol{\Omega})$ is the data fidelity term reflecting the consistency between the forward-modeled and the measured data, and $R(\mathbf{C},\boldsymbol{\Omega})$ is the regularization term incorporating prior information or constraints on the solution.

1) Data Fidelity: The overall measurement noise can be modeled as a Gaussian process with heteroscedastic variance [6], i.e.,$\mathbf{Y} \sim \mathcal{N}(\bar{\mathbf{Y}}(\mathbf{C},\boldsymbol{\Omega}), \boldsymbol{\Sigma})$, where the covariance matrix $\boldsymbol{\Sigma}$ reflects the signal-dependent noise level of each detector element. This statistical assumption naturally leads to the weighted least squares (WLS) data fidelity formulation:

$$L_1(\mathbf{C},\boldsymbol{\Omega}) = \frac{1}{2}(\mathbf{Y}-\bar{\mathbf{Y}}(\mathbf{C},\boldsymbol{\Omega}))^\top \mathbf{W}(\mathbf{Y}-\bar{\mathbf{Y}}(\mathbf{C},\boldsymbol{\Omega})), \mathbf{W} = \boldsymbol{\Sigma}^{-1}, \tag{5}$$

where the weighting matrix $\mathbf{W}$ effectively encodes the reciprocal of the variance of measurement noise. Additionally, an SSIM loss is incorporated in the post-log domain to complement the statistical data fidelity term, providing a perceptually guided constraint that promotes the preservation of structural information:

$$L_2(\mathbf{C},\boldsymbol{\Omega}) = 1 - \mathrm{SSIM}\left(-\ln\frac{\mathbf{Y}}{\mathbf{I}_{\text{air}}},\ -\ln\frac{\bar{\mathbf{Y}}(\mathbf{C},\boldsymbol{\Omega})}{\mathbf{I}_{\text{air}}}\right), \tag{6}$$

where $\mathbf{I}_{\text{air}}$ represents the flat-field signal. The final data-fidelity term is formulated as a weighted combination of the two fidelity terms:

$$L(\mathbf{C},\boldsymbol{\Omega}) = L_1(\mathbf{C},\boldsymbol{\Omega}) + \lambda L_2(\mathbf{C},\boldsymbol{\Omega}), \tag{7}$$

which jointly enforces statistical consistency and structural preservation, thereby improving the robustness of the reconstruction.

2) Regularization: We incorporate a pre-trained diffusion model as an implicit regularization prior in the reconstruction framework, which captures rich, data-driven anatomical priors by learning the gradient of the image distribution, thereby guiding each iteration toward anatomically plausible, high-fidelity solutions while suppressing noise and artifacts. Specifically, we adopt a physics-constrained and noise-controlled diffusion model [7], denoted as $\mathcal{D}_\phi(\cdot)$, to ensure stable convergence and robust generalization across diverse CT reconstruction settings.

C. Optimization Algorithm

Analysis of the forward model and objective function reveals some coupled variables or terms requiring joint optimization: (1) the material basis images and the spectrum, and (2) the data fidelity term and implicit regularization term. To handle these couplings, a two-stage alternating optimization strategy is adopted, and the overall framework is illustrated in Figure 1.

1) Initialization Stage: The first stage formulates a differentiable forward model and alternately optimizes the spectrum $\boldsymbol{\Omega}$ and material basis images $\mathbf{C}$. The initial spectrum $\boldsymbol{\Omega}_0$ is estimated from half-value layer (HVL) information, while the initial material images $\mathbf{C}_0$ are obtained via analytical image-domain decomposition of uncorrected data and thus contain noise and artifacts. Dual-energy forward projection is used to generate the primary signal, which is then processed by a pre-trained scatter network $\mathcal{S}_{\theta^*}$ to estimate the scatter component. The final simulated projection is formed by combining the primary and scatter signals, followed by post-logarithmic transformation and statistical weighting to construct the data fidelity term. In this stage, $\mathbf{C}$ and $\boldsymbol{\Omega}$ are jointly treated as learnable variables and alternately updated via gradient-based optimization enabled by the differentiable forward model.

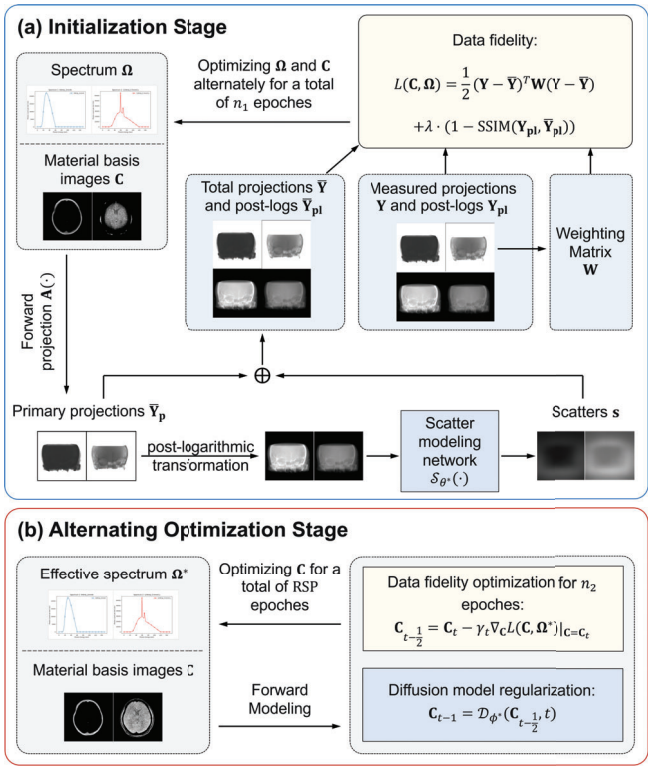


Fig. 1. The Two-Stage Alternating Optimization Workflow. Panel (a), corresponding to the **Initialization Stage**, involves the formulation of the forward model, as well as the alternating optimization of the spectrum $\mathbf{\Omega}$ and the material basis images $\mathbf{C}$, **resulting in the calibrated effective spectrum $\mathbf{\Omega}^*$ and the warm initialization of the material basis images $\mathbf{C}_{ini}$, which are used to initialize the next stage**. Panel (b), corresponding to the **Alternating Optimization Stage**, illustrates a process that **optimizes $\mathbf{C}$ by alternately applying data fidelity and implicit regularization**.

$$\mathbf{C}_{t+1} = \mathbf{C}_t - \alpha_t\nabla_{\mathbf{C}}L(\mathbf{C},\mathbf{\Omega}_t)|_{\mathbf{C}=\mathbf{C}_t} \tag{8}$$

$$\mathbf{\Omega}_{t+2} = \mathbf{\Omega}_t - \beta_t\nabla_{\mathbf{\Omega}}L(\mathbf{C}_{t+1},\mathbf{\Omega})|_{\mathbf{\Omega}=\mathbf{\Omega}_t} \tag{9}$$

$$t = 0, 2, 4, \ldots, (n_1 - 2)$$

This procedure is performed for a total of n_1 iterations, resulting in the calibrated effective spectrum $\mathbf{\Omega}^*$ and the warm initialization of the material basis images $\mathbf{C}_{ini}$.

2) Alternating Optimization Stage: The second stage performs iterative refinement by alternating data fidelity optimization and implicit regularization for RSP iterations. Initialized with $\mathbf{C}_{ini}$ and the fixed effective spectrum $\mathbf{\Omega}^*$, the material basis images are updated via the data-fidelity term for n_2 times and refined by a pre-trained diffusion model $\mathcal{D}_{\phi^*}(\cdot)$ at each iteration, to suppress noise and artifacts while preserving anatomical structures.

$$\mathbf{C}_{t-\frac{1}{2}} = \mathbf{C}_t - \gamma_t\nabla_{\mathbf{C}}L(\mathbf{C},\mathbf{\Omega}^*)|_{\mathbf{C}=\mathbf{C}_t} \tag{10}$$

$$\mathbf{C}_{t-1} = \mathcal{D}_{\phi^*}\left(\mathbf{C}_{t-\frac{1}{2}},t\right) \tag{11}$$

$$t = \mathrm{RSP}, \mathrm{RSP}-1, \ldots, 1$$

D. Implementation Details

1) Pre-training Followed by System-specific Fine-tuning Strategy: Owing to the learnable forward model, our method can be adapted to different real systems and acquisition protocols via a pre-training and system-specific fine-tuning strategy. Specifically, the scatter modeling network is first pre-trained on large-scale Monte Carlo-simulated dual-energy data. For deployment, a small amount of system data acquired from representative phantoms is used to fine-tune the scatter model, yielding a system-adaptive forward model and high-quality material basis reconstruction.

III. Materials and Methods

The simulated data were generated by combining the forward-projection of material basis images with Monte Carlo-simulated scatter, where the material basis images were derived from clinical head CT data. In physical experiments, the imaging platform was assembled using commercial-grade X-ray sources and detectors, with an anthropomorphic head phantom employed as the sample.

We compared three categories of material basis reconstruction methods. The Analytic Method applies scatter and beam-hardening pre-corrections followed by denoising post-processing [1]. The learning-based methods map non-ideal data to ideal ones in either image domain [3] or projection domain [4]. The Joint MBIR Method explicitly incorporates scatter into the forward model [2], but is limited by the accuracy and adaptability of scatter estimation. Several evaluation metrics, including MAE (Mean Absolute Error), SSIM (Structural Similarity Index Measure), and MAPE (Mean Absolute Percentage Error), are employed to assess the performance of material basis image reconstruction or scatter prediction. The quantitative metrics are evaluated over 500 slice samples from 5 digital phantoms, as illustrated in table I.

IV. Results

Figure 2: Numerical Simulations.

Figure 3: Evaluation of Scatter Prediction Accuracy for Physical Experiments.

Figure 4: Numerical Simulation and Physical Experiment Results Under Different Non-ideal Conditions.

References

[1] M. Sun and J. M. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Phys. Med. Biol.*, vol. 55, no. 22, pp. 6695–6707, Nov. 2010.

[2] A. Lorenzon, S. Z. Liu, X. Jiang, G. J. Gang, and J. W. Stayman, "Joint material decomposition and scatter estimation for spectral CT," in *Proc. Int. Conf. Image Form. X-Ray Comput. Tomogr.*, 2024, p. 186.

[3] W. Zhang et al., "Image domain dual material decomposition for dual-energy CT using butterfly network," *Med. Phys.*, vol. 46, no. 5, pp. 2037–2051, May 2019.

[4] J. Zhu et al., "Feasibility study of three-material decomposition in dual-energy cone-beam CT imaging with deep learning," *Phys. Med. Biol.*, vol. 67, no. 14, Art. no. 145012, Jul. 2022.

[5] R. F. Barber, E. Y. Sidky, T. G. Schmidt, and X. Pan, "An algorithm for constrained one-step inversion of spectral CT data," *Phys. Med. Biol.*, vol. 61, no. 10, pp. 3784–3806, May 2016.

[6] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography," *IEEE Trans. Med. Imaging*, vol. 25, no. 10, pp. 1272–1283, Oct. 2006.

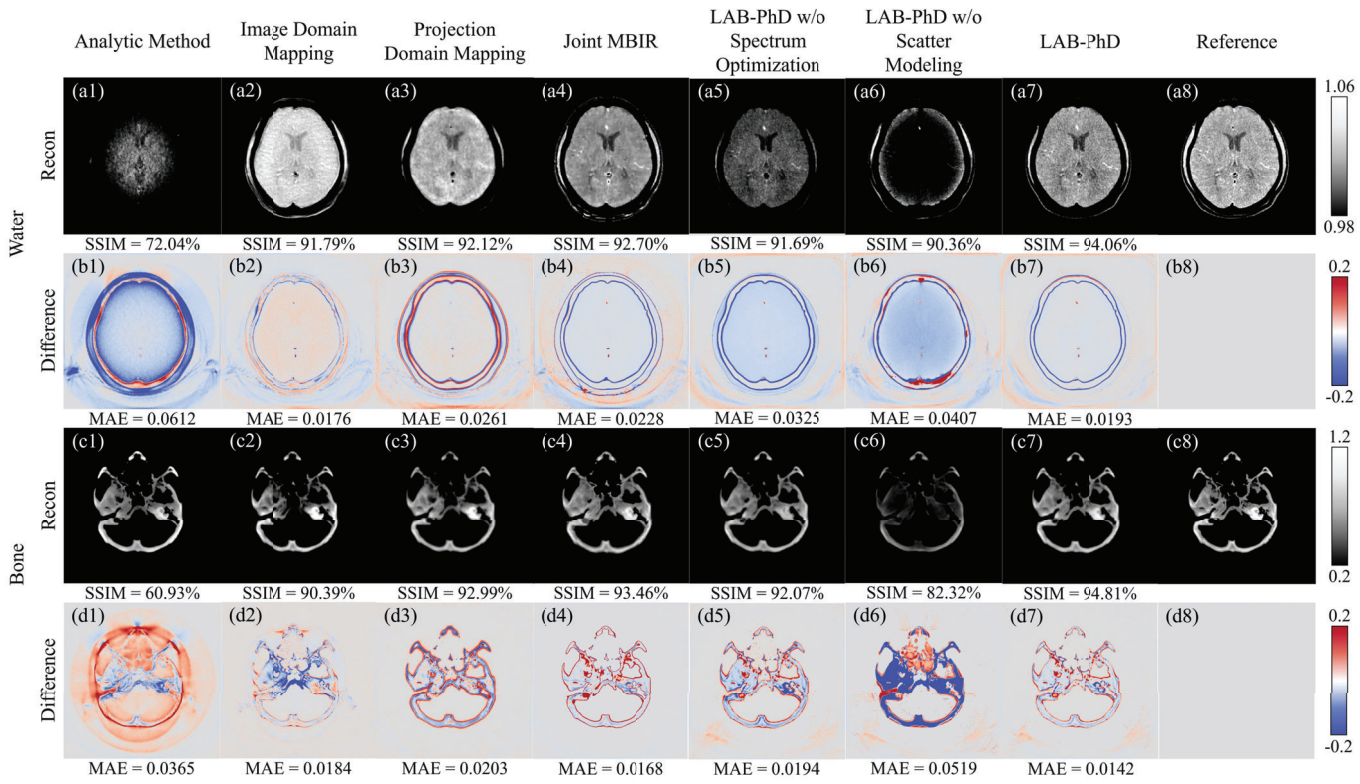


Fig. 2. Numerical Simulations. Each column corresponds to the **Analytic Method**, **Image-Domain Network Mapping**, **Projection-Domain Network Mapping**, **Joint MBIR**, **LAB-PhD**, and the **Ground Truth**, respectively. Series (a) and (b) present the reconstructed water images and their corresponding difference maps with respect to the ground truth. Series (c) and (d) show the bone images and their difference maps. The LAB-PhD method achieves **high-quality material-basis reconstruction under scatter contamination, spectral distortion, sparse-view, and low-dose acquisition**.

TABLE I
QUANTITATIVE METRICS (MEDIAN ± IQR/2) OF NUMERICAL SIMULATION EXPERIMENTS OVER 500 SLICE SAMPLES FROM 5 DIGITAL PHANTOMS

		Analytic Method	Image Domain Mapping	Projection Domain Mapping	Joint MBIR	LAB-PhD
Water	MAE ↓	0.0674 ± 0.0079	0.0236 ± 0.0075	0.0335 ± 0.0066	0.0431 ± 0.0086	0.0242 ± 0.0070
	SSIM (%) ↑	69.44 ± 3.25	89.41 ± 1.58	89.96 ± 1.60	90.53 ± 2.56	92.00 ± 2.64
Bone	MAE ↓	0.0293 ± 0.0025	0.0111 ± 0.0035	0.0153 ± 0.0027	0.0113 ± 0.0032	0.0070 ± 0.0025
	SSIM (%) ↑	65.25 ± 1.81	93.18 ± 1.84	94.46 ± 1.11	95.13 ± 1.25	98.22 ± 1.04

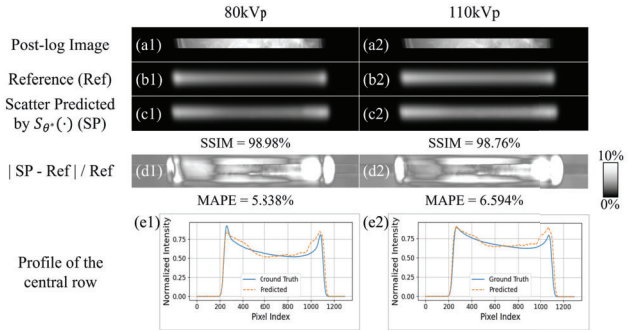


Fig. 3. Evaluation of Scatter Prediction Accuracy for Physical Experiments. (a1) and (a2) show the dual-energy post-log projections of the real head phantom within the collimator range. (b1) and (b2) display the extracted reference scatters. (c1) and (c2) show **the scatters predicted by the fine-tuned network** $\mathcal{S}_{\theta^*}(\cdot)$. (d1) and (d2) illustrate the relative error maps. (e1) and (e2) show central-row scatter profiles.

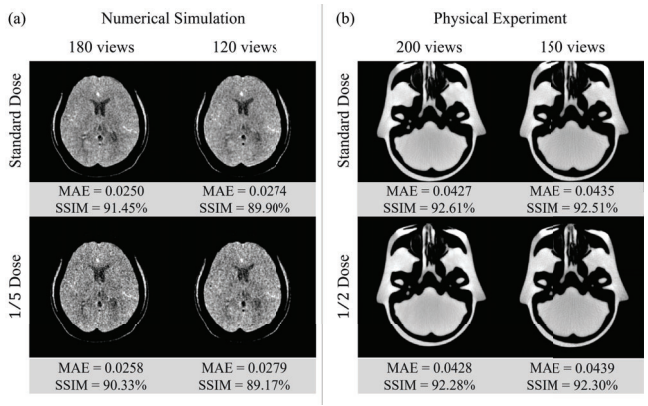


Fig. 4. Numerical Simulation and Physical Experiment Results Under Non-ideal Conditions. Panel (a) presents the numerical simulation experiments under **low-dose (1/5 dose) and sparse-view (180 and 120 views) acquisition settings**. Panel (b) presents the physical phantom experiments under **low-dose (1/2 dose) and sparse-view (200 and 150 views)** acquisition settings.

[7] R. Chen, Z. Zhang, G. Quan, Y. Du, Y. Chen, and Y. Li, "PRECISION: A physics-constrained and noise-controlled diffusion model for photon-counting computed tomography," *IEEE Trans. Med. Imaging*, 2024.

Dual-Contrast K-Edge Imaging on a Clinical PCCT: Influence of K-Edge Spacing

Martin V. Rybertt[1,2], Katie E. Villasenor[1,2], Nicole Mirzaian[1,2], Ryan A. Fair[1], Pooyan Sahbaee[3], Tristan Nowak[4], Leening P. Liu[1], Ali H. Dhanaliwala[1], David P. Cormode[1,2], Peter B. Noël[1]

***Abstract*—Photon-counting CT (PCCT) enables K-edge imaging for quantitative separation of high-Z contrast agents and may support dual-contrast imaging without repeat acquisitions. However, performance on clinically available PCCT systems has not been systematically characterized as a function of key K-edge experimental parameters. We evaluated dual-contrast K-edge material decomposition on a clinical dual-source PCCT scanner using a thorax phantom containing pairwise mixtures of K-edge agents at controlled ratios and concentrations. Gadolinium (Gd), holmium (Ho), erbium (Er), and tungsten (W) were prepared in six pairwise combinations spanning K-edge energy separations of approximately 5–19 keV, with mixture concentrations ranging from 0–10 mg/mL plus single-agent controls. Each configuration was scanned at 140 kVp and 12 mGy CTDIvol using three distinct energy-bin settings, with three repeat acquisitions to assess reproducibility. Quantitative accuracy (concentration error) and image noise were computed from ROI analysis across multiple slices in the material-specific maps. Reliable dual-contrast decomposition was demonstrated for K-edge separations as small as 5.4 keV (Gd:Ho). Accuracy depended primarily on K-edge separation, while energy-bin selection most strongly influenced performance for closely spaced K-edges; for Gd:Ho, Ho error ranged from 0.7–1.0 mg/mL across bin settings. Image noise was consistently affected by both K-edge separation and bin configuration: Ho noise ranged from 0.6–1.0 mg/mL for Gd:Ho and decreased to 0.2–0.5 mg/mL for Ho:W (ΔE=13.9 keV). These results delineate practical tradeoffs between K-edge separation and energy-bin design on a clinical PCCT platform, expanding the range of feasible high-Z contrast pairs and supporting quantitative dual-contrast imaging applications.**

***Index Terms*—Photon-counting CT, Contrast-enhanced CT, K-edge Imaging, Dual-contrast Imaging**

I. Introduction

The ability to distinguish and quantitatively map multiple contrast agents in a single CT acquisition would enable multiparametric imaging while reducing scan time, motion, and misregistration associated with sequential injections or repeat scans [1], [2], [3], [4], [5], [6]. Such capability could support applications where two physiological processes must be interrogated concurrently, for example, separating blood-pool enhancement from targeted or delayed uptake, or disentangling perfusion-related signal from tissue-specific accumulation. However, conventional CT provides a single attenuation measurement per voxel, making mixed compositions difficult to interpret and limiting specificity when multiple agents are present.

Dual-energy CT extends this paradigm by improving discrimination among materials and enabling basic decomposition approaches [7]. Yet, simultaneous separation of multiple high-Z agents remains challenging in practice when attenuation spectra are similar, when contrast concentrations are low, or when more than two basis materials are required. These scenarios can lead to ill-conditioned decompositions with increased cross-talk and noise amplification, limiting both detectability and quantitative accuracy [8]. Photon-counting CT (PCCT) offers a path forward by directly measuring energy-resolved photon counts and enabling spectral imaging with multiple energy bins [9]. Importantly, PCCT can exploit K-edge discontinuities as material-specific signatures, potentially enabling robust identification and quantification of high-Z agents that are otherwise difficult to separate using conventional dual-energy approaches [10], [11].

Despite this promise, the practical performance for dual-contrast K-edge imaging on clinically available PCCT systems is not yet well defined. In particular, given the fixed kVp setting, two factors are expected to strongly determine whether simultaneous decomposition is feasible: (i) the separation between contrast-agent K-edge energies and (ii) the placement of energy-bin thresholds that define the measured spectral channels. As K-edges become more closely spaced, the spectral signatures overlap more strongly and decomposition conditioning worsens; in that regime, bin selection may become a dominant driver of cross-talk and uncertainty. Conversely, larger K-edge separations may improve separability but still interact with bin placement through dose partitioning and bin-dependent noise. A systematic experimental characterization of these tradeoffs is needed to guide contrast-pair selection and acquisition design, especially for translation of dual-contrast PCCT protocols beyond proof-of-concept demonstrations.

In this work, we evaluate dual-contrast K-edge material decomposition on a clinical dual-source PCCT platform using controlled pairwise mixtures of high-Z contrast agents spanning a wide range of K-edge separations and multiple energy-bin configurations. We quantify quantitative accuracy

We acknowledge support through the National Institutes of Health (R01EB035908) and Siemens Healthineers. M.V. Rybertt, K. E. Villasenor, N. Mirzaian and D.P. Cormode are with the Department of Bioengineering and Department of Radiology at the University of Pennsylvania, Philadelphia, PA 19104 USA. R.A. Fair, L.P. Liu, A.H. Dhanaliwala and P.B. Noël are with the Department of Radiology at the University of Pennsylvania, Philadelphia, PA 19104 USA. P. Sahbaee is with Siemens Healthineers, Malvern, PA 19355 USA. T. Nowak is with Siemens Healthineers AG, Forchheim, Germany. (email: mrybertt@seas.upenn.edu)

and noise across repeat acquisitions and assess how performance varies with both K-edge separation and bin design. By intentionally including closely spaced K-edge pairs and multiple bin configurations, we delineate where decomposition remains stable and where performance degrades, providing practical guidance for selecting contrast pairs and scan settings for quantitative dual-contrast PCCT imaging.

II. Methods

A. Experimental setup

Four K-edge elements that have been used in prior spectral and PCCT studies [12] - gadolinium (Gd, K-edge: 50.2 keV), holmium (Ho, 55.6 keV), erbium (Er, 57.5 keV), and tungsten (W, 69.5 keV) - were selected to span a range of K-edge energies and pairwise separations. Stock solutions were combined to form six pairwise mixtures (**Table 1**). For each material pair, a fixed total concentration of 10 mg/mL was maintained while the composition was varied across seven mixing ratios (A:B = 10:0, 8:2, 6:4, 5:5, 4:6, 2:8, and 0:10 mg/mL). In addition, single-material controls at 5 mg/mL were prepared for each agent. Each solution was loaded into 5 mL syringes and arranged in a custom holder surrounded by a water-equivalent background to provide consistent attenuation conditions. The holder was then positioned within a commercial thorax phantom (QRM Thorax, QRM, Möhrendorf, Germany) and scanned under the study's acquisition protocols.

Table 1. Material contrast pairs and K-edge energy difference

Material A	Material B	K-edge energy difference [keV]
Holmium	Erbium	1.9
Gadolinium	Holmium	5.4
Gadolinium	Erbium	7.3
Erbium	Tungsten	12.0
Holmium	Tungsten	13.9
Gadolinium	Tungsten	19.3

B. Scanning protocol

The phantom was scanned using a clinically available dual-source PCCT scanner (NAEOTOM Alpha.Peak, Siemens Healthineers, Forchheim, Germany). Scans were performed using a tube voltage of 140 kVp at a volumetric CT dose index (CTDI vol) of 12 mGy (**Table 2**). To evaluate the impact of bin configuration on material decomposition, the phantom was scanned with each available energy bin configuration: 20/40/56/75 (A), 20/52/75/82 (B), 20/56/72/90 (C) keV. All scans were repeated three times to ensure reproducibility.

C. Material decomposition

Material decomposition was performed using the calibration-based approach described in Neumann et al [13]. During calibration, stock solutions of K-edge materials were prepared and placed within a solid water-equivalent phantom provided by the vendor. Solutions were repeatedly scanned while varying phantom diameter, tube current, and isocenter distance to establish a set of expected values for estimation. Subsequently, during reconstruction, water correction was applied to the images, followed by a least squares estimation using the ground truth derived from calibration to generate material-specific images for each target material.

D. Image reconstruction

CT raw data was transferred to a vendor-specific workstation (ReconCT, Siemens Healthineers, Forchheim, Germany) for reconstruction into conventional and spectral post-processing (SPP) images. Reconstruction parameters included a Qr40 kernel, quantum iterative reconstruction level of 3, a field of view (FOV) of 150 mm, and matrix size of 512 x 512 px (**Table 2**). Material decomposition was performed for each target material (Water, Material A, Material B) using vendor-specific software (eXamine, Siemens Healthineers, Forchheim, Germany).

Table 2. Scanning and reconstruction parameters

Tube voltage [kVp]	140
$CTDI_{vol}$ [mGy]	12
Bin configuration [keV]	20/40/56/75, 20/52/75/82, 20/55/72/90
Collimation [mm]	96 x 0.4
Rotation time [s]	0.25
Reconstruction kernel/QIR	QIR/3
Slice thickness [mm]	2
Field of view [mm]	150 x 150
Matrix size [px]	512 x 512

E. Image analysis

Regions of interest (ROI) were placed on each sample solution at conventional images, covering 60% of the sample diameter. ROI coordinates were saved and applied across the ten central slices of each material map to measure the mean concentration and standard deviation (noise).

Material-specific images of different contrast pairs were generated to enable the visual assessment of the effect of the difference in K-edge energy. For each contrast pair, the average concentration error and noise were estimated and plotted with respect to the K-edge energy separation between the material. Samples containing W at a 2:8 mg/mL ratio were excluded from the analysis as they sedimented during the acquisition.

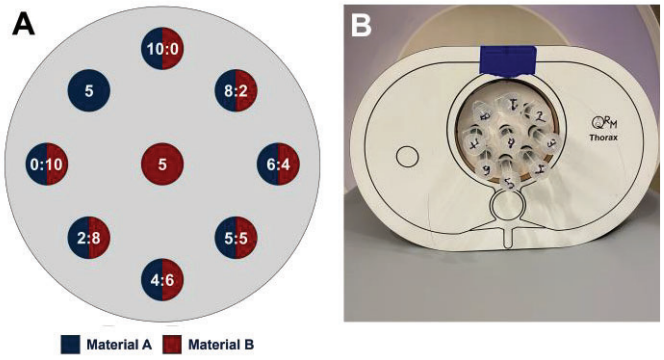


Figure 1. Phantom configuration and experimental setup. 5 mL syringes loaded with samples were placed inside a water-equivalent holder following the concentrations described in A. The holder was inserted into a thoracic phantom (B) and scanned.

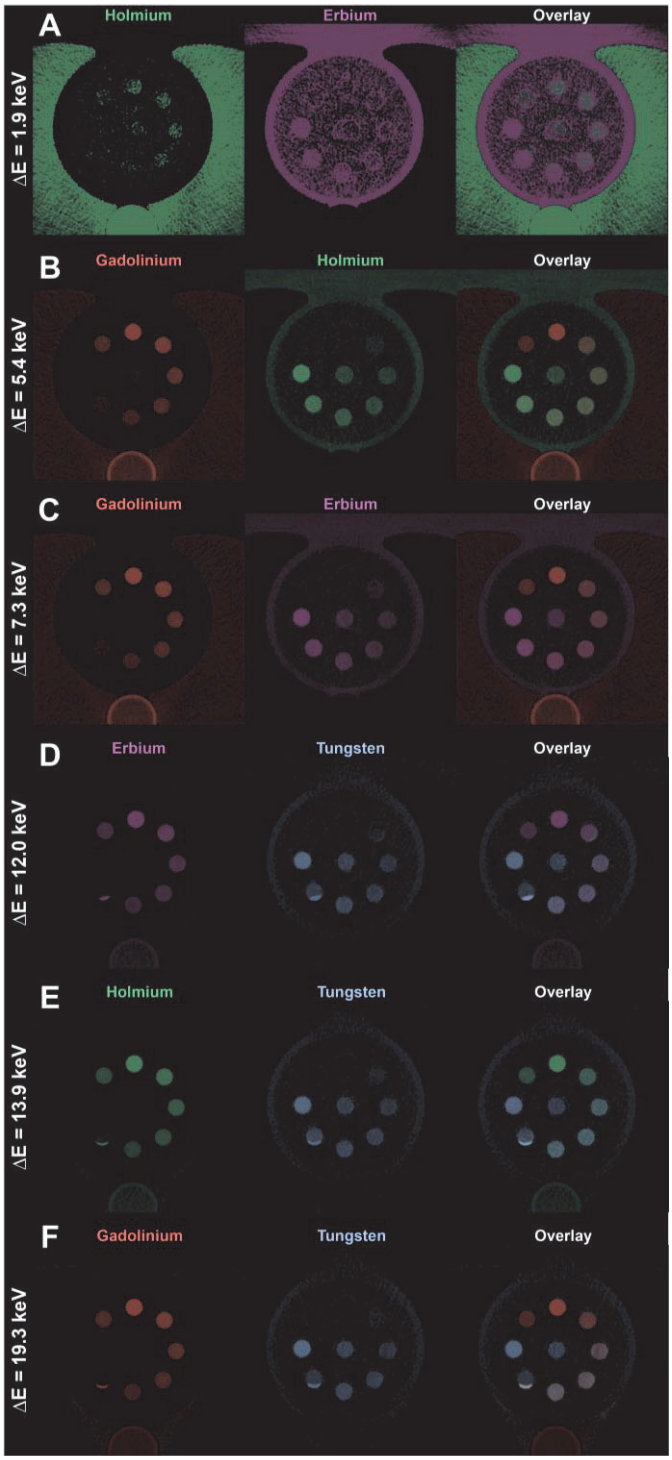


Figure 2. Material-specific images of contrast pairs using energy bin configuration of 20/55/72/90 keV. Contrast pairs of Gd:Er (A, ΔE: 1.9 keV) exhibit higher background noise and mischaracterization of samples, successful material decomposition can be observed starting at K-edge difference of 5.4 keV (B, Gd:Ho), performance of material decomposition improves for pairs with larger K-edge differences such as Gd:Er (C), Er:W (D), Ho:W (E), and Gd:W (F). Samples of 2:8 mg/mL containing W sedimented and were excluded from the analysis. WL/WW: 5/10 mg/mL.

III. Results

Overall feasibility and qualitative performance. Dual-contrast material decomposition was feasible for most material pairs with K-edge separations >5 keV across the evaluated energy-bin configurations, producing material-specific maps that correctly localized targets and reproduced the expected concentrations. Visual assessment indicated (**Figure 2**) that closely spaced K-edges were associated with elevated background noise, which reduced conspicuity for low-concentration samples, whereas pairs with larger K-edge separation showed improved visual detectability.

Quantitative accuracy as a function of K-edge separation. Mean concentration error decreased with increasing K-edge separation (**Figure 3**), consistent with improved conditioning of the decomposition for more spectrally distinct materials. For the closest feasible pair, Gd:Ho (ΔE = 5.4 keV), the Gd-map errors were 0.8 ± 0.4, 0.7 ± 0.6, and 1.0 ± 0.4 mg/mL for energy-bin configurations A, B, and C, respectively. For a widely separated pair, Gd:W (ΔE = 19.3 keV), corresponding Gd-map errors were smaller: −0.1 ± 0.1, 0.2 ± 0.1, and 0.2 ± 0.1 mg/mL across the same bin configurations. Notably, the highest accuracy within the reported examples occurred for Gd:Er (ΔE = 7.3 keV), where Gd-map error was near zero across bin settings (0.0 ± 0.1, −0.2 ± 0.2, 0.1 ± 0.1 mg/mL for A, B, C).

Impact of energy-bin configuration on quantitative accuracy. Across most contrast pairs, energy-bin configuration had a comparatively modest effect on mean error when K-edge separation was moderate-to-large, whereas it exerted a larger influence for closely spaced K-edges. An important exception was tungsten quantification: for W-containing pairs, **Figure 3** indicates a pronounced bin-dependent systematic bias in the tungsten maps, with certain bin settings producing substantially larger negative bias than others even at larger K-edge separations (e.g., Er:W and Gd:W). This suggests that, beyond spectral overlap considerations, protocol optimization (i.e., bin placement) can influence tungsten accuracy, likely through how the selected thresholds partition photons above and below the high-energy K-edge region, so bin selection remains critical not only for closely spaced pairs but also material specific.

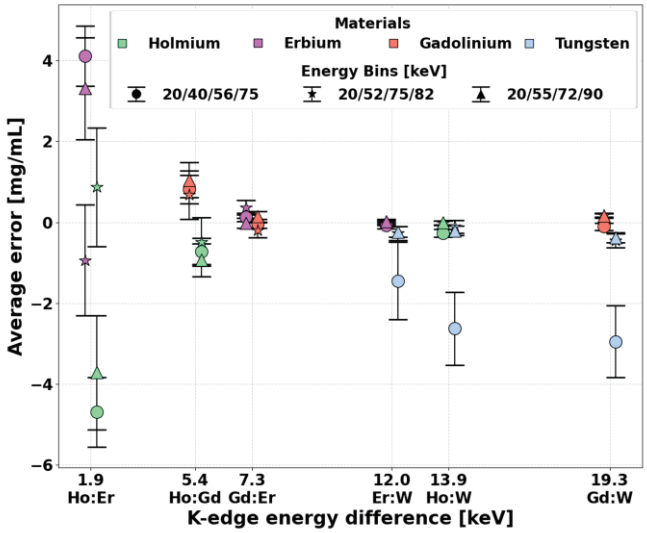


Figure 3. Average error of material contrast pairs with respect to K-edge energy difference and energy bin configuration. Error demonstrated a decreasing trend as K-edge energy separation increased. Error bars represent the standard deviation of mean values collected across all slices.

Noise and image quality trends. Noise in the material-specific maps also decreased as K-edge separation increased (**Figure 4**), and unlike mean error, noise was consistently sensitive to energy-bin configuration. For Gd:Ho (ΔE = 5.4 keV), noise in the Ho maps was 0.6 ± 0.1, 1.0 ± 0.2, and 0.6 ± 0.1 mg/mL for configurations A, B, and C, respectively. For Ho:W (ΔE = 13.9 keV), Ho-map noise was lower overall and configuration-dependent (0.4 ± 0.1, 0.2 ± 0.0, 0.5 ± 0.1 mg/mL for A, B, C).

Together, these results show that (i) clinically available dual-source PCCT can achieve dual-contrast decomposition for K-edge separations down to 5.4 keV, (ii) mean quantification error generally improves with increasing K-edge separation, and (iii) energy-bin selection is a key determinant of noise/precision across pairs and becomes especially consequential for closely spaced K-edges.

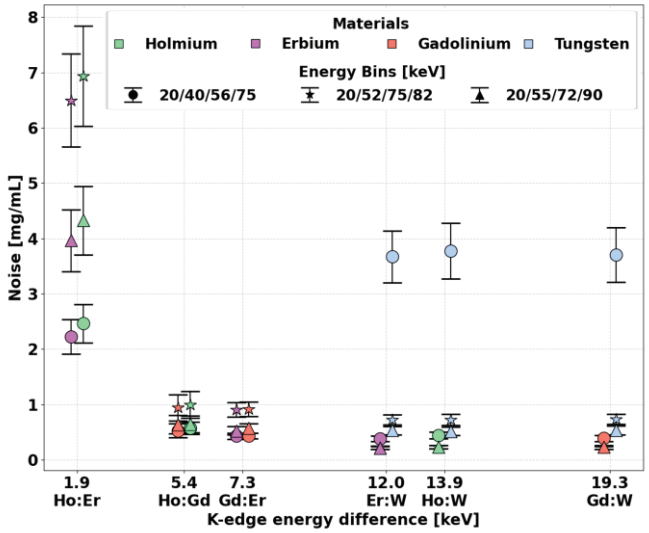


Figure 4. Noise on material-specific maps of contrast pairs with varying differences in K-edge energies. Contrast pairs with K-edge energy differences greater than 5 keV display lower noise as differences increase. Error bars represent the standard deviation of noise collected across all slices.

IV. Conclusion

This phantom study demonstrates that dual-contrast K-edge decomposition is feasible on a clinical dual-source PCCT system for multiple high-Z pairs and remains achievable for K-edge separations down to 5.4 keV (Gd:Ho) under a clinical dose protocol. Quantitatively, accuracy and visual conspicuity generally improved with increasing K-edge separation.

Energy-bin configuration impacted performance in two keyways. First, bin selection had the largest effect for closely spaced pairs (e.g., Gd:Ho), indicating that threshold placement is most critical when decomposition conditioning is poorest. Second, tungsten showed pronounced bin-dependent bias, suggesting that high-energy K-edge agents may require tailored bin choices even when K-edge separation is large. Noise decreased with K-edge separation but remained consistently sensitive to bin configuration, underscoring a practical bias–variance tradeoff when selecting thresholds.

Key limitations include the controlled phantom environment and reliance on a vendor decomposition workflow; future work should report cross-talk, linearity, and limits of detection to better support protocol selection and translation.

Clinically available dual-source PCCT can support single-scan dual-contrast K-edge imaging for high-Z pairs with K-edge separations as small as 5.4 keV, but quantitative performance depends on energy-bin selection, with material-specific sensitivity that should be considered when designing and standardizing clinical protocols.

References

[1] D. Muenzel *et al.*, "Simultaneous dual-contrast multi-phase liver imaging using spectral photon-counting computed tomography: a proof-of-concept study," *Eur. Radiol. Exp.*, vol. 1, p. 25, Dec. 2017, doi: 10.1186/s41747-017-0030-5.

[2] D. Muenzel *et al.*, "Spectral Photon-counting CT: Initial Experience with Dual–Contrast Agent K-Edge Colonography," *Radiology*, vol. 283, no. 3, pp. 723–728, Jun. 2017, doi: 10.1148/radiol.2016160890.

[3] S. Si-Mohamed *et al.*, "Spectral Photon-Counting Computed Tomography (SPCCT): in-vivo single-acquisition multi-phase liver imaging with a dual contrast agent protocol," *Sci. Rep.*, vol. 9, no. 1, p. 8458, Jun. 2019, doi: 10.1038/s41598-019-44821-z.

[4] J. Dangelmaier *et al.*, "Experimental feasibility of spectral photon-counting computed tomography with two contrast agents for the detection of endoleaks following endovascular aortic repair," *Eur. Radiol.*, vol. 28, no. 8, p. 3318, Feb. 2018, doi: 10.1007/s00330-017-5252-7.

[5] R. Symons *et al.*, "Photon-counting CT for simultaneous imaging of multiple contrast agents in the abdomen: An in vivo study," *Med. Phys.*, vol. 44, no. 10, pp. 5120–5127, Oct. 2017, doi: 10.1002/mp.12301.

[6] C. T. Badea, D. P. Clark, M. Holbrook, M. Srivastava, Y. Mowery, and K. B. Ghaghada, "Functional imaging of tumor vasculature using iodine and gadolinium-based nanoparticle contrast agents: a comparison of spectral micro-CT using energy integrating and photon counting detectors," *Phys. Med. Biol.*, vol. 64, no. 6, p. 065007, Mar. 2019, doi: 10.1088/1361-6560/ab03e2.

[7] C. H. McCollough, S. Leng, L. Yu, and J. G. Fletcher, "Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications," *Radiology*, vol. 276, no. 3, pp. 637–653, Sep. 2015, doi: 10.1148/radiol.2015142631.

[8] X. Liu, L. Yu, A. N. Primak, and C. H. McCollough, "Quantitative imaging of element composition and mass fraction using dual-energy CT: Three-material decomposition," *Med. Phys.*, vol. 36, no. 5, pp. 1602–1609, May 2009, doi: 10.1118/1.3097632.

[9] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical Principles and Clinical Prospects," *Radiology*, vol. 289, no. 2, pp. 293–312, Nov. 2018, doi: 10.1148/radiol.2018172656.

[10] E. Roessl and R. Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Phys. Med. Biol.*, vol. 52, no. 15, pp. 4679–4696, Aug. 2007, doi: 10.1088/0031-9155/52/15/020.

[11] J. P. Schlomka *et al.*, "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, no. 15, p. 4031, Jul. 2008, doi: 10.1088/0031-9155/53/15/002.

[12] J. C. Hsu *et al.*, "Nanomaterial-based contrast agents," *Nat. Rev. Methods Primer*, vol. 3, p. 30, 2023, doi: 10.1038/s43586-023-00211-4.

[13] J. Neumann, T. Nowak, B. Schmidt, and J. von Zanthier, "An Image-Based Prior Knowledge-Free Approach for a Multi-Material Decomposition in Photon-Counting Computed Tomography," *Diagnostics*, vol. 14, no. 12, Art. no. 12, Jan. 2024, doi: 10.3390/diagnostics14121262.

Enable K-Edge Material Decomposition Without K-Edge Material Calibration: A Feasibility Study

Yirong Yang, Wenying Wang

***Abstract*—Photon-counting computed tomography (PCCT) with energy discrimination capabilities enables quantitative material characterization. Compared to conventional dual-energy CT (DECT), PCCT with multiple energy bins can detect and quantify K-edge materials through three material decomposition (3MD). A common projection-domain approach measures energy bin counts using step-wedge phantoms of varying thicknesses during calibration. For K-edge imaging, such calibration requires slabs composed of the K-edge materials of interest, such as gadolinium, which are difficult to fabricate at the thin thicknesses required to achieve the desired attenuation.**

Image-domain MD avoids the need for K-edge calibration slabs but suffers from spectral distortions due to beam hardening, which can reduce quantitative accuracy for varying patient sizes. To address this limitation, we propose a projection-domain MD (PDMD) method that enables K-edge imaging without physical K-edge calibration slabs by estimating the effective energy bin spectra from calibration scans using only standard materials such as aluminum (Al) and PMMA. With the estimated spectra, both maximum likelihood estimation and simple forward models can perform 3MD with K-edge materials. This method does not require physical K-edge calibration slabs in the PDMD workflow.

We first tested the proposed method in simulation for a water-iodine-gadolinium MD task, where using one, two, or three calibration materials produced comparable MD accuracy and noise performance. We further evaluated this method on a prototype PCCT system with four fixed energy thresholds, using only Al and PMMA slabs for calibration. A GAMMEX phantom with iodine and gadolinium inserts was scanned. The proposed method successfully separated gadolinium from iodine and provided accurate gadolinium concentration estimates. These results demonstrate that PDMD with K-edge materials is feasible without physical K-edge calibration slabs.

***Index Term*—photon-counting CT, material decomposition, K-edge imaging, spectrum estimation**

I. Introduction

PHOTON counting CT with spectral imaging capability allows simultaneous multi-contrast imaging using K-edge materials such as gadolinium or bismuth, and has shown strong clinical potential [1]. To realize these benefits, accurate material decomposition (MD) is required when K-edge materials are presented. A common approach is image-domain MD (IDMD). This approach is straightforward, requires minimal calibration, and performs well for moderate patient sizes. However, there remain concerns of beam hardening effects for larger patient sizes, which can degrade spectral separation and reduce material quantification accuracy[1].

Projection-domain MD (PDMD) is preferred because it avoids these spectral distortion issues from beam-hardening effects. This approach directly maps energy bin counts to basis material thicknesses through a forward physical model. A straightforward workflow calibrates the system with a step-wedge phantom that contains all basis materials, including the K-edge materials of interest. With these energy bin measurements, the forward model can be built. However, this approach is not always practical. The high atomic number K-edge materials have high attenuation, so the calibration slabs must be thin and uniform, which are challenging to manufacture and work with. Moreover, multi-contrast-agent imaging requires a large number of material combinations with varying thicknesses for accurate spectral response calibration, further complicating the calibration process.

In many PCCT simulation studies, accurate MD results have been achieved by applying maximum likelihood estimation (MLE) when the correct effective energy bin spectra are provided [2]. This observation suggests that, if one can accurately estimate these spectra, PDMD with K-edge materials becomes feasible without physical K-edge material calibration slabs. The three material decomposition (3MD) problem then becomes a spectrum estimation problem under limited calibration measurements. Prior works have shown that spectrum estimation is ill-conditioned, and that including more calibration materials, such as K-edge filters, can improve numerical stability [2], [3]. In practice, however, PCCT systems with multiple energy bins provide multiple energy-resolved measurements, increasing the spectral information available for material decomposition tasks, and the additional benefit of using more calibration materials for MD tasks may be limited. This raises the question of whether simple calibration setups using only standard materials are sufficient to estimate the spectra with adequate accuracy for reliable 3MD.

In this work, we propose a projection-domain 3MD (PD3MD) method that estimates energy bin spectra using only standard calibration materials, without physical K-edge slabs. We evaluate its performance on 3MD tasks using both simulation and real data acquired from a prototype CZT-based PCCT system.

This work was supported by United Imaging Healthcare North America Inc.
Yirong Yang (yirong.yang@united-imaging.com) and Wenying Wang are with United Imaging Healthcare North America Inc.

II. Methods

A. Measurement Model and Spectra Estimation

The spectral measurement model for PCCT can be written as

$$\lambda_i = \int_{U_i}^{U_{i+1}} dE' \int_0^{\infty} dE \cdot S(E) \exp\left(-\sum_{m=1}^{M} \mu_m(E) A_m\right) R(E'; E), \quad (1)$$

where λ_i is the detected counts in the ith energy bin, $S(E)$ denotes the source spectrum. $A_m = \int_l t_m(r) dr$ stands for the line integral of the mth basis material, and $R(E'; E)$ is the spectral response of the PCD describing the probability that a photon of energy E is resolved at energy E'. One can rewrite equation (1) as:

$$\lambda_i = \int_0^{\infty} dE \cdot \big(R(i; E) S(E)\big) \exp\left(-\sum_{m=1}^{M} \mu_m(E) A_m\right), \quad (2)$$

where $R(i; E) = \int_{U_i}^{U_{i+1}} R(E'; E) dE'$, and $S_i(E) = R(i; E) S(E)$ represents the effective spectrum of the ith energy bin.

By discretizing equation (2), we obtain the linear system

$$y_i = T s_i, \quad (3)$$

where y_i is the discretized measurement vector of the ith energy bin measurements, s_i is the discrete effective spectrum of energy bin i, and $T[p, q] = \exp\left(-\sum_{m=1}^{M} \mu_m(q) \int_{L_p} t_m(r) dr\right)$.

Sidky *et al.* showed that T is highly ill-conditioned. Therefore, estimating spectrum s_i from transmission measurements y_i is an ill-conditioned problem. They proposed a robust method to solve this problem [2]:

$$s_i^{(k+1)}[q] = s_i^{(k)}[q] \frac{\sum_p T^T[q, p] \dfrac{y_i[p]}{\sum_{q'} T[p, q'] s_i^{(k)}[q']}}{\sum_p T^T[q, p]}. \quad (4)$$

We use these estimated effective spectra $\hat{s}_i$ to simulate energy bin measurements with additional K-edge filters and fit simple feed-forward models for material decomposition. We can also directly use these spectra in a maximum likelihood estimator for MD. The overall calibration and correction workflow is summarized in Figure 1.

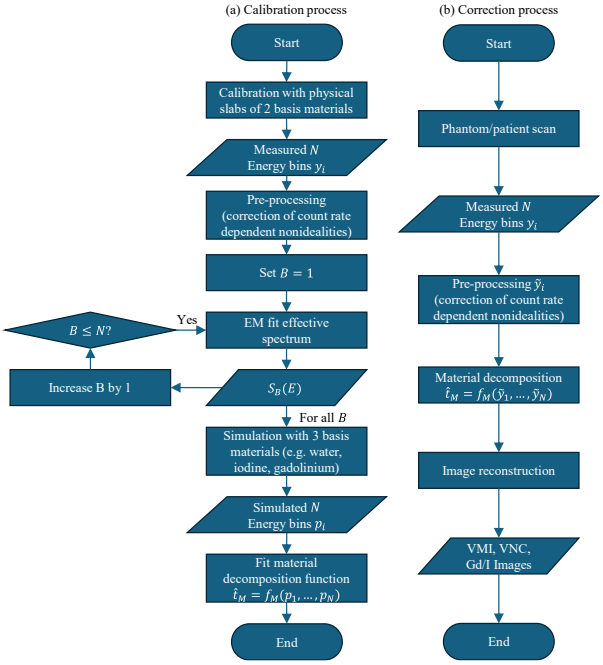


Figure 1. Flowchart of PD3MD method with K-edge materials.

B. Calibration Materials and Effective Rank Analysis

Both Sidky *et al.* and Xu *et al.* showed that including additional calibration filters, especially K-edge filters, increases the rank of the transmission matrix T and improves the numerical stability of this optimization problem [2], [3]. We used three material sets: {Water [0, 320] mm}; {Water [0, 320] mm, Iodine [0,45] mg/cm^2}; {Water [0, 320] mm, Iodine [0,45] mg/cm^2, Gadolinium [0, 30] mg/cm^2} to compare the singular values of T across different calibration material combinations. For each set, we uniformly sampled 64 thickness combinations within the spaces spanned by these ranges and evaluated the transmission model at 46 energy points between 30 and 120 keV. The normalized singular values are shown in Figure 2.

We observed increased singular value with additional materials, however, the singular values decay quickly for all material sets. In these cases, the standard matrix rank does not reflect the true dimensionality of transmission matrix T. Roy *et al.* introduced the effective rank of matrix T, defined as

$$\text{erank}(T) = \exp\left(-\sum_{k=1}^{Q} p_k \log p_k\right), \quad (5)$$

where $p_k = \frac{\sigma_k}{\|\sigma\|_1}$, $\sigma = [\sigma_1, \ldots, \sigma_Q]^T$ are the nonzero singular values of matrix T with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_Q > 0$ [4]. The effective rank measures the effective spread of singular values using spectral entropy, and $1 \leq \text{erank}(T) \leq Q$. A small effective rank indicates that the transformation defined by T preserves only a few dominant directions.

We calculated the effective ranks of T for the three different calibration material sets: 1.4085, 1.6413, and 1.7337 for one, two, and three calibration materials (

Table *1*). The effective rank increases with the inclusion of more materials, consistent with observations in [2], [3]. However, the increment is small, suggesting that including additional calibration materials may not significantly improve spectrum estimation accuracy.

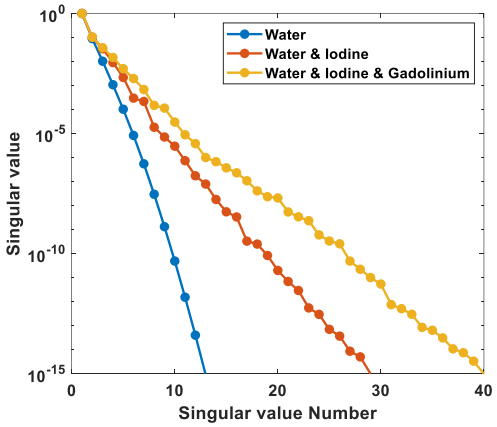


Figure 2. Normalized singular values of transmission matrix T for 64 thickness combinations of material spaces spanned by [0, 320] mm water (blue), [0, 320] mm water + [0, 45] mg/cm^2 iodine (red), and [0, 320] mm water + [0, 45] mg/cm^2 iodine + [0, 30] mg/cm^2 gadolinium (yellow).

C. Illustration with Simulation Study

We applied the PD3MD pipeline described in Figure 1 on simulated data using a realistic CdTe-based PCD energy response with energy thresholds at 35, 50, 65 keV [5]. In this

study, we used MLE as the MD method. MD accuracy was evaluated over a test space spanned by {Water [0, 240] mm, Iodine [0,45] mg/cm^2, Gadolinium [0, 30] mg/cm^2}. For each calibration material set, we computed the RMSE and relative variance of the estimated material thickness and compared the results with those obtained from the material set with the largest number of materials. The results are summarized in

Table *1*.

We observed that the RMSE and relative variance of the estimated thickness remained comparable across all material sets. This small difference in accuracy and noise performance suggests that the benefit of additional calibration materials is limited for 3MD tasks. In practice, a simple calibration setup without physical K-edge materials may be sufficient to achieve reliable MD results.

D. Experimental Study on A PCCT Prototype

With the above conclusions, we then validated and evaluated the proposed method on a prototype CZT PCCT system using a tube voltage of 120 kVp and fixed energy thresholds at 30, 55, 65, 85 keV. This setting was used throughout the experiment. The calibration scan was performed using Al and PMMA slabs of known thicknesses that cover a range of patient sizes. We formed three-energy-bin configurations with threshold sets [30, 55, 65] keV and [30, 55, 85] keV using these measurements.

After preprocessing to correct count rate dependent nonidealities, including pulse pileup effect, we estimated the effective spectrum for each energy bin. These spectra were then used to simulate energy bin measurements for x-ray beams attenuated by water, iodine, and gadolinium filters with various thickness combinations. A second-order polynomial function $\hat{t}_m = f_m(p_1, p_2, p_3)$ was fit to map the simulated bin counts p_i to the estimated basis material thickness t_m.

A GAMMEX head phantom containing iodine inserts of 2, 5, 10 mg/ml, gadolinium inserts of 5, 10 mg/ml, and syringes with gadolinium solutions of 1, 2, 4, 8 mg/ml was scanned using the same multi-bin protocol at 200 mA. The phantom layout and materials/concentrations of inserts are shown in Figure 3.

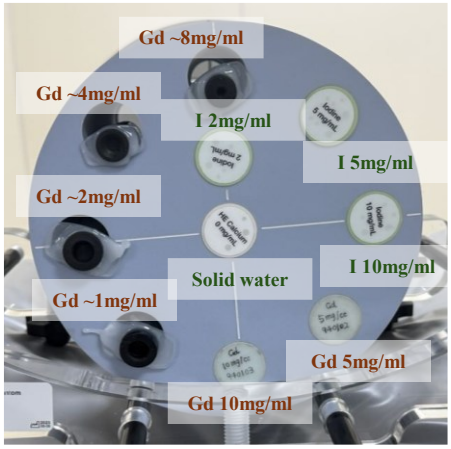


Figure 3. Layout and concentration of iodine and gadolinium samples in the GAMMEX head phantom.

The projection data was acquired with the four preset energy thresholds and then recombined into three-energy-bin datasets using the following two threshold sets: [30, 55, 65] keV, and [30, 55, 85] keV. Images reconstructed from these three-bin configurations were also linearly combined with weights chosen to minimize noise variance in the gadolinium regions [6]. This approach provided an estimate of the performance that is achievable with the original four-energy-bin measurements.

III. RESULTS

The reconstructed basis material maps from the three-energy-bin and four-energy-bin configurations are shown in Figure 4. All images were reconstructed using FBP without denoising or postprocessing. The results show that the proposed method successfully separates gadolinium from iodine and performs consistently across different threshold settings. The four-energy-bin configuration (combined threshold, Figure 4 (c)(f)) exhibits visually lower noise in both iodine and gadolinium maps, highlighting the benefit of using an additional energy bin for 3MD tasks.

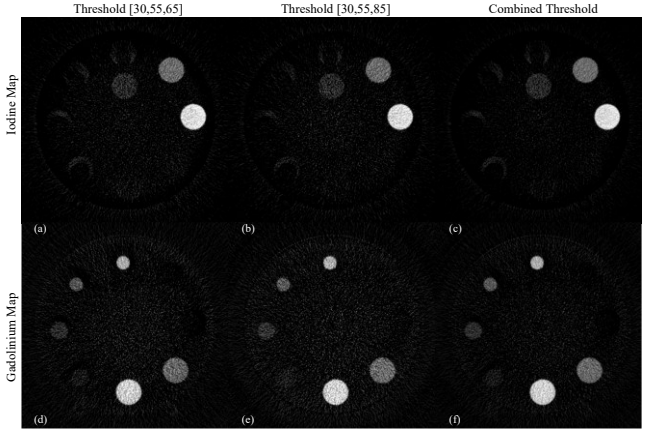


Figure 4. Iodine (a-c) and gadolinium (d-f) basis material maps using three-energy-bin with thresholds [30,55,65] keV (a,d) and [30,55,85] keV (b,e) and using four-energy-bin with thresholds [30,55,65,85] keV (c,f). The display window is [0, 12]mg/ml.

Virtual monoenergetic images (VMIs) at 40, 51 (Gd K-edge), 69 keV were generated (Figure 5). The four-energy-bin configuration yields improved MD performance with lower noise and fewer streaks. As expected, the HU values in gadolinium regions increase at 51 keV due to the K-edge.

To quantify the improvement, we compared the CNR of iodine and gadolinium ROIs between the two configurations. Across all VM energies, using additional energy bin increases CNR by more than 20% (Figure 6).

We also evaluated the quantitative accuracy of the estimated iodine and gadolinium concentrations. Table 2 shows that both materials have small bias relative to the ground truth. The concentrations of gadolinium solutions are slightly overestimated from the nominal values, likely due to uncertainties in manual solution preparation and syringe filling.

Table 1. Effective rank, RMSE, and relative variance of three material sets using one, two, and three calibration materials.

Calibration Materials	ERank	RMSE			Relative Variance					
		H_2O (mm)	I(mg/cm^2)	Gd (mg/cm^2)	H_2O	I	Gd	VM@50 keV	VM@65 keV	VM@90keV
H_2O	1.4085	0.0340	0.7276	0.4838	1.0013	1.0247	1.0073	1.0004	1.0111	1.0018
H_2O+I	1.6413	0.0304	0.6893	0.4635	0.9985	1.0266	1.0008	0.9963	1.0082	1.0017
H_2O+I+Gd	1.7337	0.0308	0.4573	0.2942	1	1	1	1	1	1

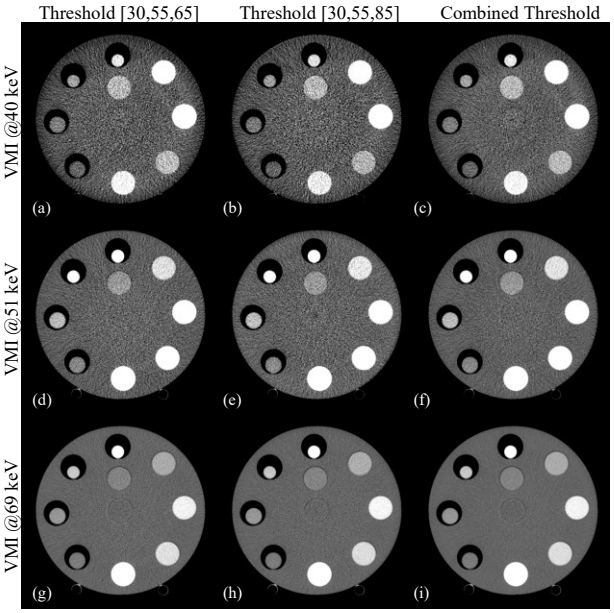


Figure 5. Virtual monoenergetic images at 40 keV (a-c), 51 keV (d-f), and 69 keV (g-i) using three-energy-bin with thresholds [30,55,65] keV (a,d,g) and [30,55,85] keV (b,e,h) and using four-energy-bin with thresholds [30,55,65,85] keV (c,f,i). The display window is [-200, 300] HU.

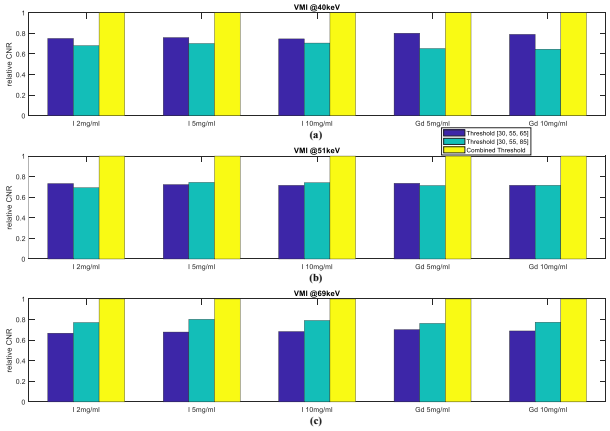


Figure 6. Relative CNR of iodine and gadolinium insert ROI compared to the four-energy-bin configuration.

Table 2. Measured iodine and gadolinium concentrations of ROIs on the GAMMEX head phantom.

	Iodine (mg/ml)			Gadolinium (mg/ml)					
	2	5	10	~1	~2	~4	~8	5	10
[30,55,65]	1.77	4.98	10.37	1.08	2.19	4.19	8.29	4.97	9.92
[30,55,85]	1.75	4.90	10.20	1.15	2.24	4.18	8.26	5.05	9.89
Combined	1.76	4.95	10.29	1.12	2.21	4.18	8.28	5.00	9.91

IV. Discussion

We proposed and evaluated a PD3MD method that enables K-edge material decomposition without the use of physical K-edge material for calibration. This workflow is practical for K-edge imaging with PCCT systems as it removes the need for thin and uniform K-edge slabs, which are difficult to manufacture and handle. Both simulation and experimental studies using a prototype CZT-based PCCT system demonstrate the feasibility of this approach.

The simulation results show that adding additional calibration materials increases the effective rank of the transmission matrix. However, the increment is limited. Both RMSE and variance of the 3MD outputs remain comparable across all calibration material sets. These results suggest that the spectra estimation problem remains highly ill-conditioned even with additional calibration materials, and that the benefit of additional K-edge filters in the calibration process may be limited in practice. Instead, two standard non-K-edge materials, such as Al and PMMA, may already provide adequate information to estimate the energy bin spectra with the accuracy required for 3MD tasks including K-edge contrast agents.

The experimental results from the GAMMEX phantom scans on the prototype PCCT system further validated these findings. The spectra estimated using only Al and PMMA step-wedge phantoms enabled successful separation of iodine and gadolinium with accurate quantification of their concentrations. The consistent performance across different threshold configurations also suggests that the proposed method is robust to various acquisition protocols. In addition, the results showed significantly improved CNR when using an additional energy bin in 3MD tasks, highlighting the value of increased spectral information from more energy bins.

Future work will focus on the following aspects. First, systematic guidelines are needed for calibration materials selection and phantom design that balance calibration phantom complexity, calibration workflow efficiency, and MD performance. We also plan to test this method using anthropomorphic phantoms with iodine, gadolinium, and their mixtures. Finally, we plan to extend this approach to multiple K-edge material decomposition tasks, such as bismuth or gold, to evaluate its generalizability to diverse spectral imaging applications.

V. Conclusion

We proposed a PD3MD method that estimates energy bin spectra using standard step-wedge calibration phantoms without the need for physical K-edge material slabs. Both

simulation and experimental results from a prototype PCCT system show accurate separation and quantification of iodine and gadolinium, with consistent performance across different energy threshold settings.

REFERENCES

[1] L. Ren *et al.*, "Performance evaluation of single- and dual-contrast spectral imaging on a photon-counting-detector CT," *Medical Physics*, vol. 51, no. 11, pp. 8034–8046, 2024, doi: 10.1002/mp.17367.

[2] E. Y. Sidky, Y. Lifeng, P. Xiaochuan, Z. Yu, and V. Michael, "A robust method of x-ray source spectrum estimation from transmission measurements: Demonstrated on computer simulated, scatter-free transmission data," *Journal of Applied Physics*, vol. 97, no. 12, June 2005, doi: 10.1063/1.1928312.

[3] X. Xu, L. Zhang, and S. Wang, "Response Function Estimation for the Photon Counting Detector using Multiple Balanced KEdge Filters," in *2017 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, Oct. 2017, pp. 1–4. doi: 10.1109/NSSMIC.2017.8532857.

[4] O. Roy and M. Vetterli, "The effective rank: A measure of effective dimensionality," in *2007 15th European Signal Processing Conference*, Sept. 2007, pp. 606–610. Accessed: Dec. 04, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/7098875

[5] M. Persson, A. Wang, and N. J. Pelc, "Detective quantum efficiency of photon-counting CdTe and Si detectors for computed tomography: a simulation study," *JMI*, vol. 7, no. 4, p. 043501, July 2020, doi: 10.1117/1.JMI.7.4.043501.

[6] Y. Yang, X. Zhang, and W. Wang, "Impact of energy threshold placement on spectral imaging performance in multi-bin photon counting CT," in *Medical Imaging 2026: Physics of Medical Imaging*, SPIE, Feb. 2026.

A Digital Twin of Photon-Counting CT System

Ruiran Lai, Ran Zhang, Linying Zhan, Xiaopeng Yu, Roy Nilsen, Tyler Curtis-Nguyen, Brian Yanoff, Bruno De Man, Ke Li, Guang-Hong Chen

***Abstract*—A digital twin of a CT system requires physics-grounded and experimentally validated representations of its constituent subsystems, particularly the detector. While x-ray source, object attenuation, and reconstruction models for photon-counting CT (PCD-CT) are largely mature, the lack of a validated photon-counting detector digital twin has limited existing system-level simulation platforms from qualifying as true digital twins of real scanner systems.**

In this work, we present a system-level PCD-CT digital-twin framework anchored by a physics-based and experimentally validated detector digital twin. The detector model captures x-ray interaction, space-charge-dependent charge transport, signal induction, pulse shaping, and multi-threshold photon counting. Validation using CdTe and CZT detectors demonstrates quantitative agreement in multi-energy-bin counting statistics across multiple tube-current settings, with a maximum bin-fraction error of 3.2%. This validated detector digital twin provides the critical missing component for completing a system-level digital twin of PCD-CT.

I. Introduction

Digital twins have become essential tools in modern large-scale engineered systems, including spacecraft, launch vehicles, and planetary exploration platforms. In these applications, a digital twin is not merely a collection of simulations, but a physics-grounded digital representation that is anchored by measurements and validated against the corresponding physical subsystems. Digital twins enable ground teams to diagnose anomalies, test mechanistic hypotheses, and evaluate corrective actions under operational constraints where direct access to hardware is impossible.

A credible digital twin requires physically faithful subsystem models together with experimental or operational validation within a defined operating envelope. The defining distinction between a true digital twin and a conventional simulator lies in the degree to which the model is experimentally anchored and predictively reliable for its intended tasks.

Computed tomography (CT) imaging systems share many characteristics with other complex engineered platforms, including tightly coupled subsystems, nonlinear physics, and strong interactions between hardware design, data acquisition, and computational processing. A system-level CT digital twin therefore has the potential to accelerate research and development by enabling rapid design exploration, systematic evaluation of performance trade-offs, and efficient investigation of mitigation strategies. In addition, CT digital twins can support realistic data generation for algorithm development and machine learning, as well as closed-loop optimization of acquisition and reconstruction strategies for task-specific imaging applications.

Recent advances in computational power have made comprehensive CT digital twins increasingly feasible. Over decades of CT development, several system components—including x-ray source and geometry modeling, object attenuation, and reconstruction pipelines—have reached a high level of maturity, with many validated software tools and public implementations available [1]. Building on these components, multiple simulation frameworks have been developed to support virtual imaging trials and system evaluation for PCD-CT. For example, DukeSim and related platforms [2], [3], [4] generate scanner-specific PCD-CT images by modeling source and scanner geometry and incorporating user-supplied detector response models, while more recent efforts integrate Geant4-based transport and analytical detector-response mechanisms within CT simulation workflows [5].

Despite these advances, a physics-grounded and experimentally validated digital twin of the photon-counting detector itself remains largely absent. In existing frameworks, detector behavior is typically represented by simplified or empirical response models rather than a validated digital twin that explicitly captures charge transport, space-charge effects, signal induction, and electronic processing. This limitation is consequential: without a validated detector digital twin, even highly sophisticated system-level PCD-CT simulation platforms do not qualify as digital twins of the real scanner system, as similar system-level performance can arise from different combinations of component-level models that do not faithfully represent the underlying detector physics.

In this work, we address this critical missing component by developing and experimentally validating a physics-based digital twin of pixelated photon-counting detectors. The proposed detector digital twin captures the full response chain from x-ray interaction and charge transport, including space-charge effects, to signal induction, pulse shaping, and multi-threshold photon counting. By establishing detector-level predictive fidelity, this work provides the enabling foundation required to complete a system-level digital twin of a PCD-CT imaging system through integration with existing, mature source, attenuation, and reconstruction modules.

II. Materials and Methods

We develop a physics-based digital twin framework for a PCD-CT system that integrates x-ray source and object attenuation modeling, a validated photon-counting detector digital twin, and sinogram generation and reconstruction. Figure 1 summarizes the system-level workflow and the internal

R. Lai, L. Zhan, R. Zhang, X. Yu, and G.-H. Chen are with the Department of Medical Physics, University of Wisconsin–Madison, USA.

R. Nilsen, B. Yanoff, T. Curtis-Nguyen, and B. De Man are with GE Healthcare.

K. Li is with The University of Texas MD Anderson Cancer Center, USA.

structure of the detector digital twin, which constitutes the core component of the framework.

A. X-ray Source and Object Attenuation Modeling

X-ray source and object attenuation modeling provide the spectral and fluence input to the photon-counting detector digital twin. Given the incident x-ray spectrum (fluence per unit energy) $\Omega_0(E)$ at the object entrance, the transmitted spectrum after object attenuation is computed using the Beer–Lambert law:

$$\Omega(E) = \Omega_0(E) \exp\left(- \int_L \mu(E, \vec{x}) \, \mathrm{d}l \right), \tag{1}$$

where $\mu(E, \vec{x})$ is the energy- and position-dependent linear attenuation coefficient, and the line integral is evaluated along the ray path L from the x-ray source to the detector pixel. Energy-dependent attenuation coefficients are obtained from the NIST XCOM database and combined according to the object material map to form $\mu(E, \vec{x})$. The resulting spectrum $\Omega(E)$ is used as input to the photon-counting detector model to generate multi-threshold (energy-bin) measurements.

B. Photon-Counting Detector Model

The photon-counting detector digital twin constitutes the core component of the proposed framework. It models the complete chain of physical and electronic processes that transform incident x-ray photons into energy-binned photon counts. The detector model is organized into three tightly coupled modules: pixelated electrodynamics, signal train formation, and counting statistics with energy binning.

1) Module 1: Pixelated Electrodynamics: This module models charge transport and signal induction in pixelated detector elements with lateral dimensions (W_x, W_y) and thickness d. Two electrodynamic boundary-value problems are solved. First, the weighting field $\vec{E}_w(x, y, z)$ is computed, which enters the Shockley–Ramo theorem for determining induced current and charge on the readout electrodes. Second, the physical electric field $\vec{E}(x, y, z)$ inside the detector is obtained by solving the Poisson equation

$$\nabla \cdot (\varepsilon \nabla \phi) = -\rho_{h/e}(x, y, z) - q \, \delta(\vec{r} - \vec{r}_0), \tag{2}$$

subject to a bias voltage V_0 applied between the anode and cathode, where $\rho_{h/e}(x, y, z)$ denotes the background space-charge distribution and $\vec{r}_0$ is the carrier location.

Using the resulting electric field, the quasi-classical equation of motion is solved to obtain the carrier velocity $\vec{v}_{h/e}(x, y, z; t)$ under the Lorentz force. Together with the weighting field, the Shockley–Ramo theorem is applied to compute the induced current $I(t)$ and induced charge $Q_{\mathrm{ind}}(t)$. This module captures the coupled effects of detector geometry, material properties, space charge, and bias conditions on charge transport and signal induction.

Key parameters include detector pitch, dielectric constant ε, carrier mobility μ, mean electron–hole pair creation energy W_{e-h}, space-charge profile, bias voltage V_0, photon energy, and interaction location.

2) Module 2: Signal Train Formation: This module models the stochastic conversion of an incident x-ray photon flux into a time-resolved electrical pulse train. It consists of two coupled sub-modules: detector energy response modeling and electronic pulse shaping.

Sub-module 2.1: Detector Energy Response

Incident photon arrivals are modeled as a Poisson process with rate Ψ, yielding an exponential inter-arrival time distribution $P(t) \sim e^{-\Psi t}$. For an exposure time T, the total number of photons n is sampled from a Poisson distribution with mean $\bar{n} = \Psi T$, and individual arrival times are sampled conditionally. Photon energies E are independently drawn from the post-object spectrum $\Omega(E)$.

For each photon, the deposited energy E' is sampled from a detector-specific energy response function $p(E'|E)$, accounting for stochastic energy deposition mechanisms such as Compton scattering and charge sharing.

Sub-module 2.2: Pulse Shaping and Signal Generation

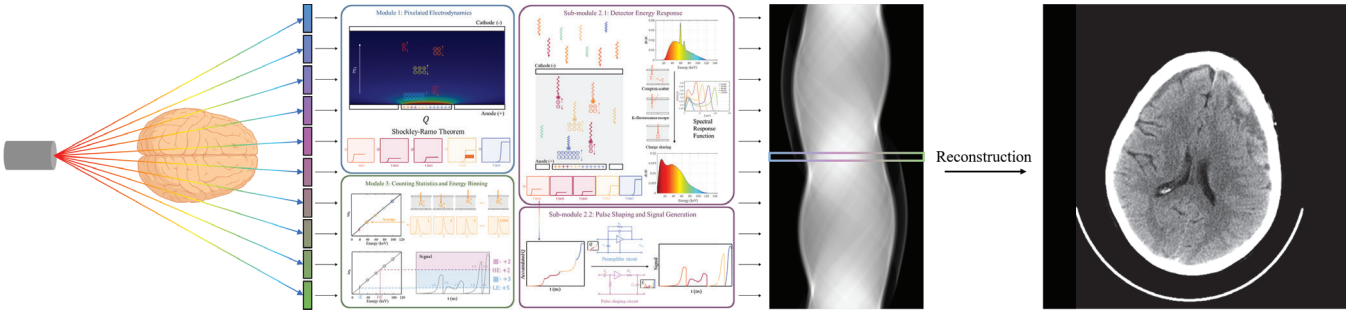
For each deposited energy E' and arrival time t, the induced charge signal from Module 1 is processed through an electronic pulse-shaping circuit model. CRRC and CRRC^4 shaping circuits are considered. The shaping capacitance is treated as the primary adjustable parameter controlling pulse width, while a finite reset time models non-paralyzable detector operation. Each photon thus produces a shaped electrical pulse with a defined amplitude and temporal profile. Repeating this process for all n photons yields the complete electrical pulse train. Pileup naturally emerges from this module.

3) Module 3: Counting Statistics and Energy Binning: This module converts the electrical pulse train into photon counts in discrete energy bins. A Monte Carlo–based calibration is performed to map pulse amplitude to deposited energy. For each energy E', 1000 photon interaction locations are sampled uniformly in the (x, y) plane and exponentially along z according to the Beer–Lambert law. For each location, Modules 1 and 2 generate an electrical pulse, and the peak pulse amplitude is recorded.

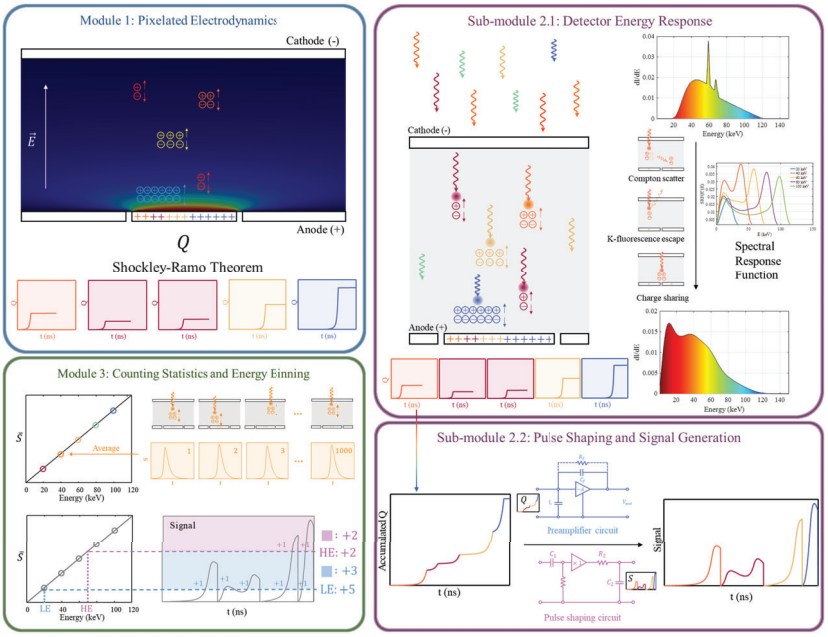
The mean pulse amplitude is computed, and 95% of this value is selected as the representative signal amplitude for energy E'. This procedure is repeated for $E' = 10, 20, 30, \ldots, 140$ keV to obtain a monotonic calibration curve $\bar{S}(E')$. Energy thresholds are mapped to corresponding signal amplitude thresholds, and pulses exceeding a threshold contribute one count. Counts in individual energy bins are obtained from differences between adjacent threshold counts.

C. Sinogram Generation and Reconstruction

The x-ray source and object attenuation model and the photon-counting detector digital twin are evaluated for each projection view and detector pixel. The resulting multi-energy-bin measurements form a set of energy-resolved sinograms indexed by projection angle, detector channel, and energy bin. These sinograms serve as inputs to subsequent reconstruction algorithms, enabling integration of the validated detector digital twin into a system-level PCD-CT simulation workflow.



(a) End-to-end photon-counting CT (PCD-CT) digital-twin workflow. X-ray source and object attenuation modeling determine the incident spectrum and fluence at the detector surface; the detector digital twin generates energy-resolved measurements; energy-resolved sinograms are formed and reconstructed to produce images.



(b) Structure of the photon-counting detector digital twin, illustrating the three coupled modules: pixelated electrodynamics, signal train formation, and counting statistics with energy binning.

Fig. 1: Schematic overview of the proposed photon-counting CT system digital twin and its detector-level core.

D. Experimental Validation Studies

To assess the predictive fidelity of the proposed photon-counting detector digital twin, experimental validation studies were conducted using a CZT-based PCD (Redlen Technologies, Vancouver, Canada). The detector thickness was 2.0 mm, with a pixel pitch of $0.3425\ \text{mm} \times 0.3687\ \text{mm}$. X-rays were generated using a tube at 120 kV potential. Experimentally measured energy spectra, count-rate behavior, and energy-bin resolved counting statistics were compared against corresponding digital twin predictions using matched detector geometry, material properties, bias conditions, and x-ray spectra.

III. Results

The results are organized around two representative use cases that reflect the role of the proposed detector digital twin within a system-level PCD-CT digital twin framework. The first use case establishes detector-level predictive fidelity through experimental validation of energy-resolved counting statistics. The second use case demonstrates how a validated detector digital twin enables task-relevant system design operations that are difficult to perform experimentally.

A. Use Case I: Energy-Bin Count Distributions

We first evaluate the ability of the detector digital twin to predict energy-resolved photon-counting statistics across six energy bins in a CZT-based detector. This comparison probes whether the model captures the coupled effects of photon interaction physics, charge transport, signal formation, and threshold-based counting under varying photon flux conditions.

Figure 2 compares simulated and experimentally measured energy-bin count distributions across multiple tube current settings. The digital twin reproduces both the overall spectral shape and the relative distribution of counts among energy bins over the investigated operating range. The maximum bin-wise

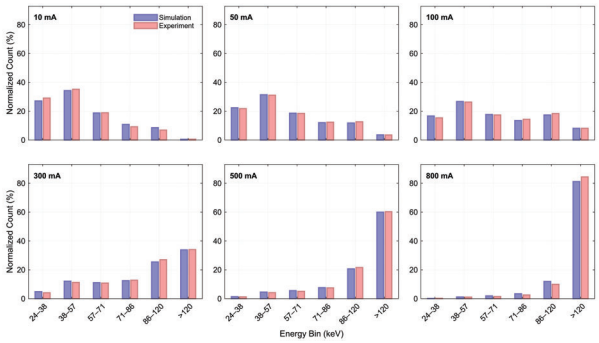


Fig. 2: Quantitative comparison between simulated and experimental energy-bin count distributions across multiple tube current levels for a CZT photon-counting detector (Redlen Technologies, Vancouver, Canada).

absolute error is 3.2%, with most energy bins exhibiting errors below 2%.

This agreement indicates that the detector digital twin captures the dominant physical mechanisms governing energy-dependent detector response and counting behavior. Importantly, the agreement is maintained across changes in tube current, demonstrating predictive performance under varying photon flux conditions rather than tuning to a single operating point. These results establish the detector digital twin as a quantitatively validated component suitable for integration into a system-level PCD-CT digital twin.

B. Use Case II: Energy Threshold Determination

We next demonstrate the use of the validated detector digital twin for determining energy thresholds in an equal-count, multi-bin photon-counting readout. This task is representative of system design and protocol optimization problems that benefit from a predictive detector model, as threshold selection depends on detector response, spectral shape, and object-dependent beam hardening.

For a given acquisition condition, the detector digital twin predicts the detected energy distribution $s(E)$ over $[E_{\min}, E_{\max}]$. Energy thresholds are selected such that each energy bin contains approximately the same expected number of counts. The bin edges $\{T_k\}_{k=0}^{5}$, with $T_0 = E_{\min}$ and $T_5 = E_{\max}$, are determined by

$$\int_{T_{k-1}}^{T_k} s(E)\,\mathrm{d}E = \frac{1}{5}\int_{E_{\min}}^{E_{\max}} s(E)\,\mathrm{d}E, \qquad k = 1,\ldots,5. \quad (3)$$

Figure 3 shows an example for a 120 kV spectrum. The procedure is repeated for spectra transmitted through 20 cm and 40 cm of water to account for beam-hardening effects, yielding the threshold sets summarized in Table I. The systematic upward shift of thresholds with increasing attenuation reflects the expected spectral hardening and demonstrates the utility of the detector digital twin for object-dependent system configuration.

IV. Conclusion

In this work, we developed and experimentally validated a physics-based digital twin of pixelated photon-counting detectors that captures the full detector response chain from x-ray interaction and charge transport to signal induction, pulse shaping, and multi-threshold, energy-binned photon counting. Quantitative agreement with experimentally measured six-bin counting statistics in a CZT detector establishes the proposed model as a predictive and reliable detector-level digital twin. As a representative system-design application, we demonstrated its use for equal-count energy-threshold determination under different spectral and beam-hardening conditions.

From a system-level perspective, this validated detector digital twin addresses the principal missing and most nonlinear component required to complete a digital twin of a photon-counting CT (PCD-CT) system. While x-ray source modeling, object attenuation, and reconstruction pipelines are largely mature, detector physics governs energy discrimination and counting behavior unique to PCD-CT. Integration of this validated detector core with existing system components enables a comprehensive system-level digital twin for photon-counting CT.

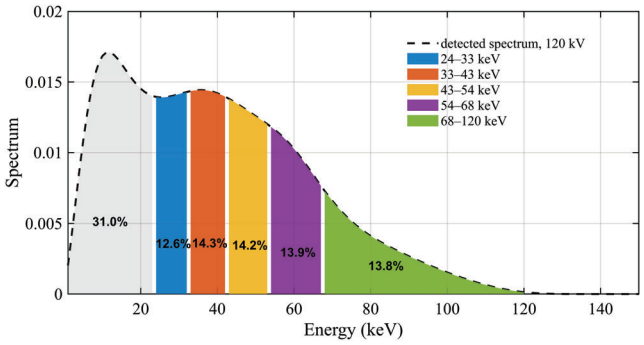


Fig. 3: Example of equal-count energy-bin threshold determination for a 120 kV spectrum over $[E_{\min}, E_{\max}]$.

TABLE I: Equal-count energy-bin thresholds (keV) for three beam-hardening conditions. The lower and upper bounds are fixed at 24 and 120 keV.

Condition	T_0	T_1	T_2	T_3	T_4	T_5
Unfiltered	24	33	43	54	68	120
20 cm water	24	37	50	62	78	120
40 cm water	24	40	56	69	85	120

References

[1] M. Wu, P. FitzGerald, J. Zhang, W. P. Segars, H. Yu, Y. Xu, and B. De Man, "XCIST—an open access x-ray/CT simulation toolkit," *Physics in Medicine & Biology*, vol. 67, no. 19, p. 194002, 2022.

[2] E. Abadi, B. Harrawood, J. R. Rajagopal, S. Sharma, A. Kapadia, W. P. Segars, K. Stierstorfer, M. Sedlmair, E. Jones, and E. Samei, "Development of a scanner-specific simulation framework for photon-counting computed tomography," *Biomedical Physics & Engineering Express*, vol. 5, no. 5, p. 055008, Aug. 2019. [Online]. Available: https://iopscience.iop.org/article/10.1088/2057-1976/ab37e9

[3] A. Kavuri, M. Fryling, N. Felice, L. Malvin, D. P. Clark, E. Samei, and E. Abadi, "An end-to-end CT simulation framework with graphical user interface and sample scanner models," *Medical Physics*, vol. 52, no. 10, p. e70066, Oct. 2025. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.70066

[4] S. Sharma, S. Vrbaški, M. Bhattarai, E. Abadi, R. Longo, and E. Samei, "A framework to model charge sharing and pulse pileup for virtual imaging trials of photon-counting CT," *Physics in Medicine &*

Biology, vol. 69, no. 22, p. 225001, Nov. 2024. [Online]. Available: https://iopscience.iop.org/article/10.1088/1361-6560/ad8b0a

[5] E. Fredenberg, D. Collin, L. Carbonne, M. Wu, B. D. Man, and F. Grönberg, "Simulating and correcting the pileup effect in deep-silicon photon-counting CT," *Medical Physics*, vol. 52, no. 8, p. e18075, Aug. 2025. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/10.1002/mp.18075

Combining Poisson random variables to generate and analyze the statistics of photon counting detectors

Karl Stierstorfer, Siemens Healthineers, Forchheim, Germany

***Abstract*—Photon counting detectors with several thresholds or bins show a non-trivial statistic that substantially deviates from Poisson. Reason for this are effects like double counting of photons due to charge sharing or K fluorescence. This makes generating random data following this statistic a non-trivial task.**

The paper shows that a linear combination of Poisson random variables with small integer factors can be used to tackle this problem efficiently.

The paper also demonstrates that the corresponding decomposition of the statistic into independent Poisson random variables reveals information about the underlying detector physics.

Results indicate that current simulation models of CdTe photon counting detectors need improving.

***Index Terms*—Photon counting detector, Poisson random variables, Poisson excess, Monte Carlo simulation, detector statistics**

I. Introduction

PHOTON COUNTING DETECTORS show an interesting statistic that substantially deviates from Poisson. Naturally, since all photons counted in a higher threshold will also be seen in a lower threshold, different thresholds are always strongly correlated but, due to effects like charge sharing, Compton scatter or K-fluorescence, different bins may also be correlated, and some photons are counted twice or even more times in neighboring pixels. Quantifying these effects is relevant since e.g. double counting substantially affects the DQE (detective quantum efficiency) of the detector. Correlations between bins or thresholds will also affect the spectral performance of photon counting detectors [2].

For Monte Carlo simulations, we want to create noise realizations a random variable x_j describing photon counting detector data following a given statistic

$$\mathrm{E}(x_i) = \mu_i, \qquad \mathrm{cov}(x_i, x_j) = \sigma_{ij}, \tag{1}$$

Where E symbolizes the expectation value. As predicted e.g. by detector simulations [1]. The index *i* is the number of the spectral channel (energy threshold or energy bin). For the reasons mentioned above, this is non-trivial for photon counting data, and it seems that there is no straightforward recipe to create such data.

Solving this problem leads to an interesting decomposition of the statistics into independent processes that allows some insight into the inner workings of the detector.

The author is with Siemens Healthineers AG, Computed Tomography, Siemensstr. 3, 91301 Forchheim, Germany, karl.stierstorfer@siemens-healthineers.com

II. Main idea

Count values from a photon counting detector are positive numbers. Hence, when generating such random data following a given statistic, i.e. certain mean values and covariances, we would like to create *n* random variables *x* as a linear combination of $m > n$ Poisson random variables *y*:

$$x_i = \sum_{k=1}^{m} M_{ik} y_k\,, \quad i = 1 \ldots n, \tag{2}$$

where M_{ik} is an arbitrary $m \times n$ matrix (typically $m \geq n$) and the y_k are *m* independent Poisson variables:

$$\mathrm{E}(y_k) = \lambda_k, \quad \mathrm{cov}(y_k, y_l) = \delta_{kl}\lambda_k, \quad k, l = 1 \ldots m \tag{3}$$

Then we have $n(n+3)/2$ *linear* equations connecting the statistics of the *x* and *y* random variables:

$$\mathrm{E}(x_i) = \mu_i = \sum_{k=1}^{m} M_{ik}\lambda_k\,, \qquad i = 1 \ldots n, \tag{4}$$

$$\mathrm{cov}(x_i, x_j) = \sigma_{ij} = \sum_{k,l=1}^{m} M_{ik}\mathrm{cov}(y_k, y_l)M_{jl} = \sum_{k=1}^{m} M_{ik}M_{jk}\lambda_k\,, \qquad 1 < i \leq j < n\,. \tag{5}$$

This ansatz, however, creates the problem that we end up with $m(n+1)$ unknown quantities: the *m* mean values of the *y* variables and the $m \cdot n$ entries of the *M* matrix. The requirement to reproduce the mean values and the covariances, however, yields only $n + n(n+1)/2 = n(n+3)/2$ independent constraints. For example, if we wanted to create $n = 2$ correlated random variables from $m = 3$ uncorrelated Poisson variables, we would need to determine 9 unknowns (the three λ_k and the six entries of the *M* matrix)

from only 5 equations (two equations for the mean values, two for the variances and one for the covariance).

The main idea is this: elementary processes in photon counting detectors will always lead to *small positive integer* multiplicities. A single photon will create 0, 1, 2 or possibly 3 counts in a threshold or bin. Hence it makes sense to restrict the matrix M_{ik} to small positive integers. This means that only a finite number of candidate matrices M need to be considered. Then we can identify the columns of M with the possible *outcomes* from a photon hitting the detector. Hence we call M the *outcome matrix*.

III. Algorithm

This suggests the following procedure:
For *each* meaningful candidate M_{ik}, solve (4) and (5) with the given μ_i and σ_{ij} for the mean values of the independent Poisson random variables λ_k in a least-square sense. Ignore solutions with any negative λ_k.
For all-positive λ_k, calculate the sum of square differences

$$e = \sum_i [\mu_i - \hat{\mu}_i]^2 + \sum_{i \le j} [\sigma_{ij} - \hat{\sigma}_{ij}]^2$$
$$= \sum_i [\mu_i - \sum_{k=1}^m M_{ik}\lambda_k]^2 + \sum_{i \le j} [\sigma_{ij} - \sum_{k=1}^m M_{ik}M_{jk}\lambda_k]^2. \quad (5)$$

A M matrix resulting in a lower quadratic error e is considered more suitable to reproduce the desired statistics of the x random variables. A full search through all candidate M_{ik} returns the most suitable matrix M_{ik}^{best} and the corresponding Poisson parameters λ_k^{best}.

The term "*meaningful candidate* M_{ik}" above implies that not all matrices M need to be considered. For a photon counting detector involving thresholds, for example, we can assume that
$M_{ik} \le M_{jk}$ if threshold i is higher than threshold j. In other words, the columns of M should be non-increasing. If one threshold fires, all thresholds below also must fire. This rule may have to be reconsidered in situations where pile-up plays a role.
From the most suitable matrix M_{ik}^{best} and the corresponding λ_k^{best}, we can calculate the mean resulting values and covariances from equations (4) and (5)

$$\hat{\mu}_i = \sum_{k=1}^m M_{ik}^{best}\lambda_k^{best}, \ \hat{\sigma}_{ij} = \sum_{k=1}^m M_{ik}^{best}M_{jk}^{best}\lambda_k^{best}, \quad (6)$$

which can be compared to the given μ_i and σ_{ij}.

IV. Results

A. Data from simulations

From simulations [1], we can obtain mean values and covariance matrices for the interaction of a single photon with a detector.

1) Example 1: 100 keV photon

For a single 100 keV photon hitting a CdTe detector with four thresholds [20,35,65,80] keV, simulations have predicted the following mean values and covariances:

$$\mu_i = [0.992, 0.807, 0.618, 0.451], \sigma_{ij} = \begin{bmatrix} 1.507 & 1.120 & 0.740 & 0.457 \\ 1.120 & 0.948 & 0.623 & 0.453 \\ 0.740 & 0.623 & 0.620 & 0.453 \\ 0.457 & 0.453 & 0.453 & 0.454 \end{bmatrix}.$$

Note that this deviates markedly from a Poisson statistics: the variance of the lowest threshold is more than 50% larger than its mean. We want to reproduce the statistics from six Poisson random variables, allowing factors up to 3. The optimization algorithm returns the most suitable outcome matrix M_{ik}^{best} and the corresponding λ_k^{best}:

```
M(T20):    1    2    2    2    1    3
M(T35):    1    1    2    1    1    0
M(T65):    1    1    0    0    1    0
M(T80):    1    0    0    0    0    0

lambda:  0.454  0.122  0.070  0.048  0.044  0.005
```

shows that the desired mean values and covariances can be reproduced to high precision:

$$\hat{\mu}_i = [0.994, 0.810, 0.620, 0.454], \hat{\sigma}_{ij} = \begin{bmatrix} 1.507 & 1.120 & 0.740 & 0.457 \\ 1.120 & 0.948 & 0.623 & 0.453 \\ 0.740 & 0.623 & 0.620 & 0.453 \\ 0.457 & 0.453 & 0.453 & 0.454 \end{bmatrix}.$$

2) Example 2: 70 keV photon

Running the same experiment with 70 keV photons where mean and covariance have the form

$$\mu_i = [1.195, 0.920, 0.440, 0], \quad \sigma_{ij} = \begin{bmatrix} 1.651 & 1.114 & 0.439 & 0 \\ 1.114 & 0.925 & 0.439 & 0 \\ 0.439 & 0.439 & 0.439 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

with five Poisson variables only, returns for the M_{ik}^{best} and the λ_k^{best}:

```
M(T20):    1    1    2    2    1
M(T35):    1    1    1    0    0
M(T65):    1    0    0    0    0
M(T80):    0    0    0    0    0

lambda:  0.439  0.292  0.191  0.037  0.007
```

which reproduce the statistics almost perfectly:

$$\hat{\mu}_i = [1.195, 0.922, 0.440, 0], \quad \hat{\sigma}_{ij} = \begin{bmatrix} 1.651 & 1.114 & 0.439 & 0 \\ 1.114 & 0.922 & 0.439 & 0 \\ 0.439 & 0.439 & 0.439 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

B. Example 3: Measured data

It is also possible to analyze measured data in this manner. Data were obtained from a Siemens NAEOTOM Alpha CT scanner with CdTe detector using thresholds [20,40,56,75] at 140 kV, 10 mA, attenuated by 20 cm of water. Combining the 4x6 subpixels pertaining to one ASG opening, mean values and covariances were measured to

$$\mu_i = [17.99, 12.59, 8.80, 4.17], \quad \sigma_{ij} = \begin{bmatrix} 21.93 & 13.82 & 9.42 & 4.40 \\ 13.82 & 12.77 & 8.86 & 4.17 \\ 9.42 & 8.86 & 8.59 & 4.17 \\ 4.40 & 4.17 & 4.17 & 4.17 \end{bmatrix}.$$

The data come from a polychromatic spectrum and are not normalized, but we can run the same procedure with nine independent Poisson variables to obtain the most suitable outcome matrix M_{ik}^{best} and the corresponding λ_k^{best}:

```
M(T20):   1   1   1   1   2   2   2   2   2
M(T40):   1   1   1   0   0   1   1   1   2
M(T56):   1   1   0   0   0   0   1   1   1
M(T75):   0   1   0   0   0   0   0   1   0

lambda: 4.23 3.94 3.21 2.66 0.83 0.51 0.32 0.23 0.08
```

From this outcome matrix containing nine possible outcomes, the measured mean values and covariances can be reproduced quite well:

$$\hat{\mu}_i = [17.99, 12.60, 8.80, 4.17], \quad \hat{\sigma}_{ij} = \begin{bmatrix} 21.93 & 13.82 & 9.42 & 4.40 \\ 13.82 & 12.76 & 8.88 & 4.17 \\ 9.42 & 8.88 & 8.80 & 4.17 \\ 4.40 & 4.17 & 4.17 & 4.17 \end{bmatrix}.$$

C. Interpreting the results

The results can be interpreted directly: Each column of the *M* matrix pertains to one elementary process governed by one of the independent Poisson random variables. The non-zero values in the column show the thresholds firing if this process occurs. Factors bigger than one pertain to processes where a photon is counted multiple times in neighboring pixels.

1) 100 keV photon

The statistics of the interaction of the 100 keV photon with the detector can be reproduced considering six elementary processes: the desired case where the 80 keV threshold fires (45%), the case of the 65 keV threshold firing while some energy goes to the next pixel, producing a double count in the 20 keV threshold (12%). The energy can be split roughly by half so that in two neighboring pixels the 35 keV thresholds fire (7%), the energy can be split such that the 35 keV threshold fires in one and the 20 keV threshold in the next pixel (4.8%), the case where some energy is lost so that only the 65 keV threshold fires (4.4%) and finally the rare but interesting case where a photon is counted three times in the 20 keV threshold (0.5%). The percentages do not add up to 100% since there is also the case where the photon hits the detector unobserved.

2) 70 keV photon

For the 70 keV photon, since the 80 keV threshold can never fire, only five processes are required to reproduce the statistics: The desired case where the 65 keV thresholds fires (44%), the case of the 35 keV threshold firing (29%), the double counting case with the 35 keV threshold firing in one pixel and the 20 keV threshold in the next pixel (19%), the double counting case where only the 20 keV thresholds fire in two pixels (3.7%) and finally the case where most of the energy is lost so that only the 20 keV threshold fires (0.7%).

3) Measured data

The measured data can be interpreted similarly. The main difference is that the data are not normalized so that it is not possible to talk about percentages relative to the number of incident photons. Also, the incoming photons have a polychromatic spectrum so that the interpretation is not that straightforward.

The most frequent process is the firing 56 keV threshold (including all lower thresholds), followed by the 75 keV threshold, the 40 keV threshold and the lowest (20 keV) threshold. Double counting contained in the last five outcomes accounts for only ~12% of total outcomes. This is surprising since our current simulation models suggest a much higher probability (more than 25%) of double counting. This suggests that current models need improving.

V. Summary

The paper shows that the complex statistics of photon counting detectors can be modelled to high precision as a linear combination of several uncorrelated Poisson variables with small integer weights. The matrix used to transform the uncorrelated Poisson random numbers into the threshold statistics can be interpreted as a matrix of potential outcomes.

The presented approach can be useful in two ways:

1. It defines a simple strategy to accomplish the non-trivial task of *generating random variables* for realistic simulations following the statistics of a photon counting detector *including correlations and multiple counting*. A detector simulation (e.g. following [1]) provides expectation values and covariances for each detector pixel. To create noisy sinogram data, the procedure needs to be performed for each detector pixel in every reading. However, it should be sufficient to keep the outcome matrix *M*, once found, fixed, so that only the (pseudo-) inversion of equations (4) and (5) needs to be

performed for each detector pixel which can be done efficiently.

2. Using the *measured* statistics (μ_i, σ_{ij}), it is possible to find a meaningful outcome matrix M_{ik} and probabilities λ_k that can reproduce this statistic. This can be a tool to look deeper into the inner workings of the detector and serve to further evaluate and improve simulation models.

VI. Future work

The approach can be extended in many directions. It can be used to analyze the statistics of photon counting detectors at elevated count rates where pile-up starts playing a role. It will be interesting to analyze the combined statistics of neighboring pixels or the statistics of pixels at different locations in the detector.
To provide realistic simulations, the method can be extended to include also spatial correlations (crosstalk). This would imply a larger *M* matrix that also contains outcomes affecting neighboring pixels.

The talk at the conference will present a simpler and more direct approach showing that the λ_k can be identified with the product of the average number of (Poisson distributed) input photons *N* and the probability of outcome *k*: $\lambda_k = Np_k$. The p_k can be directly obtained from simulations of the detection process.
This alternative approach also shows the connection to the work of K. Taguchi [3] where double counting is accounted for using extra Poisson random numbers taking care of correlations with neighboring pixels.

Acknowledgment

The author is indebted to Frédéric Noo and Martin Hupfer for inspiring and clarifying discussions.

References

[1] K. Stierstorfer, M. Hupfer, *Modeling the DQE(f) of photon-counting detectors: impact of the pixel sensitivity profile*, Phys. Med. Biol. **64** (2019) 105008.
[2] K. Stierstorfer, M. Hupfer, *A Monte Carlo method to assess the spectral performance of photon counting detectors*, Med. Phys. 2024, 1-11.
[3] K. Taguchi *et al.*, *Spatio-energetic cross talk in photon counting detectors: Detector model and correlated Poisson data generator*, Med. Phys. **43** (2016) 6587ff.

Spectral performance comparison of counting schemes for photon counting x-ray detectors with anti-scatter grids

Katsuyuki Taguchi, Patrick Riehl, and Scott S. Hsieh

I. Introduction

Photon counting detectors (PCDs) for x-ray computed tomography (CT) are the future of CT imaging [1, 2]. The performance of PCDs for various spectral tasks may be degraded due to charge sharing and pulse pileup. In this study, we focus on the effect of charge sharing as it occurs with all of the detection events, whereas pileup may be severe only for 5–10% of data.

Various detection schemes have been proposed to tackle these problems from different angles, and recently two new schemes have been added to the mix: Quantitative Pulse Selection (QPS) and Charge Sharing Rejection (CSR). QPS records single-counting events separately from other counts that may have been affected by charge sharing or pulse pileup (Fig.1); CSR rejects events that may be recorded with charge sharing. Thus, both QPS and CSR schemes are expected to produce more accurate x-ray spectrum than conventional methods but may record fewer events. The sensitivity to pileup effect is not clear.

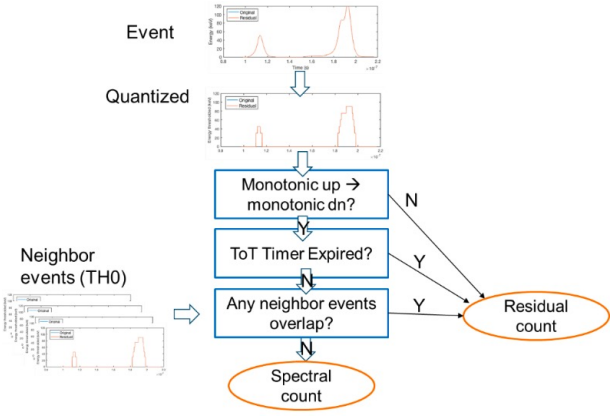


Figure 1. The quantitative pulse selection (QPS) detection scheme.

In this study, we compared the performances of eight detection schemes including QPS and CSR using Monte Carlo (MC) simulation. In this study, we incorporated the effect of anti-scatter grid (ASG) for the first time, which is a critical component for CT systems.

II. Methods and Materials

A. PCDs

We chose the following eight PCDs for this study: (1) threshold–subtract (**TS**), which counts up-crossing events for each threshold energy and subtract the output of adjacent energy thresholds after data acquisition; (2) direct energy binning (**DB**) [3], which counts events within each energy window; (3) **CSR**, which counts events only if there is no coincident events with adjacent pixels; (4) analog charge summing (**ACS**) [4], which detects coincident events and reconstruct the incident energy for each event, simulating Medipix3RX; (5) digital count summing (**DCS**) [5], which is a digital version of ACS. DCS works on digital counters signals whereas ACS uses analog signals; (6) coincidence counting bin (**CCB**) [6]; (7) multi-energy inter-pixel coincidence counter (**MEICC**) [7]; and (8) "**QPS-lite**," which incorporates coincidence part of QPS but without pulse analysis mechanism.

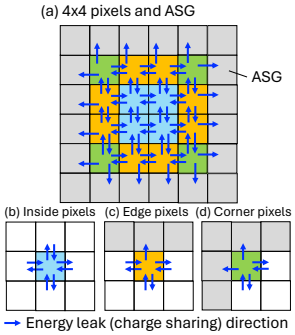


Figure 2. 4×4 PCD pixels, anti-scatter grid (ASG) and 3 types of PCD pixels based on the number of adjacent "silent" pixels, which changes the amount of spill-in charge sharing and whether coincident events are detected or not.

B. Monte Carlo simulator

We used the Monte Carlo simulation program developed for previous studies [3], which cascaded the following processes: (1) photon generation with Poissonian randomized number and time intervals, (2) energy deposition based on a randomized incident location and detection probability, (3) pulse train generation with electronic noise, (4) comparator detection signal generation with up-crossing events of energy thresholds, and (5) the corresponding detection scheme. We had 4×4 pixels surrounded by ASG (Fig. 2), and the detector pixels underneath the ASG received no x-rays, did not detect events (thus, did not contributed to coincidence detection), and were grounded, isolating the 4×4 pixels.

We used CdTe with a pixel size of $(300\ \mu m)^2$, a thickness of 2.0 mm, a Gaussian charge cloud with a diameter of 48 μm at full-width-at-half-maximum (FWHM) [5], pulse duration of 29 ns at FWHM, and the electronic noise of 2.5 keV at FWHM.

KT is with Johns Hopkins University; PR is with Analog Devices; SSH is with Mayo Clinic.

The energy deposition for Step (2) included charge sharing and stochastic temporal effects used in [8].

C. Evaluation methods

We evaluated the performances of eight PCDs outlined in Sec. I in terms of the fidelity of the recorded spectrum, counting capabilities, and three imaging tasks. The three imaging tasks are (a) the conventional CT imaging, (b) water–bone material decomposition, and (c) K-edge imaging with gold.

It has been endless debate on how to characterize their performances, especially in the context of CT imaging. We believe that a PCD's performance can be best assessed by quantifying the statistical information content in the PCD output data using the Cramér–Rao lower bound (CRLB). We believe that a few imaging tasks should be used to capture the diversity of possible clinical tasks and we therefore studied (a) the conventional CT imaging, (b) water–bone material decomposition, and (c) K-edge imaging. The performance of other spectral applications such as virtual mono-energetic CT imaging or virtual non-contrast-enhanced imaging can be derived from (a)–(c), because these applications use the output of material decomposition such as basis line integrals or basis density maps.

The following settings were used: the flat-field x-rays of 140 kVp spectrum with 2 mm aluminum and 20 cm water filtration operated at tube current values of 1–3,800 mA. X-rays were incident onto a set of 4×4 pixels. We assumed that the anti-scatter grid surrounded these pixels, and that no photons landing on pixels outside the 4x4 would create charge within the 4x4. Threshold energies were set at [20, 45, 65, 80, 120] keV. The output of each pixel was used for the assessment of the spectrum and counting capability, whereas the outputs of 4×4 pixels were summed to create a macro-pixel (or super-pixel) output with (1.2 mm)^2 aperture. 500–2,000 noise realizations were used for each tube current setting. Bootstrap resampling was performed 100 times for each setting and the standard deviation was computed over the bootstrap samplings.

III. Results

Spectrum. The recorded spectra with 80-keV input (Fig. 3) shows different PCDs having different effects of three locations. With TS, DB, and their family PCDs, less spill-in charge sharing from PCD pixels underneath ASG resulted in lower fluorescent peaks at corner pixels compared to inside pixels (arrow, Fig. 3a). With ACS and DCS, PCD pixels underneath ASG did not report coincident events, even though partial photon energy was leaked to them. This absence of coincidence counting pair resulted in worse energy recovery, demonstrated by lower photo-peaks and higher K-escape peaks (curved arrows and line arrows, respectively, Fig. 3c,d). QPS-lite and CSR had similar effect (Fig. 3b). An interesting finding was that the photo-peak height of QPS-lite/CSR was comparable to that of TS/DB and lower than those of ACS and DCS, because QPS-lite/CSR does prevent spectral contamination by coincident events but it does NOT recover the energy of charge sharing events.

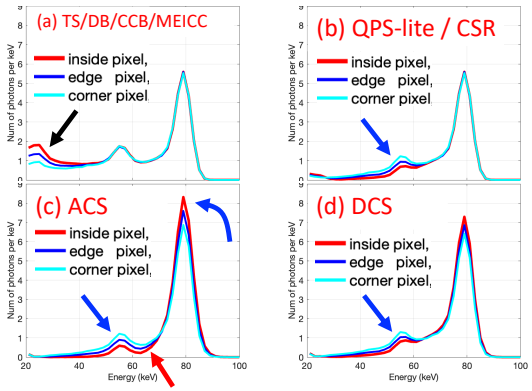


Figure 3. The spectra recorded by PCDs at 3 different locations.

Counting capabilities. Recorded count-rates of DB was higher than TS, and those of ACS and DCS were lower (Fig. 4a), which agreed with previous reports. The count rate curves (CRCs) of CSR and QPS-lite spectral bins were even lower than ACS/DCS, because photons with true coincident (i.e., charge sharing) or incidental false-coincidence (via random incident) detections were removed from their spectral bins, with increasing probability with increasing count-rates. Comparing CRCs of corner pixels against those of inside pixels, TS/DB had a lower CRC due to less spill-in events, whereas the other PCDs had more linear CRCs and more counts, as they had less pileup effects. One can summarize that corner pixels had mild pileup effects while inside pixels had severer pileup effects.

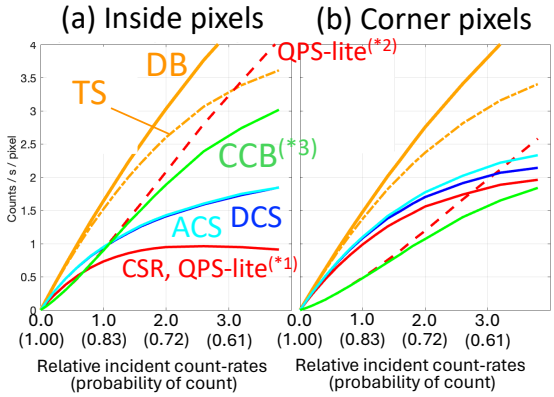


Figure 4. Count-rate curves at 2 different locations. (*1,2) Spectral bins (*1) and residual bin (*2) of QPS-lite. (*3) The coincidence counter of CCB; the spectral bins of CCB and MEICC are the same as DB.

Spectral imaging tasks. The normalized CRLB values for the three imaging tasks are presented in Fig. 5. We made a few observations for each imaging task.

For conventional CT imaging, five PCDs were clustered together and had better CRLBs than TS and DB by ~10%. CSR had a worst CRLB (worse than TS/DB by 13%) probably because fewer counts were recorded as charge sharing events were excluded.

For water–bone material decomposition imaging, first, in the absence of pileup, CRLBs of TS and DB were similar (normalized to 1.0) and 35–55% higher than the other PCDs. Second, with no pileup, the CRLB values were from the best to worst, ACS, MEICC, DCS, CCB, QPS-lite, then CSR with some separation, then far away at 1.0 were DB and TS. The top

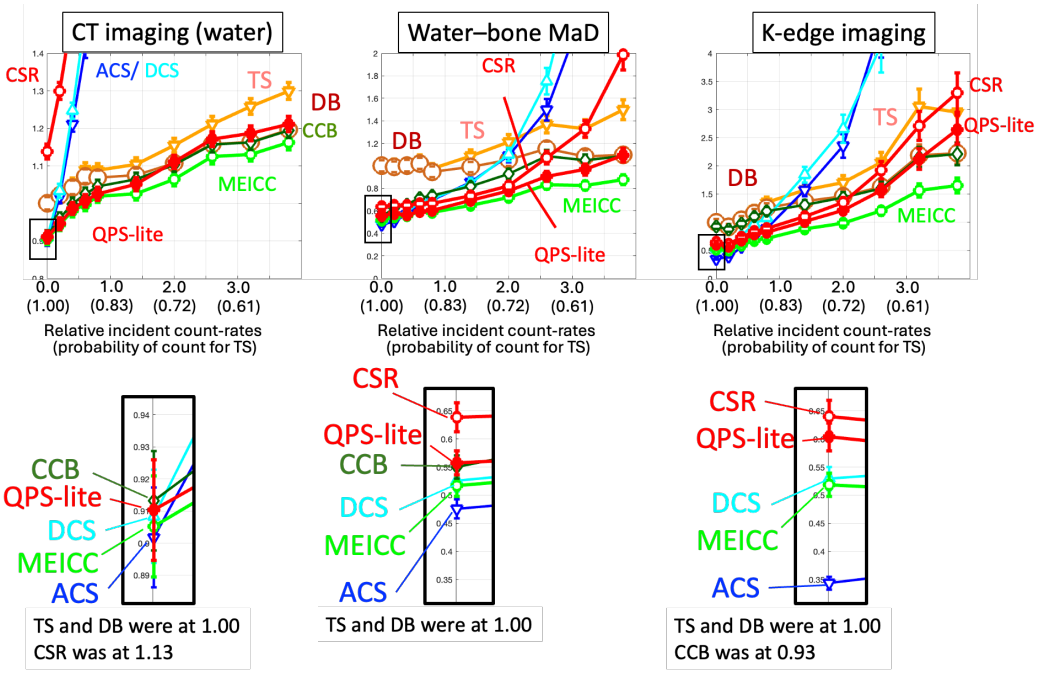


Figure 5. Normalized CRLB values against that of TS at 1 mA. Panels in the top row present data over a large range; those in the bottom row show data inside small boxes at the count-rates of ~0.0 (when the probability of count for TS was 1.00).

five PCDs seemed to address charge sharing effect effectively. Third, CRLB values of any PCDs increased almost monotonically with increasing incident count-rates (i.e., tube current values), due to pulse pileup. Fourth, ACS, DCS, and CSR were more sensitive to pileup and increased CRLBs values more quickly than the other PCDs. Fifth, MEICC stayed better than DB and TS throughout the count-rate range investigated; and CCB, and QPS-lite remained better until the very end.

For K-edge imaging, ACS was the best at 0.39; both MEICC and DCS formed the second group at 0.52–0.53; the third group were QPS-lite and CSR at 0.60–0.65. The other three PCDs (TS, DB, and CCB) were the worst at ~1.00. The sensitivity to count-rates (hence, pileup) was similar to water–bone material decomposition, although ACS/DCS passed TS/DB at lower count-rates (at count-rates of at 1.0 and the probability of count at 0.83 versus 2.0 and 0.72). Spectral coincidence counters used in MEICC somehow provided statistical information to combat both charge sharing and pulse pileup. These two effects may appear in different parts of spectral coincidence counters, and it will be of interest in the future work.

IV. Conclusion

In this paper, we studied the expected performance of eight PCDs in the context of CT imaging and the effect of ASGs independently. We obtained several interesting findings and would like to share them with the community.

References

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical Physics,* vol. 40, no. 10, p. 100901, 2013. [Online]. Available: http://dx.doi.org/10.1118/1.4820371.

[2] K. Taguchi, I. Blevis, and K. Iniewski, *Spectral, Photon Counting Computed Tomography* (Devices, Circuits, and Systems). New York: CRC Press. Taylor & Francis Group, 2020.

[3] K. Taguchi and S. S. Hsieh, "Direct energy binning for photon counting detectors: Simulation study," *Medical Physics,* vol. 51, no. 1, pp. 70-79, 2024, doi: 10.1002/mp.16841.

[4] R. Ballabriga, M. Campbell, E. H. M. Heijne, X. Llopart, and L. Tlustos, "The Medipix3 Prototype, a Pixel Readout Chip Working in Single Photon Counting Mode With Improved Spectrometric Performance," *IEEE Transactions on Nuclear Science,* vol. 54, no. 5, pp. 1824-1829, 2007.

[5] S. S. Hsieh and M. Sjolin, "Digital count summing vs analog charge summing for photon counting detectors: A performance simulation study," *Medical Physics,* vol. 45, no. 9, pp. 4085-4093, 2018, doi: 10.1002/mp.13098.

[6] S. S. Hsieh, "Coincidence counting for charge sharing compensation in spectroscopic photon counting detectors," *IEEE Transaction on Medical Imaging,* vol. 39, no. 3, pp. 678-687, 2020, doi: 10.1109/TMI.2019.2933986.

[7] K. Taguchi, "Multi-energy inter-pixel coincidence counters for charge sharing in photon counting detectors," *Medical Physics,* vol. 47, pp. 2085-2098, 2020, doi: 10.1002/mp.14047.

[8] K. Taguchi and J. S. Iwanczyk, "Assessment of multi-energy inter-pixel coincidence counters for photon-counting detectors at the presence of charge sharing and pulse pileup: A simulation study," *Medical Physics,* vol. 48, no. 9, pp. 4909-4925, 2021, doi: 10.1002/mp.15112.

Extension of an Open and Vendor-neutral DICOM-CT-PD Format for Photon-Counting-Detector CT

Zhongxing Zhou, Scott Hsieh, Zitong Yu, Jarod Wellinghoff, Solveig Narum, Cynthia McCollough, Lifeng Yu

***Abstract*—Access to raw CT projection data from commercial scanners remains highly restricted, limiting independent research in image reconstruction and AI-based methods. The DICOM-CT projection data (CTPD) format developed previously has enabled public sharing of raw data from real patient exams in an open and vendor-neutral format, which has supported researchers across the globe to develop, train, validate, and test various AI-based algorithms. However, the original CTPD format does not support photon-counting-detector CT (PCD-CT), despite its increasing clinical adoption. In this work, we extend DICOM-CT-PD to accommodate PCD-CT projection data. The proposed framework incorporates multi-energy threshold and multi-resolution information using a DICOM multi-frame structure to manage data volume efficiently. Restricted access to raw count data is addressed through noise-equivalent estimates and a calibration process. In addition, validated projection-domain lesion and noise insertion techniques are integrated, enabling realistic simulation datasets across disease and dose conditions. This work provides a standardized, vendor-neutral platform for PCD-CT projection data sharing, facilitating future collaborative research in image reconstruction, image quality assessment, and protocol optimization on PCD-CT.**

***Index Terms*— Computed Tomography (CT), Photon-counting-detector CT (PCD-CT), dual-energy CT, multi-energy CT, image reconstruction, image quality assessment**

I. Introduction

ACCESS to CT projection data acquired from commercial CT scanners is typically restricted due to its proprietary nature. This restriction has limited the development and optimization of image reconstruction mostly to CT manufacturers. With the increased interest of various AI-based algorithms for image reconstruction, data correction, and downstream postprocessing, the restricted access to CT projection data excludes a broader research community from contributing effectively to this field. During the last decade or so, our research group has built a CT patient projection data library using an open and vendor-neutral format [1]. This format, known as DICOM-CT projection data (CTPD), is an extension of DICOM that describes the acquisition geometry of CT projections in metadata. The library, freely available through The Cancer Image Archive (TCIA), contains de-identified patient projection data, which were decoded with manufacturer assistance and converted into DICOM-CTPD [2]. It also includes reconstructed images, patient age and gender, and pathology annotations. This data library, which includes 299 paired lower-dose and original full-dose datasets, has been widely used by researchers across the globe to develop, train, validate, and test various AI-based algorithms, including many deep-learning-based reconstruction and noise reduction (DLR) methods, as well as AI-based segmentation methods and radiomics.

Despite its success, the original DICOM-CTPD only supports energy-integrating-detector (EID) CT. Photon-counting-detector CT (PCD-CT) has many potential benefits over EID-CT such as improved signal, spatial resolution, and radiation dose efficiency, and intrinsic spectral imaging capability. With the increased adoption of PCD-CT into clinical practice, there is an urgent need to extend the original format to accommodate the projection data acquired from PCD-CT such that a wide research community can effectively contribute to the development of various reconstruction and AI-based algorithms.

Corresponding author: Lifeng Yu (Email: yu.lifeng@mayo.edu)

Zhongxing Zhou, Scott Hsieh, Jarod Wellinghoff, Solveig Narum, Cynthia McCollough, and Lifeng Yu are with Department of Radiology, Mayo Clinic, Rochester, MN, 55905, USA

This work was supported by NIH R01 EB036541. We would like to thank Drs. Bernhard Schimdt and Karl Stierstorfer for providing the tools to read the PCD projection data.

In this work, we will extend the DICOM-CT-PD format to PCD-CT. The multi-energy threshold data and multi-resolution data will be incorporated into a DICOM multi-frame structure to consolidate the number of DICOM files. Restricted access to the count data will be addressed using a noise equivalent estimate and a calibration measurement. We will further incorporate validated projection-domain techniques for synthetic lesion and noise insertion [3, 4], facilitating the creation of realistic, annotated simulation datasets for a wide range of multi-energy CT imaging studies.

II. METHODS

A. Development of DICOM-CT-PD Format for PCD-CT

1) General Structure

The extended DICOM-CT-PD format for PCD-CT builds upon the original vendor-neutral DICOM-based format developed for EID-CT projection data. In the original format, each DICOM-CT-PD file stores a single CT projection view as a DICOM image consisting of a header and an image component. The header contains all information required for reconstruction (e.g., gantry geometry, scan parameters) using standard DICOM tags supplemented by private tags for CT-specific geometry. The image component stores the projection matrix as 16-bit unsigned integers, with optional rescaling via Tags (0028,1052) Rescale Intercept and (0028,1053) Rescale Slope to represent line integrals. The matrix has dimensions (N_col, N_row), where N_col is the number of detector columns and N_row is the number of detector rows.

A key feature of the original format is the use of multiple series when projections from the same scan have incompatible properties. Multiple series are generated for multisource (e.g., dual-source CT), fast kV-switching, or multi-layer detectors, with Tags (7033,1061) and (7033,1063) identifying the total number and specific index, respectively.

For PCD-CT, the same principles are applied uniformly. Data from each energy threshold or bin are stored in separate series (distinct Series Instance UID, Series Number, and description, e.g., “PCD Threshold 1” or “PCD Bin 3”). This approach accommodates cases with differing matrix dimensions (e.g., due to binning in threshold modes) as well as uniform multi-bin configurations, ensuring consistency and simplicity. To support both dual-energy/multi-energy CT (DE/ME-CT) and PCD-CT in a general and backward-compatible manner, the original multi-energy tags are extended (see Table 1). For example, when TypeOfMECT is set to 3 (PCD), ListOfSpectrum provides a list of energy windows, such as “[25 120] [65 120]” for 2 energy thresholds, “[25 65] [65 120]” for 2 energy bins, or “[25 45][45 65][65 90][90 120]” for 4 energy bins, allowing flexible accommodation of more than 2 energies. If available, Spectrum describes the detailed array for specific energy as indicated by IndexOfSpectrum.

TABLE I
NEW DICOM TAGS TO INCLUDE DE/ME-CT AND PCD-CT

Tag	Attribute name	Description
(7033, 1059)	TypeOfMECT	A flag used to indicate the type of MECT (0: EID-None; 1: EID-ME-kV; 2: EID-ME-mLayerDetector; 3: PCD)
(7033, 1061)	ListOfSpectrum	List of tube voltages/ detector layers/ energy thresholds/ energy bins
(7033, 1063)	IndexOfSpectrum	Index of tube voltages/ detector layers/ energy thresholds/ energy bins
(7033, 1065)	BHC	A flag to indicate if the data has been corrected for beam hardening or not
(7033, 1066)	Spectrum	An array describing the spectrum of a given tube voltage/ detector layer/ energy threshold/bins

Within each series, to improve storage efficiency and reduce file count for large datasets, a multi-frame DICOM structure (e.g., CT Image Storage SOP Class) is adopted. A single file contains multiple frames, with each frame corresponding to one projection view at a specific gantry angle. Pixel data are encapsulated as fragmented items within the PixelData tag (7FE0,0010), where each item holds the raw 2D projection matrix (N_col × N_row) for one frame. The NumberOfFrames tag specifies the total number of views in the file.

Header parameters follow multi-frame DICOM organization: 1). Shared parameters (constant across frames, e.g., detector spacing, energy level value) are stored in the base dataset or Shared Functional Groups Sequence (5200,9229); 2). Per-frame varying parameters (e.g., source/detector positions, timestamps, focal spot offsets) are stored in the Per-frame Functional Groups Sequence (5200,9230).

Current commercial PCD-CT systems generally do not provide access to raw photon count data due to proprietary restrictions. To overcome this limitation while maintaining broad usability, the extended DICOM-CT-PD format currently stores calibrated noise-equivalent photon number (NEPN) estimates derived from air-scan variances [3]. The NEPN was validated against a limited set of raw count data provided by the manufacturer. This enables accurate noise simulation and realistic low-dose studies without requiring proprietary count information.

The format is designed to remain compatible with direct photon count data should any manufacturer choose to share

it in the future. In such cases, raw counts can be stored using the same 16-bit unsigned integer PixelData representation already described, with appropriate RescaleIntercept and RescaleSlope tags for scaling to physical units. Future extensions may adopt an alternative strategy if higher precision becomes necessary. Thus, the extension supports both the currently practical NEPN-based approach and potential future direct-count datasets, thereby minimizing workflow complexity while maximizing reusability for the research community.

2) *Geometry Definition*

Geometry tags follow the original specifications, including cylindrical/spherical/flat-panel definitions with tags for detector shape, spacing, and source positions. In the multi-frame format, static geometry (e.g., (7029,100B) DetectorShape) is shared within the series. Dynamic elements (e.g., gantry angle per view) are per-frame, ensuring accurate forward/back-projection across views.

B. Projection Preprocessing and Correction

The original projections are stored as linear attenuation values after log transformation and basic corrections. In addition to beam hardening correction (BHC), the projections have been corrected for pixel nonuniformities and bad pixel replacement.

C. Lesion Insertion for PCD DICOM-CT-PD Data

Lesion insertion adapts the projection-domain techniques validated for EID-CT [5, 6], extended to PCD-CT by incorporating energy-dependent attenuation handling [4]. Virtual lesions are modeled from segmented real CT images to ensure realistic shapes and textures.

Linear attenuation coefficients for lesion materials are derived using monoenergetic x-ray spectra tailored to each energy threshold, with effective beam-hardening correction energies calibrated via phantoms containing tissue and contrast inserts. Forward projections are computed analytically (e.g., Siddon's algorithm for primary paths), aligned with scan geometry, and digitally added to baseline projections in the target energy threshold/bin series.

After insertion, per-frame pixel data are updated, and annotation tags record lesion details: (7041,1003) NumberOfLesions; (7041,1004) LesionPathologyArray; and additional tags for lesion location, size, and material as needed.

Validation adapts EID-CT observer studies [5, 6], with PCD-specific metrics from [4] confirming CT number accuracy (≤4.4 HU) and indistinguishability from real lesions.

D. Noise Insertion for PCD DICOM-CT-PD Data

Noise insertion adapts a projection-domain algorithm validated for PCD-CT [3]. The method estimates noise equivalent photon numbers (NEPN) from variances in log-normalized air scan projections, then adds Poisson-like noise scaled to the desired dose reduction factor. This approach is directly compatible with the log-normalized attenuation projections stored in DICOM-CT-PD. Noise is inserted independently into each energy threshold/bin series. Due to manufacturer restrictions preventing access to raw count data, a calibration step converts estimated NEPN to effective count equivalents. Validation in [3] demonstrates excellent agreement in noise power spectra (≤6.3%) and ROI noise (≤5.9%) for simulated low-dose PCD-CT images. The Poisson approximation holds well in typical low-dose regimes with minimal pileup; however, noise correlations between adjacent pixels or bins are not captured by the pure Poisson model and may require further validation for certain spectral imaging applications.

III. RESULTS

The proposed extension effectively manages the large volume of PCD-CT projection data while remaining backward compatible with the original single-frame approach. By consolidating multiple projection views into multi-frame files within each series and using separate series for different energy thresholds or bins, the format significantly reduces file count and overhead compared to storing one view per file. Figure 1 illustrates this structure, showing how Shared and Per-frame Functional Groups Sequences organize constant and varying parameters efficiently across frames.

DICOM-CT-PD (Multi-frame PCD-CT Extension

Header
(0008,0660) Modality
...
(0022,0008) Number of Frames
...
(5200,9229) Shared Functional Group Sequence
(5200,9230) Per-frame Functional Group Sequence
...
(7029,1010) Number of Detector Rows
...
(7033,1059) Type Of MECT
(7033,1061) List Of Spectrum
...
(7037,100A) Type of Projection Geometry

Image
(7FE0,0010) Projection data
(Inside it:
Frame 1: Projection View 1
Frame 2: Projection View 2
...
Frame N: Projection View N
)

Figure 1. Structure of a multi-frame DICOM-CT-PD file for PCD-CT projection data. The file extends the original single-frame format by incorporating multiple projection views (frames) within one DICOM instance, using Shared and Per-frame Functional Groups Sequences for efficient organization. Each energy threshold or bin is stored in a separate series.

Figure 2 demonstrates a key step of estimating Noise Equivalent Photon Number (NEPN) at each energy threshold from routine-dose air scan data. Log-normalized projections are first computed from the air scans, then the variance in these projections is used to derive the NEPN profile for both T1 (low-threshold) and T2 (high-threshold) channels. The plot shows example columns from T1 and T2 projections alongside their corresponding NEPN values.

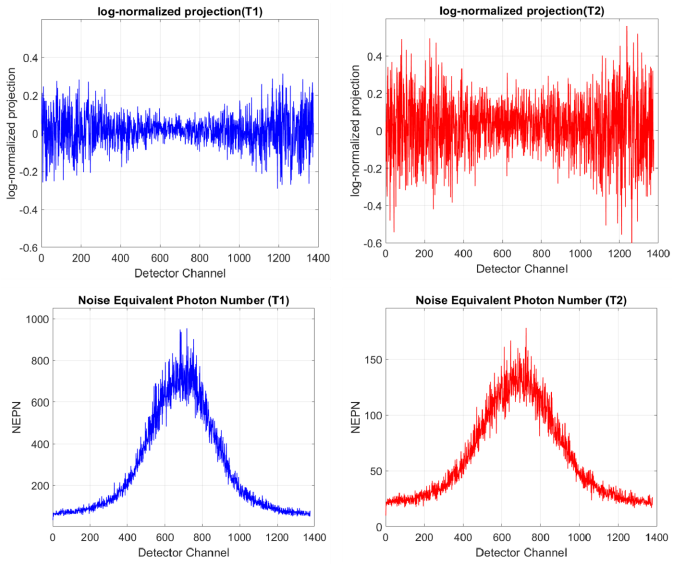


Figure 2. Log-normalized projection of T1 and T2 from air scan and the corresponding NEPN profile of T1 and T2, calculated from the variance in the log-normalized projection data of the air scan.

Quarter-dose projections were simulated from a routine-dose PCD-CT abdomen dataset in DICOM-CT-PD format by inserting NEPN-based Poisson-like noise independently into the low- (T1) and high- (T2) energy threshold series. The original and simulated data were reconstructed using filtered back-projection (FBP). Figure 3 shows representative images from both T1 and T2 series, visually confirming realistic noise texture and scaling in the quarter-dose data. Uniform ROIs (e.g., 20 mm² in liver parenchyma) were placed in the reconstructed images; noise levels (standard deviation, SD) followed the expected inverse square-root dose relationship. For example, SD values were 22.6 HU (routine) and 44.9 HU (25% dose) for T1, and 33.0 HU (routine) and 62.1 HU (25% dose) for T2, consistent with [4].

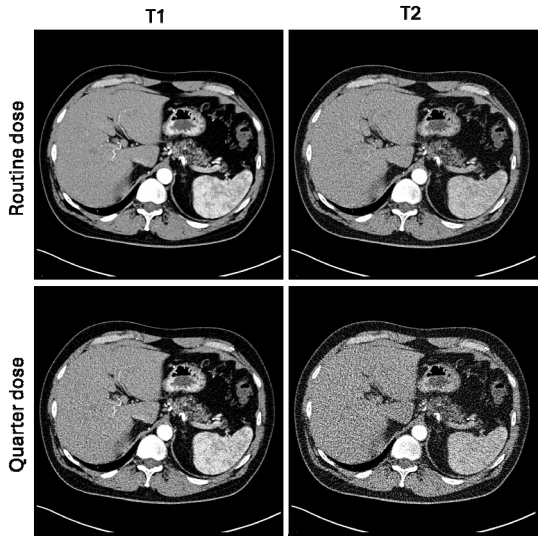


Fig. 3. Reconstructed images from the original routine-dose data and simulated quarter-dose data using filtered back-projection (FBP) for both T1 and T2 series.

IV. Conclusion

The extended DICOM-CT-PD format for PCD-CT offers a standardized, vendor-neutral platform for raw projection data sharing. Multi-frame organization within series efficiently handles large datasets, with new tags for multi-energy threshold data storage. Lesion and noise insertion tools for the DICOM-CTPD data enable simulation of various disease and dose conditions for reconstruction algorithm development, image quality evaluation, and protocol optimization. In the future, we will develop a public and customizable PCD-CT patient projection data library based on this open data format to support collaborative research.

REFERENCES

[1] Chen, B.Y., X.H. Duan, Z.C. Yu, et al., Technical Note: Development and validation of an open data format for CT projection data. Medical Physics, 2015. 42(12): p. 6964–6972.

[2] Moen, T.R., B.Y. Chen, D. Holmes, et al., Low dose CT image and projection dataset. Medical Physics, 2021. 48(2): p. 902–911.

[3] Winfree, T., C. McCollough, and L.F. Yu, Development and validation of a noise insertion algorithm for photon-counting-detector CT. Medical Physics, 2024. 51(9): p. 5943–5953.

[4] Gong, H., S. Kharat, J. Wellinghoff, et al., Insertion of hepatic lesions into clinical photon-counting-detector CT projection data. Physics in Medicine and Biology, 2025. 70(19).

[5] Chen, B.Y., S. Leng, L.F. Yu, et al., Lesion insertion in the projection domain: Methods and initial results. Medical Physics, 2015. 42(12): p. 7034–7042.

[6] Chen, B.Y., C. Ma, S. Leng, et al., Validation of a Projection-domain Insertion of Liver Lesions into CT Images. Academic Radiology, 2016. 23(10): p. 1221–1229.

[7] Zhou, Z.X., H. Gong, S.T. Hsieh, et al., Image quality evaluation in deep-learning-based CT noise reduction using virtual imaging trial methods: Contrast-dependent spatial resolution. Medical Physics, 2024. 51(8): p. 5399–5413.

Design and Characterization of the Pyxis™ X-Ray Detection Platform for CdTe- and CZT-based Photon-Counting Computed Tomography

Josh Star-Lack, Mattias Urech, Devang Savaliya, Simo Veijola, Henrik Lohman, Fredrik Reinholdsen, Kyle Holland, Kevin Holt, Steve Hoelzer, Alex Stewart, Mark Dinh, Thomas Frach, Ivan Mollov, Tuomas Pantsar, Christer Ullberg, Richard Colbeth

Varex Imaging, 1678 Pioneer Rd, Salt Lake City, UT 84104

***Abstract*—Energy-resolving photon-counting (PC) X-ray computed tomography (CT) is emerging as a new standard in diagnostic imaging, offering superior spatial resolution, improved signal-to-noise ratios, and the potential for enhanced spectral reconstructions compared to conventional CT. Direct conversion detectors based on Cadmium Telluride (CdTe) or Cadmium Zinc Telluride (CZT) are well-suited for PCCT due to their high stopping powers, high count-rate capabilities, and strong spectral performance, but can exhibit nonlinearities and instabilities that degrade image quality. Thus, to meet the challenging requirements of PCCT, not only are specialized high-bandwidth detection and counting electronics needed, but also "pre-processing" software-based corrections that address nonlinearities arising from either the sensor or electronics. This work describes the Pyxis™ PC platform, designed to accelerate PCCT development times, comprising: 1) Programmable detector module assemblies with 19.2 mm x 76.8 mm active surface areas and 0.15 mm pixel pitches that can be mounted side by side into a detector arc. Each module contains four custom ASICs bump-bonded to CdTe/CZT tiles and a high-speed interface to the gantry electronics. 2) A software toolkit containing a suite of plugins to linearize the signal in real time. The design and performance of the platform are described.**

***Index Terms*—Photon counting computed tomography (PCCT), CdTe, CZT, X-Ray Detector Design, Pre-Processing.**

I. Introduction

THE detector performance requirements for diagnostic PCCT are formidable **[1,2]**. Linearity and homogeneity errors should be less than 0.5%, stability must be maintained over long periods of time, spectral resolution should be better than 10 keV, and the maximum resolvable count rate should be on the order of 5×10^8 photons/mm^2/sec.

Factors that distort measured count rates and spectral responses and make it challenging to meet image quality requirements include **[3,4]:**

1. Pulse height variations, which cause photons to be assigned to the wrong energy bin or not be counted at all. These variations can have multiple causes, including pulse pile-up, polarization, ballistic deficit, crystalline inhomogeneities, systemic localized electric fields, and temperature sensitivities.
2. Detector electronics instabilities (e.g., energy threshold reference voltage sensitivities to temperature).
3. Charge sharing between pixels, including k-fluorescence.

All authors are affiliated with Varex Imaging Incorporated.

Here we describe the Pyxis PC platform, designed to meet the requirements for PCCT. The platform consists of two key components: 1) Programmable detector modules, containing custom ASICs bonded to CdTe or CZT tiles, which can be readily assembled into a full arc CT detector, and 2) Easy-to-use, real-time, pre-processing software for linearization and artifact reduction.

II. Methods

A. Pyxis Detector Module

The Pyxis CT module (Figure 1) has an active sensor area of 19.2 mm × 76.6 mm, corresponding to 128×512 pixels with a pitch of 150 µm. Each module contains four (128×128 pixel) 4-side buttable photon-counting ASICs bump-bonded to the CdTe/CZT sensor material. For each pixel, the ASIC provides six independently adjustable energy thresholds with separate counters. Data acquisition rates up to 10,000 frames per second are supported. The ASICs have a unique integrated charge-sharing correction capability, which improves energy discrimination accuracy **[5]**.

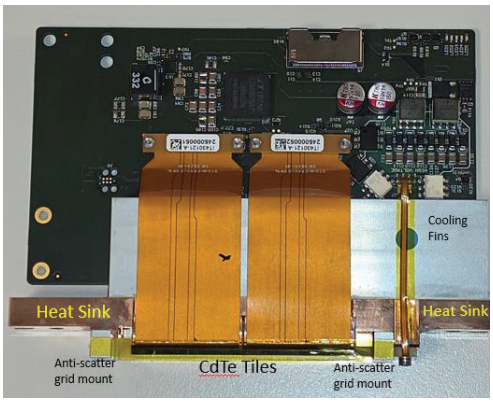


Figure 1. Pyxis detector modules can be placed side-by-side to construct a full-arc detector. Key mechanical components include the CdTe tiles, anti-scatter grid mounts, heat sink, and cooling fins.

A CT module is powered by a 24 V input, from which all component voltages and power needs—including those for the FPGAs, ASICs, and the high-voltage bias circuit—are met. Each module can be connected to a backplane with data rates up to 5 Gbits/sec. Binning is performed outside of the ASIC on the FPGA, and the firmware supports trading off between frame rate, amount of pixel binning, and number of energy bins to stay within the data rate limit. Each module contains a water-coolable heat sink as well as cooling fins. To form a CT arc detector, modules can be placed side by side

with a minimum spacing of 300 μm.

B. *Pre-Processing Software*

The Pyxis pre-processing software is integrated into the Varex Cone-Beam Software Tools (CST) toolkit, which provides a flexible platform for creating and running computationally intensive CT-related processing pipelines [6]. The pipelines comprise modular plugins enabling users to select the specific pre-processing, reconstruction, and post-processing operations they wish to perform. Pipelines are XML file-driven and can be launched by the user's application software through an API.

Many of the corrections are calibration-driven and use tables that CST generates and stores in a "CalBundle" .zip file (Figure 2). Below is a summary of the CST PC-related calibration and correction pre-processing plugins:

1. **Multi-Material Pile-Up and Multi-Point Gain (MPG) Calibration and Correction.** These address: 1). Count rate errors and spectral distortions due to pulse pile-up, and 2) Spatial inhomogeneities caused by pixel-to-pixel variations in recorded pulse heights and in ASIC threshold voltages. Calibration is performed with flood field scans taken at different beam currents of flat, iso-dense slabs of varying thicknesses comprising either a water-like (e.g., PMMA) or bone-like (e.g., titanium) material. A companion paper has been submitted to the conference that describes the algorithm in detail **[7].**
2. **"Vein" Calibration and Correction.** Defects and impurities in the conversion material, which tend to be vein-shaped, create localized electric field gradients that affect charge transport and distort pulse height measurements. Pixels within a vein are less stable and more susceptible to temperature and polarization fluctuations than other pixels. Because these fluctuations can occur over relatively long time periods, a separate correction is needed to augment the temporally invariant MPG algorithm. Calibration for vein correction is performed with longer flood field scans than those used for MPG calibration.
3. **Tile-Edge (TE) Correction**. Electric field gradients affect pixels near tile edges in a manner that is similar to how vein pixels are affected. Tile edge correction is based on analyzing object scan data and does not require calibration.
4. **IntraScan (IS) Calibration and Correction** adjusts for low spatial frequency (non-vein-shaped) count-rate variations caused by polarization and temperature fluctuations occurring over multiple gantry rotations.
5. **Bad Pixel (BP) Calibration and Correction**. The calibration operation identifies pixels with excessively high or low count rates, as well as disproportionally noisy pixels. Correction is performed using 2-D inpainting.

The left part of Figure 2 shows the full **calibration** process, which involves running four pipelines sequentially:

1. Calibration for MPG correction is performed first since the subsequent calibration pipelines require the use of an up-to-date MPG calibration table. The pipeline starts by inputting multiple calibration flood field scans from either disk or memory with the ReadProjectionSets plugin. The MPG *calibration* plugin then analyzes the scan data, computes new calibration tables, and writes them to the CalBundle file.
2. The Bad Pixel Calibration pipeline is next run. After the calibration scans are input, MPG *correction* is performed, and the Bad Pixel *calibration* plugin marks pixels in the output calibration table as defective that were not adequately corrected by the MPG instance.

3,4. The Vein Calibration and IntraScan calibration pipelines are then run. Like in the Bad Pixel calibration pipeline, MPG *correction* executes before the *calibration* plugins.

The right part of Figure 2 shows an example **correction** pipeline that can be run after calibration is completed. This pipeline comprises MPG, TE, Vein, and BP *correction* plugins that read appropriate tables from the CalBundle file as necessary. The ordering of the plugins is important, with MPG placed first to provide basic linearization and BP placed last to inpaint with the best possible replacement data.

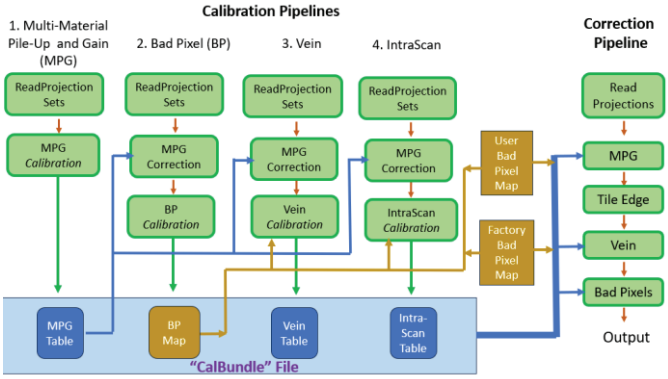


Figure 2. (left) A full CST calibration process involves running four pipelines sequentially, during which time the CalBundle file is both written to and read from. The correction pipeline (right) only reads from the CalBundle file. The Bad Pixel tables (in brown) are user-editable.

C. *Experimental Studies*

To evaluate the Pyxis platform, experiments were performed at 120 kV using a plug-and-play development kit that can contain up to 4 Pyxis modules offering a total active area of 76.6 x 76.6 mm^2. CdTe tile thickness was 1.6 mm, the applied bias voltage was ~-800V, and no anti-scatter grid was attached. Due to x-ray tube power constraints, smaller source-to-imager distances (SIDs) were used than are typical for diagnostic CT. Hence, beam currents are reported in units of "CT-equivalent" mA (CTeq mA), corresponding to the mA that would yield an equivalent flux at a typical 100 cm SID. Except when noted, charge-sharing correction was not used.

The following studies were performed:

1. RQA9 **DQE measurements** were made at 28 and 45 deg C according to the IEC 62220 protocol [8]. Data were collected at 14.5, 145, and 346 CTeq mA at a frame rate of 1440 Hz, corresponding to input doses of 0.05, 0.53, 1.6 uGy/frame.
2. **Detector linearity** was measured for 4 energy bins before and after MPG correction. Calibration was with Air and PMMA slab scans, as described in Section IIB. Tube currents ranged from 18 to 533 CTeq mA, yielding a maximum input count rate of 7.63×10^8 photons/mm^2/sec.

3. **CBCT scans** were acquired of a uniform phantom and a mini-spine phantom. To maximize the imaging field of view, the offset-detector geometry was used with a 28 mm shift applied in the u-direction. Other acquisition parameters were Frame Rate = 1000 Hz, Frames/Rotation = 2000. SAD = 570mm, and SID = 760 mm. The Beam current was 260 CTeq mA. FDK reconstruction with ring correction **[9]** was performed onto a 1000x1000x500 matrix with an isotropic voxel size of 0.1 mm. The detector was calibrated with scans of 25-, 50-, and 100-mm thick PMMA slabs.
4. The **energy resolution** of the detector was measured with and without charge sharing correction enabled using an Am241 radioactive source (59.5 keV peak). ASIC thresholds were swept from 0 to 100 keV in 1-keV steps, and the spectral response was obtained by numerically differentiating the measured counts with respect to the threshold energy.

III. Results

A. MTF and DQE

Figure 2 shows MTF and DQE plots for a 25 keV "open" energy bin containing counts from all photons with energies greater than 25 keV. DQE is largely insensitive to the input flux rate and temperature. The average oversampled MTF of 0.45 at the Nyquist frequency of 3.33 lp/mm is noteworthy but below the ideal (Sinc) value of 0.63. DQE(0) is over 80%.

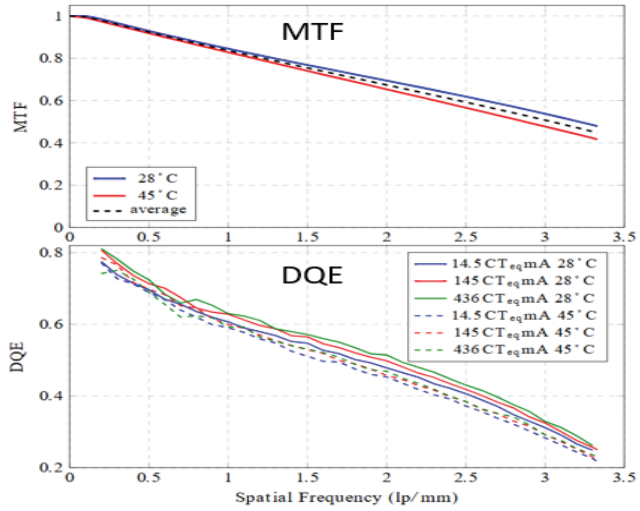


Figure2: RQA9 MTF and DQE curves for the 25 keV open energy bin. Performance is largely in line with expectations for a 1.6 mm-thick CdTe sensor and a 0.15 mm pixel size.

B. Linearity

Figure 3 shows linearity curves of the air scans before and after MPG correction for each of the four energy channels sampled. Pulse pile-up distortions, which cause the measured count rates of lower energy channels to be depressed and count rates of higher energy channels to be elevated, are corrected. Accurate count rate measurements for input flux rates greater than 700×10^6 photons/mm^2/sec are obtained.

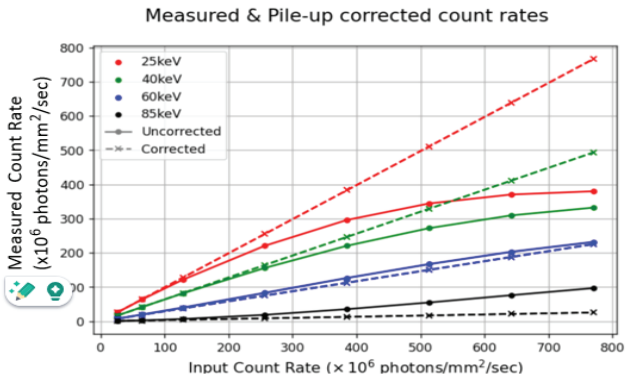


Figure 3: Linearity curves before (solid lines) and after (dashed lines) MPG correction.

C. CT Imaging Studies

Figure 4 shows an example 25 keV open bin projection of the spine phantom before (left) and after (right) corrections with the MPG, TE, Vein, and BP algorithms. With CST pre-processing, homogeneity is greatly improved and dynamic range is increased (air is considerably brighter), which is crucial for achieving good Hounsfield Unit (HU) accuracy.

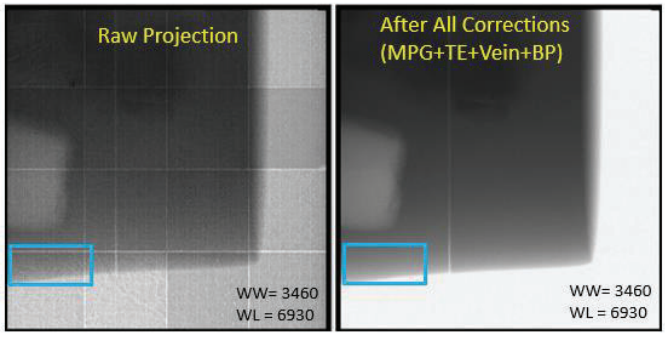


Figure 4. Before (left) and after (right) MPG, TE, Vein, and BP corrections.

Figure 5 shows a zoomed-in area of the part of the projections outlined in blue in Figure 4. The four sub-images show the effects of each correction separately. The raw (uncorrected) sub-image (top left) contains inhomogeneities from structured noise, including veins (yellow arrows), and tile edges (red arrows). After MPG correction, background noise is reduced by 28% with vein and tile edge artifacts removed entirely after the application of their respective corrections.

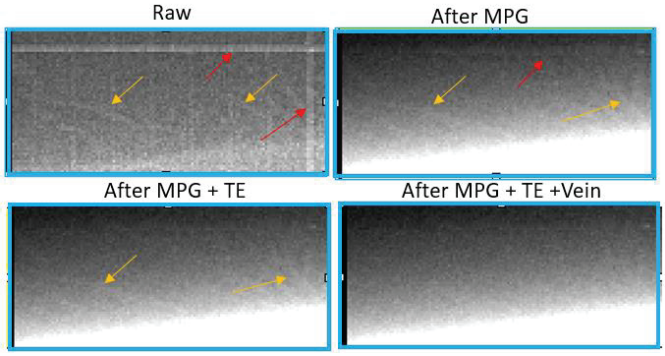


Figure 5: Detail of the area outlined in blue in Figure 4 showing the impact of each correction. The yellow arrows point to vein artifacts, and the red arrows point to tile edge artifacts. The display window width is approximately twice the display level.

Figure 6 shows example slices of the spine phantom reconstructed from projections that are air-normalized without CST pre-processing (left) and or that are normalized after CST pre-processing (right). Figure 7 shows a histogram of the CT numbers of voxels in the slice. The "Soft Tissue" and "Bone" profiles from the reconstruction of CST-pre-processed projections are significantly narrower, and HU values are considerably more accurate than in the reconstructions of air-normalized-only projections.

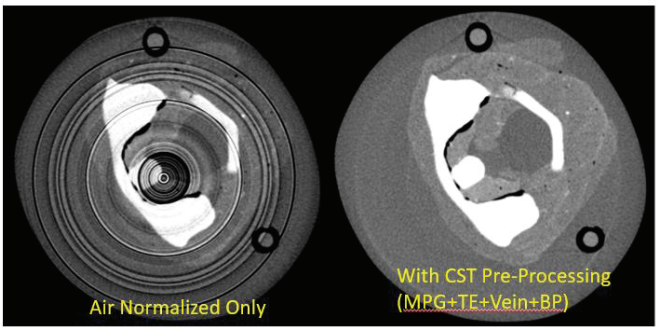


Figure 6: Example reconstructed 25keV open bin axial slice of the spine phantom normalized without (left) and after (right) MG, TE, Vein, and BP corrections were applied. The Window Width = 700 HU

.

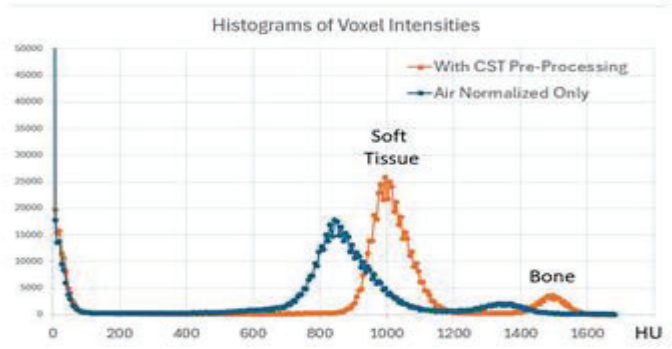


Figure 7. Histogram of voxels in the "Air-Normalized Only" and CST-preprocessed reconstructions in Fig 6. Significantly better HU accuracy and soft tissue and bone separation are achieved with CST pre-processing.

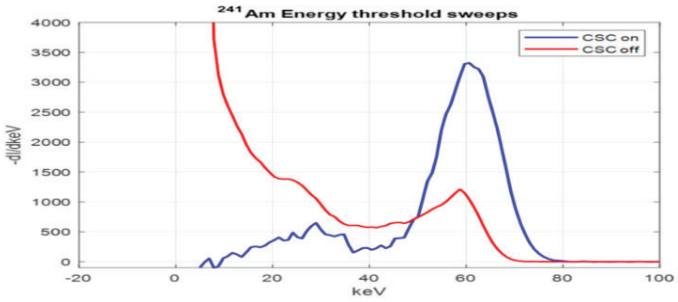


Figure 8: Measured energy response with and without charge sharing correction.

IV. Discussion

Measured DQE was insensitive to input dose as expected for a PC detector. Future studies are required to determine if DQE(0) can be improved beyond 0.8 and why the MTF is not more Sinc-shaped.

The small pixel sizes lead to very high MTFs and count rate capabilities, but they can decrease spectral resolution due to charge sharing between pixels. As confirmed by this study (Figure 8), Pyxis' charge sharing correction feature restores spectral fidelity at the expense of reducing the maximum count rate **[5]**. Investigations are currently underway to explore a hybrid mode where charge-sharing correction is enabled for some energy channels and not for others. Measurements would then be combined in a way that both high count rates and good spectral resolution can be obtained.

The reconstructions of the spine phantom were significantly improved with CST pre-processing. However, some ring and banding artifacts remained, which are believed to be primarily due to temperature- and polarization-related fluctuations. Further algorithmic improvements are under development to address these artifacts.

The CST corrections are CPU-based, with the most computationally intensive parts implemented using the Advanced Vector Extensions instruction set to maximize processing speed. On an advanced workstation with at least 16 cores, throughputs are on the order of 5×10^8 pixels per second (8 Gbits/sec). If necessary, the corrections will be migrated to a GPU platform for faster throughputs.

V. Conclusion

The combined hardware and software components of the Pyxis platform were shown to provide good single-channel CT image quality at high count rates that may be suitable for diagnostic applications. Ongoing work is focused on further artifact reduction and on spectral imaging.

Acknowledgements

The authors thank Yuki Yamahata, Rudiger Schwartz, and Isaias Job for their important contributions.

References

[1] McCollough, Cynthia H., et al. "The technical development of photon-counting detector CT." *European radiology* 33.8 (2023): 5321-5330.

[2] Flohr T, Petersilka M, Henning A, Ulzheimer S, Ferda J, Schmidt B. Photon-counting CT review. *Phys Med.* 2020 Nov;79:126-136.

[3] Ballabriga, R., et al. "Photon counting detectors for X-ray imaging with emphasis on CT." *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.4 (2020): 422-440.

[4] Danielsson, Mats, Mats Persson, and Martin Sjölin. "Photon-counting x-ray detectors for CT." *Physics in Medicine & Biology* 66.3 (2021): 03TR01.

[5] Ullberg, Christer, et al. "Measurements on a novel 4-side buttable photon counting ASIC with integrating charge sharing correction." *Medical Imaging 2022: Physics of Medical Imaging*. Vol. 12031. SPIE, 2022.

[6] Holt, Kevin, et al. "Design and optimization of 3D VSHARP® scatter correction for industrial CBCT using the Linear Boltzmann Transport Equation." *7th International Conference on Image Formation in X-Ray Computed Tomography*. Vol. 12304. SPIE, 2022.

[7] Simo Veijola et. al. "Material-Aware Pulse Pile-Up and Pixel Gain Correction for Spectral Photon Counting Detectors (PCDs)" Submitted to 2026. CT Meeting, Salt Lake City, Utah.

[8] IEC 62220-1:2003 "Medical electrical equipment - Characteristics of digital X-ray imaging devices - Part 1: Determination of the detective quantum efficiency".

[9] Star-Lack, J., et al. "SU-FF-I-04: a fast variable-intensity ring suppression algorithm." *Medical Physics* 33.6Part2 (2006): 1997-1997.

Material-Aware Pulse Pile-Up And Pixel Gain Correction for Spectral Photon Counting Detectors

Simo Veijola, Kyle Holland, Henrik Lohman, Fredrik Reinholdsen,
Mattias Urech, Alex Stewart, Devang Savaliya, Richard Colbeth, Josh Star-Lack
Varex Imaging, 1678 Pioneer Rd, Salt Lake City, UT, 84104

***Abstract*—The additional spectral information provided by photon counting detectors (PCDs) can reduce noise in CT imaging and enable material decomposition. A key limitation at high flux is the nonlinearity from pulse pile-up. In pile-up, photons arriving close in time overlap within the detector shaping time, leading to loss of counts and redistribution of counts between energy channels. Pixel gain variation results from spatially dependent inhomogeneities in the detector response that can be attributable to various physical processes, including ASIC behavior. With pixel gain correction, a uniform detector response is produced under flood field illumination. This work presents a software-level pile-up and pixel gain correction method designed to be material-aware by using the full measured spectrum to estimate the correction factors. The approach is calibration-driven and uses aluminum and PMMA as basis materials with measurements taken over a range of flux levels and filter thicknesses to model the detector response under varying conditions. For the correction, a fast, constant-time Maximum Likelihood Estimation (MLE) step is used to obtain a stable pile-up-free estimate with a spatially uniform pixel response. The method is validated with experimental studies.**

I. Introduction

Spectral photon-counting CT generates material-specific information by measuring the photon-flux spectrum using discrete energy bins. Two non-idealities that distort the measurements and reduce the accuracy of spectral reconstructions are: 1. Pulse pile-up and 2. Pixel gain non-uniformities.

A. Pulse pile-up

During pile-up, temporally close photon detection events lead to pulse overlap, and the photons do not produce independent signals. In *peak* pile-up, a new pulse overlaps near the peak of the previous pulse, and a photon with increased energy is measured. This leads to both loss of counts and the possibility of wrongly detected photon energies. In *tail* pile-up, the overlap happens in the latter part of the previous pulse. Depending on the pulse shape, the second pulse may have an additional positive or negative offset, which also leads to a wrongly recorded photon energy or even loss of counts. Both types of pile-up events, to varying degrees, lead to cross-bin dependencies, which can be modeled if the energy spectrum and energy-dependent pile-up probabilities are known.

Software-level corrections for spectral PCDs have been studied using both analytical and calibration-based approaches. Analytical models describe pile-up formation using presumed pulse shapes and, using Poisson statistics, predict the distorted spectrum. A widely cited method modeled the total spectral distortion from different orders of pile-up [1] and has further been used in likelihood-based estimation [2]. Other approaches estimate pile-up contributions from the detected spectrum by modeling ”convolutions” of the incident spectrum with itself [3], [4]. These methods can be accurate when sufficient spectral resolution is available, which may not be the case for PCDs with a small number of energy bins.

Calibration-driven methods learn the detector response under pile-up conditions and apply a correction using lookup tables or fitted spectral response functions [5], [6]. Neural networks have also been used [7], [8], with promising results in simulation-based studies. The main challenges with these empirical methods are the need for multiple approximations and the large amount of calibration scans required.

B. Pixel gain variation

Pixel-to-pixel pulse height variations, stemming from several sources including threshold variations in the detection ASIC, heterogeneity of the detector crystal structure, and irregularities in pixel electrode contacts, introduce inhomogeneities and structured noise. These pixel gain variations are spectrally dependent, which makes a simple correction based on normalizing to a single reference scan [9] less effective. A more advanced approach is to predict the correction factors using a model calibrated with multiple thicknesses of aluminum and PMMA [10].

C. Contribution

We present a novel empirical combined pile-up and pixel gain correction that is designed to be accurate for any body tissue combination over a wide range of input fluxes. Calibration is performed with flood field scans, taken with different beam currents, of flat, iso-dense slabs of varying thicknesses comprising either a water-like (PMMA) or bone-like (aluminum) material. Conveniently, calibration does not require the use of hybrid slabs made up of combinations of the two materials. Correction is performed using an efficient Maximum Likelihood Estimation (MLE)-based technique to perform an approximate material decomposition from which pile-up and pixel gain correction factors are determined. The method is validated with scans of composite materials.

II. Methods

We consider a fixed frame rate and tube voltage without loss of generality. Assuming b basis materials, the Beer-Lambert

law for the transmitted monoenergetic photon counts $n(E)$ can be formulated as

$$n(E) = n_0(E)e^{-\int \mu(\mathbf{r}(l),E)\,dl} = n_0(E)e^{-\sum_{i=1}^{b}\mu_i(E)t_i}, \quad (1)$$

where $n_0(E)$ are the counts for an air scan, $\mu(\mathbf{r}(l), E)$ is the linear attenuation coefficient at position l along the ray path $\mathbf{r}(l) \in \mathbb{R}^3$, and $\mu_i(E)$ and t_i are the linear attenuation coefficient and the path equivalent thickness of the basis material i, respectively. For diagnostic CT, the dominant interactions between photons and matter are photoelectric absorption and Compton scattering. Ignoring K-edges, this can be modeled with two linearly independent basis materials.

For this study, the two basis materials used are PMMA (P) and aluminum (Al), modeling soft tissue and bone, respectively. We work with a Pyxis detector [11] operating with 4 *closed* energy bins $\mathbf{n} = [n_1, n_2, n_3, n_4]$, with threshold levels

$$T = [25, 40, 60, 85]\, keV,$$

where bin n_1 counts reflect photons with energies between $25keV$ and $40keV$ and so on. The last bin includes all the counts above $85keV$.

Processing is performed after taking the channel-wise natural logarithm of photon counts, $\mathbf{m} = \ln(\mathbf{n})$. Figure 1 shows the measured response $m_{2,P}$ (in blue) of the 40-60 keV bin to varying beam currents I and PMMA slab thicknesses. Due to pile-up, as the tube current increases and slab thickness decreases, the response increasingly deviates from the desired (green) curve, $m^*_{2,P}$. Note that the variable, *normalized flux* (ρ), in units of counts per mA, is monotonic with respect to PMMA thickness. If flux is linear with respect to beam current, the log of expected counts $\hat{m}$ simply equals $\ln(\rho I)$.

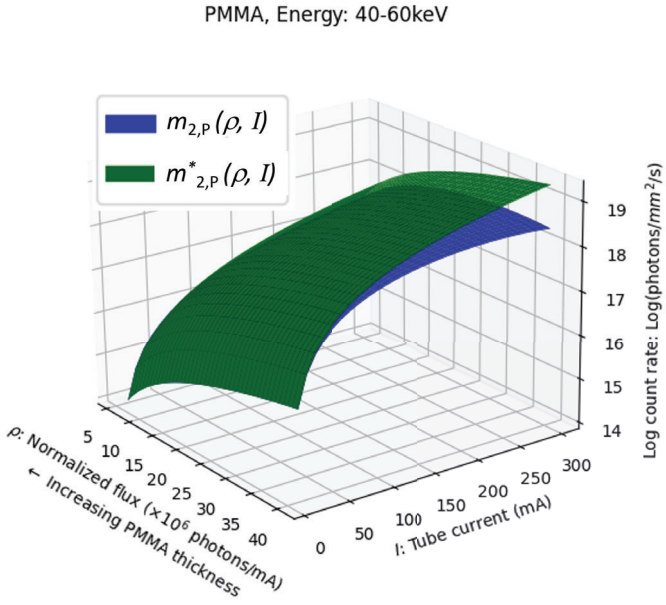


Fig. 1. Actual response $m_{2,P}$ of the 40-60 keV energy bin (in blue) in the presence of pulse pile-up compared to the desired response $m^*_{2,P}$ (in green). "Normalized Flux" is inversely correlated with PMMA thickness.

A simple approach to estimating pulse pile-up for each channel independently is to build a look-up table from the measured and desired responses, as shown in Figure 1 for the 40-60 keV energy channel. Aside from the problem that the approach may not work well for objects containing aluminum, there are other weaknesses to note, which stem from the channels being corrected independently: 1. Determining the proper entry in the calibration table depends on the sampled value of m, which can be noisy. 2. When the lower energy channels start to roll over at high flux rates, their responses become non-monotonic, and a single calibration table is not likely to yield a good solution.

To address these limitations, we take a holistic approach where we use received counts from all energy channels, and separate look-up tables for Al and PMMA, to estimate the amount of pile-up and create a spectrally correct estimate. Let the unknown parameters $\theta = [\rho, \alpha]$ represent the information that determines the expected spectral response for a given tube current I. Figure 2a shows an example of the expected total log counts as a function of PMMA and Al thicknesses. The red diagonal line roughly corresponds to a constant value of ρ. α represents the fraction of each material along ρ.

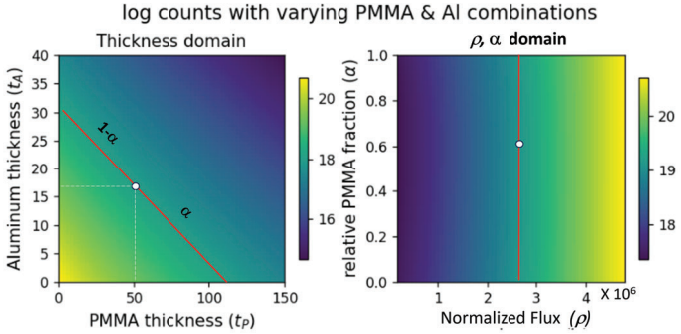


Fig. 2. For two materials, the red isocount curve approximates a line and is identified with constant "Normalized Flux" ρ, which is a good predictor of the amount of pulse pile-up. The parameter α defines the relative amounts Al and PMMA in a sample (white dot).

We will perform an approximate material decomposition to estimate flux- and material-dependent pulse pile-up. We choose to operate in the ρ, α-domain instead of the traditional material thickness domain for the following reasons: 1. Due to the monotonic relationship between the pile-up rate and the total photon count rate, for a constant value of ρ, the pile-up rate is roughly constant. 2. ρ and α define an identical, bounded rectangular domain, unlike material thickness. 3. For MLE, the cost function (Equation (5)) is nearly quadratic in the neighborhood of the solution point (white dot in Figure 2). 4. The use of a single value for α shared across all energy bins enforces consistent material mixing and keeps the parameterization low-dimensioned.

A. Calibration

We acquire Al and PMMA calibration data over a grid of parameter pairs (ρ, I) by varying tube current and material thickness. For each material and each closed energy bin, we construct smooth predictors (cubic-spline-based look-up

tables) of the expected log-measurements to model detector response and pile-up behavior:

$$\hat{\boldsymbol{m}}_{\mathrm{P}}(\rho, I) \in \mathbb{R}^4, \quad \hat{\boldsymbol{m}}_{\mathrm{A}}(\rho, I) \in \mathbb{R}^4,$$

Similarly, pile-up-free reference surfaces

$$\hat{\boldsymbol{m}}^*_{\mathrm{P}}(\rho, I) \in \mathbb{R}^4, \quad \hat{\boldsymbol{m}}^*_{\mathrm{A}}(\rho, I) \in \mathbb{R}^4$$

are generated from reference scans taken at low beam currents with very low amounts of pile-up. Linear extrapolation is used to generate the pile-up-free estimates at higher beam currents.

A separate log-measurement map $\hat{\boldsymbol{m}}_{x,y}$ is generated for each pixel at position x, y. The reference surfaces $\hat{\boldsymbol{m}}^*_{x,y}$ for each pixel are set to be the median pixel response. Hence, for each pixel and each (ρ, I) pair, the reference value is the same, which enables correction of pixel gains.

B. Correction

For each pixel x, y of each frame, taken at time t, we first estimate the unknown parameters $\theta_t = (\rho, \alpha)$ given the frame's beam current I_t. From these parameters, we generate and apply the pile-up and pixel gain corrections.

1) Parameter Estimation: Assuming a relatively high number of counts, the parameterized distribution of the pixel-wise vector of log photon counts is approximately

$$\log \mathbf{n} = \mathbf{m} \sim N(\hat{\boldsymbol{m}}(\theta, I), \Sigma(\theta, I)), \tag{2}$$

where

$$\hat{\boldsymbol{m}}(\theta, I) = \hat{\boldsymbol{m}}(\alpha, \rho, I) = \alpha \hat{\boldsymbol{m}}_{\mathrm{P}}(\rho, I) + (1 - \alpha)\hat{\boldsymbol{m}}_{\mathrm{A}}(\rho, I) \tag{3}$$

is the interpolated expected value vector of log-counts $\mathbf{m}$. The covariance matrix is approximately diagonal with entries proportional to $\Sigma(\theta, I)_{ii} \propto \exp\big(-\hat{\boldsymbol{m}}^{(i)}(\theta, I)\big)$, following from the 1st order Taylor approximation of the logarithmic transform.

We estimate the unknown parameters $\theta = (\rho, \alpha)$ for each pixel using MLE to minimize the negative log-likelihood (NLL). Using the Gaussian approximation (2) the NLL, up to a constant, is:

$$L(\theta) = \frac{1}{2}\big(\mathbf{m} - \hat{\boldsymbol{m}}(\theta)\big)^T \Sigma(\theta)^{-1}\big(\mathbf{m} - \hat{\boldsymbol{m}}(\theta)\big) + \frac{1}{2}\log\big|\Sigma(\theta)\big|. \tag{4}$$

We define $W(\theta) = \mathrm{diag}\big(\mathbf{w}(\theta)\big) = \Sigma(\theta)^{-1}$, and treat the weights as locally constant during the iterations, which yields a weighted least-squares objective in log-space:

$$L(\theta) \approx \frac{1}{2}\sum_{j=1}^{4} w^{(j)}\big(m^{(j)} - \hat{m}^{(j)}(\theta)\big)^2 \propto \|\mathbf{r}(\theta)\|^2_W, \tag{5}$$

where the sum is taken over the 4 energy channels j and we define $\mathbf{r}(\theta) := \mathbf{m} - \hat{\boldsymbol{m}}_t(\theta)$. Newton's method, based on a local quadratic approximation, is used to minimize the objective $L(\theta)$. If $\nabla L(\theta)$ and $H_L(\theta)$ are the gradient and the Hessian, the Newton update with step-size η for the measurement taken at time t for iteration i is:

$$\theta_{t,i} = \theta_{t,i-1} - \eta\, H_L^{-1}(\theta_{t,i-1})\nabla L(\theta_{t,i-1}). \tag{6}$$

We use this step formulation iteratively to estimate the parameters $\theta_t = (\rho_t, \alpha_t)$ at time t:

$$\theta_t = \arg\min_{\rho,\alpha} L(\rho, \alpha) \tag{7}$$

Since measurements do not change dramatically from frame to frame, we can initialize the starting value θ_t at time t with the estimate from the previous frame θ_{t-1}.

2) Application of Correction: Once θ_t is estimated, the combined pile-up-free and pixel gain-corrected counts estimate $\hat{\mathbf{n}}^*_t$ is computed as follows:

$$\hat{\mathbf{n}}^*_t = \mathbf{n_t} \exp\big[\big(\hat{\boldsymbol{m}}^*_t(\rho_t, \alpha_t, I_t)\big) - \big(\hat{\boldsymbol{m}}_t(\rho_t, \alpha_t, I_t)\big)\big] \tag{8}$$

where $\hat{\boldsymbol{m}}^*_t$ and $\hat{\boldsymbol{m}}_t$ are obtained via interpolation as in (3).

C. Experimental Studies

1) Data Acquisition and Processing: Studies were conducted to assess the accuracy of the method for temporally constant flat-field images. The Pyxis detector consisted of 4 tiles, each with 128×128 native pixels of size $150\mu m \times 150\mu m$, combined to a total pixel count of 512×128. These were binned into effective pixels by summing non-overlapping 4×4 blocks of native pixels.

For calibration, several thicknesses of PMMA and aluminum were scanned under varying tube currents. For validation, two thickness combinations of the materials were scanned using the same set of tube currents.

Acquisition Settings:

- Frames: 30 frames, each of $10ms$ duration.
- Source-to-Imager Distancer (SID): $0.8m$
- Tube voltage: $120kVp$
- Tube currents: $[10, 25, 50, 100, 150, 200, 250, 300]mA$

Calibration material thicknesses:

- PMMA: [0 (air), 10, 30, 50, 100]mm
- Aluminum: [0 (air), 3.6, 8, 17, 31]mm

Validation material thicknesses:

- (PMMA, Al): [(20, 7), (50, 10)]mm

The tube currents of $150mA$ and $250mA$ were left out of the calibration set and used for validation only.

For correction, parameter estimation of the first frame was generated using a comprehensive search of the PMMA response curve

$$\rho_0, \alpha_0 = \arg\min_{\rho,\alpha} \big|\mathbf{m}_1 - \hat{\boldsymbol{m}}_P(\rho, \alpha | I_1)\big|.$$

followed by two Newton's iterations, using Eq. (6). Thereafter, 2 iterations were used for subsequent frames.

2) Evaluation Methods: Linearity was evaluated with the R^2 correlation coefficient. Target (pile-up-free) reference values were estimated from the measurements made at low beam currents without pile-up using linear extrapolation.

Comparisons were made between pulse pile-up corrections generated with the proposed MLE-based method using Al and PMMA look-up tables and the target values. The simpler approach described earlier, using only a PMMA look-up table, was also evaluated.

III. Results

Figure 3 shows the median corrected and uncorrected counts for the two validation samples, averaged over all frames, as a function of input count rates. All of the R^2-values of the corrected curves were equal to 1 up to 5 decimal places, showing excellent linearity.

The relative deviations from a linear curve for the composite 7mm Al - 20 mm PMMA validation material are shown in Figure 4. With multi-material correction, errors are less than +/- 1% for all energy bins, while errors are up to 6% using the PMMA-only correction. The small amount of bias in the multi-material correction likely originates from noisy signals and a mismatch between the data and the model.

Figure 5 shows the improved uniformity offered by the correction of pixel gains.

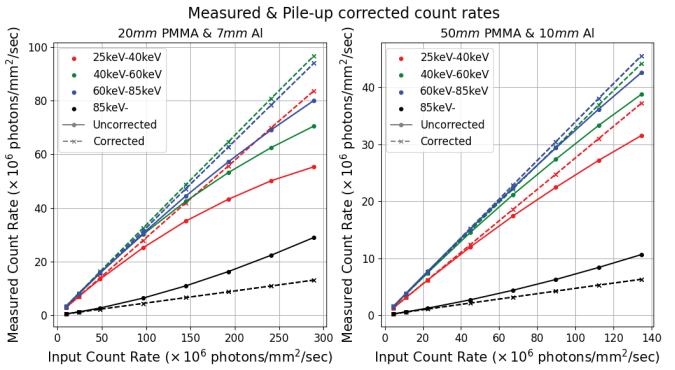


Fig. 3. Measured and pile-up-corrected median count rates plotted against the input count rate for the two validation samples.

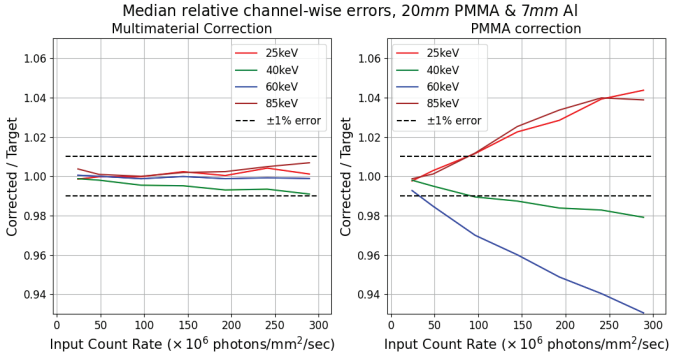


Fig. 4. The relative deviations from a linear curve for 7mm Al- 20 mm PMMA validation material for the multi-material correction (left) and the simpler PMMA-only correction

IV. Conclusion

The model shows promising results for correction of pile-up and pixel gain variation for unseen spectral data. With a limited amount of required calibration data and fast dynamic MLE steps, the method shows promise for real-time clinical CT corrections. Since the method leverages solved effective material coefficients, it can be used as a pre-estimate for a more sophisticated material decomposition (MD). The estimated parameters ρ, α and the corrected measurement vector

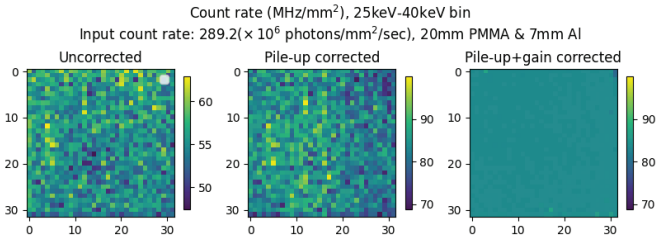


Fig. 5. Validation image ($300mA$, 20mm of PMMA, 7mm of Al) corrected in separate stages

$\mathbf{n}_t^*$ can be mapped to initial estimates of the basis material thicknesses. This can be achieved through a more extensive calibration in the pile-up-free region. The acquired estimates may be useful for solving the maximum likelihood material thicknesses [12].

There are several topics for future research. First, the forward model could be improved by incorporating analytical models and physical constraints during calibration. Second, the inclusion of a third K-edge basis material should be validated. Third, a correction of dynamic images should be validated. Finally, the effect of spectral signal variations on the estimation during long exposures should be studied.

References

[1] Taguchi, K. *et al.* "Modeling the performance of a photon counting x-ray detector for CT: Energy response and pulse pileup effects." *Medical Physics*, vol. 38, no. 2, pp. 1089–1102, 2011. doi: https://doi.org/10.1118/1.353960210.1118/1.3539602.

[2] Taguchi, K. *et al.* "Model-based pulse pileup and charge sharing compensation for photon counting detectors: A simulation study." *Medical Physics*, vol. 49, no. 8, pp. 5038–5051, 2022. doi: https://doi.org/10.1002/mp.1577910.1002/mp.15779.

[3] Plagnard, J. "Comparison of measured and calculated spectra emitted by the X-ray tube used at the Gustave Roussy radiobiological service." *X-Ray Spectrometry*, vol. 43, no. 5, pp. 298–304, 2014. doi: https://doi.org/10.1002/xrs.255410.1002/xrs.2554.

[4] Jumanazarov, D. *et al.* "The significance of the spectral correction of photon counting detector response in material classification from spectral x-ray CT." In *Quantum Optics and Photon Counting 2021*, SPIE, vol. 11771, 2021. doi: https://doi.org/10.1117/12.258929010.1117/12.2589290.

[5] Treb, K., Radtke, J., and Li, K. "Simultaneous Counting and Integrating: A New Paradigm for Pulse Pileup Correction in Photon Counting CT." In *2023 IEEE Nuclear Science Symposium, Medical Imaging Conference*, 2023.

[6] Ding, H. and Molloi, S. "Image-based spectral distortion correction for photon-counting x-ray detectors." *Medical Physics*, vol. 39, no. 4, pp. 1864–1876, 2012. doi: https://doi.org/10.1118/1.3693056.10.1118/1.3693056.

[7] Feng, R., Rundle, D., and Wang, G. "Neural-networks-based Photon-Counting Data Correction: Pulse Pileup Effect." *arXiv*, 2018.

[8] Li, M., Rundle, D. S., and Wang, G. "X-ray Photon-Counting Data Correction through Deep Learning." *arXiv preprint arXiv:2007.03119*, 2020.

[9] Procz, S. *et al.* "Flatfield Correction Optimization for Energy Selective X-Ray Imaging With Medipix3." *IEEE Transactions on Nuclear Science*, vol. 58, no. 6, pp. 3182–3189, 2011. doi: 10.1109/TNS.2011.2165732.

[10] Zhou, E. *et al.* "A cone-beam photon-counting CT dataset for spectral image reconstruction and deep learning." *Scientific Data*, vol. 12, art. 1955, 2025. doi: 10.1038/s41597-025-06246-4.

[11] Ullberg, C. *et al.* "Measurements on a novel 4-side buttable photon counting ASIC with integrating charge sharing correction." *Medical Imaging 2022: Physics of Medical Imaging.* Vol. 12031. SPIE, 2022.

[12] Sidky E. et al.*et al.* "Spectral calibration of photon-counting detectors at high photon flux." *Medical physics* 49.10 (2022): 6368-6383.

Inter-institutional consistency of spectral results from a clinical photon-counting CT: effects of tube voltage and radiation dose level

Leening P. Liu[1], Martin V. Rybertt[1,2], Peter B. Noël[1], Liqiang Ren[3], Xinhui Duan[3], Kishore Rajendran[4], Di Zhang[5], Grace H. Kim[5], Michael F. McNitt-Gray[5]

Abstract— **Photon-counting CT (PCCT) introduces detector technology enabling improved quantitative imaging, creating opportunities for enhanced disease characterization and stratification. To develop biomarkers that accurately capture biological changes, potential quantitative variation arising from detector-related differences must be characterized across institutions. To evaluate the inter-institutional consistency of PCCT, four institutions scanned their copy of a commercial multi-energy phantom on a clinical PCCT. Scans were performed at different tube voltages (120, 140 kVp) and volumetric CT dose index ($CTDI_{vol}$) levels (1 - 10 mGy). For virtual monoenergetic images (VMI), iodine, and virtual non-contrast (VNC) maps, average error for each insert, institution, tube voltage, and dose was determined relative to expected values based on manufacturer specifications. The average and standard deviation of the errors between institutions (inter-institutional) was calculated for each spectral result, insert, tube voltage, and dose. As VMI keV increased from 50 to 150 keV, inter-institutional errors for the blood insert at 140 kVp and 10 mGy decreased from 10 ± 2 to 1 ± 1 HU, respectively, highlighting the dependence of error on VMI energy and the consistency across institutions. A minimal effect of tube voltage was observed with inter-institutional errors of 6 ± 1 HU for the blood insert at both 120 and 140 kVp. Furthermore, as dose increased, the deviation remained consistent at 12 ± 2 and 10 ± 2 HU for VMI 50 keV at 1 and 10 mGy, respectively. Iodine and VNC maps similarly exhibited a stable quantification between institutions for different tube voltages and dose levels. PCCT demonstrated consistent inter-institutional quantification of spectral results with the same phantom and scanner model. The determined variability between institutions increases confidence in PCCT quantification and provides practical guidance for the development and clinical translation of biomarkers.**

Index Terms— **Photon-counting CT, Spectral CT, Quantitative imaging**

I. Introduction

Over the past decade, quantitative imaging has become increasingly adopted in clinical settings to better stratify disease and guide clinical management. Compared to standard subjective diagnostic interpretation, quantitative imaging introduces additional biomarkers that characterize disease, such as shape, size, and tissue composition. Preliminary work has demonstrated that biomarker utilization may improve diagnostic confidence and enable changes in clinical management to ensure more effective treatment [1], [2]. However, existing variability resulting from non-biological factors, such as scanner manufacturer, tube spectra, and patient size, confounds patient-specific disease change [3], limiting the widespread adoption and standardization of quantitative imaging.

The introduction of clinical photon-counting CT (PCCT) has spurred increasing interest in quantitative maps and biomarkers. PCCT utilizes a photon-counting detector that bins photons based on their incident energy, enabling basis material decomposition with every scan [4]. The reconstructed spectral results, such as virtual monoenergetic images (VMI), iodine map, and virtual non-contrast (VNC) reduce the effects of tube spectra, tube filtration, and other protocol parameters compared to conventional images available with existing conventional CT systems. Previous phantom studies demonstrated consistent spectral results across tube voltages, patient habitus, and radiation dose [5], [6], [7], [8]. PCCT's quantitative consistency highlights the reduced impact of hardware-specific variation and facilitates the development and adoption of PCCT-derived biomarkers.

Compared with existing clinical CT scanners that use energy-integrating detectors, PCCT incorporates detector technology with distinct sensitivities and environmental requirements, including detector temperature control. The increased complexity introduces possible instability that may influence quantitative variability and thus requires characterization. While our previous work has established the temporal stability of spectral results with a clinical PCCT [9], the inter-institutional variability is currently unknown. Multi-site scanner consistency is necessary to ensure the robustness of PCCT-derived biomarkers. This study aims to characterize the inter-institutional quantitative consistency with the same scanner model, phantom model, and protocol for different tube voltages and radiation dose levels.

L.P. Liu and P.B. Noël are with the Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. M.V. Rybertt is with the Department of Bioengineering and Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. L. Ren and X. Duan are with the Department of Radiology, UT Southwestern Medical Center, Dallas, TX, USA. K. Rajendran is with Department of Radiology, Mayo Clinic, Rochester, MN, USA. D. Zhang, G.H. Kim, and M.F. McNitt-Gray are with Department of Radiological Sciences, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.

II. Methods

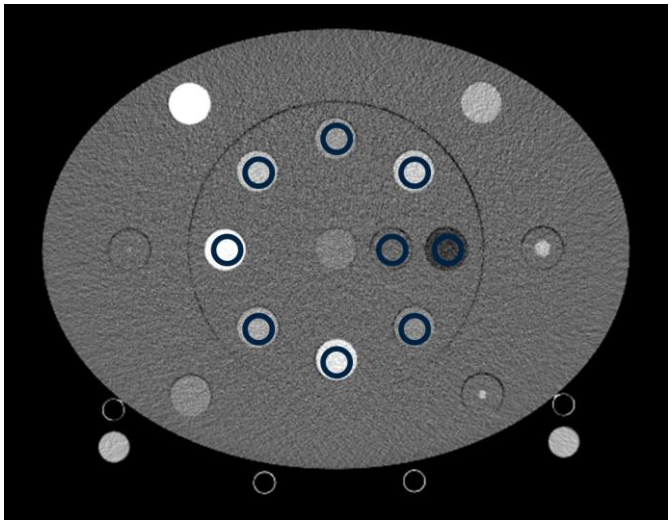


Figure 1. Multi-energy CT phantom. Regions of interest were placed on central inserts to determine the inter-institutional variability for different tube voltages and radiation dose levels. WL/WW: 40/400 HU.

A. Phantom

To evaluate inter-institutional quantitative consistency of PCCT, each institution utilized their own copy of a commercial multi-energy CT phantom (Mult-energy CT phantom, Sun Nuclear, Melbourne, FL, USA, **Figure 1**). The elliptical phantom measured 30 x 40 cm and consisted of tissue- and material-specific inserts. Key inserts were selected to incorporate soft tissue and iodine-containing rods. These inserts include adipose, brain, blood 70, solid water, blood + 2 mg/mL iodine, blood + 4 mg/mL iodine, iodine 2 mg/mL, iodine 5 mg/mL, and iodine 10 mg/mL.

B. Image acquisition and reconstruction

At each of the four institutions, the phantom was scanned with the same model of a first-generation clinical photon-counting CT (NAEOTOM Alpha.Peak, Siemens Healthineers, Forchheim, Germany) running the same software version (VB20, Siemens Healthineers, Forchheim, Germany). Scans were performed in single-source mode with two tube voltages (120 and 140 kVp) and four volumetric CT dose index ($CTDI_{vol}$ of 1, 2.5, 5, and 10 mGy). For each tube voltage and $CTDI_{vol}$ combination, three scans were performed to obtain adequate statistics. VMI from 40 to 190 keV, iodine maps, and VNC were reconstructed with a field of view of 450 mm and a slice thickness of 2 mm. All other acquisition and reconstruction parameters can be found in **Table I**.

TABLE I
Acquisition And Reconstruction Parameters

Tube voltage [kVp]	120 / 140
Exposure time [s]	0.25
Spiral pitch factor	1
$CTDI_{vol}$ [mGy]	1 / 2.5 / 5 / 10
Collimation width	144 x 0.4 mm
Slice thickness [mm]	2
Convolution kernel	Qr48
QIR	2
Field of view [mm]	450
Matrix size	512 x 512
Pixel spacing [mm]	0.88

C. Analysis

To account for manufacturing differences between each institution's copy of the same phantom model, the expected attenuation of each insert $HU_{expected}(E)$ for VMIs from 40 to 190 keV was calculated using the manufacturer provided mass fraction w for each element i, NIST [10] elemental attenuation μ_i at a given energy E, and physical density ρ:

$$HU_{expected}(E) = \left(\frac{\rho \sum_i w_i \times \mu_i(E) - \mu_{water}(E)}{\mu_{water}(E)}\right) \times 1000 \quad (1)$$

The physical density of the inserts was batch specific, and thus specific to each institution's phantom copy. The tolerance of each insert's attenuation was then determined for each energy level by utilizing the manufacturer's physical density tolerance of 0.02 g/mL. Additionally, a basis material decomposition using base pairs of iodine and water was applied to the calculated low and high attenuation (Eqn. 1) of iodine-containing inserts in order to calculate expected iodine and VNC values.

For each institution's data, VMI 70 keV images acquired with a tube voltage of 140 kVp and a $CTDI_{vol}$ of 10 mGy were utilized to automatically place circular regions of interest (ROI) on each insert with in-house software. The determined ROI locations were then used to measure quantitative values on all

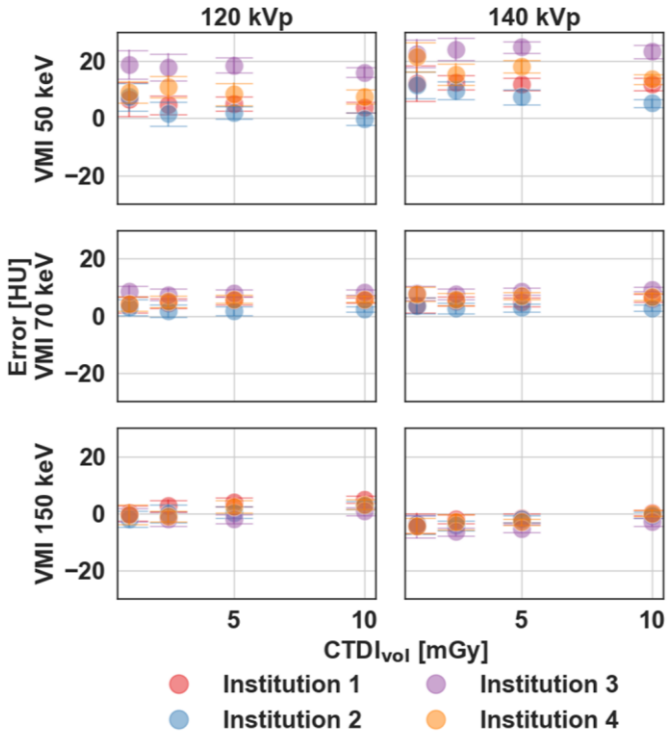


Figure 2. Attenuation error of iodine 5 mg/mL insert for select VMIs for each institution at different tube voltage and radiation dose levels. Errors for VMI 50, 70, and 150 keV were consistent between institutions across dose for each tube voltage. Quantitative values were stable but with more variation between institutions at VMI 50 keV. Error bars represent the standard deviation of measured means across image slices and scan repetitions.

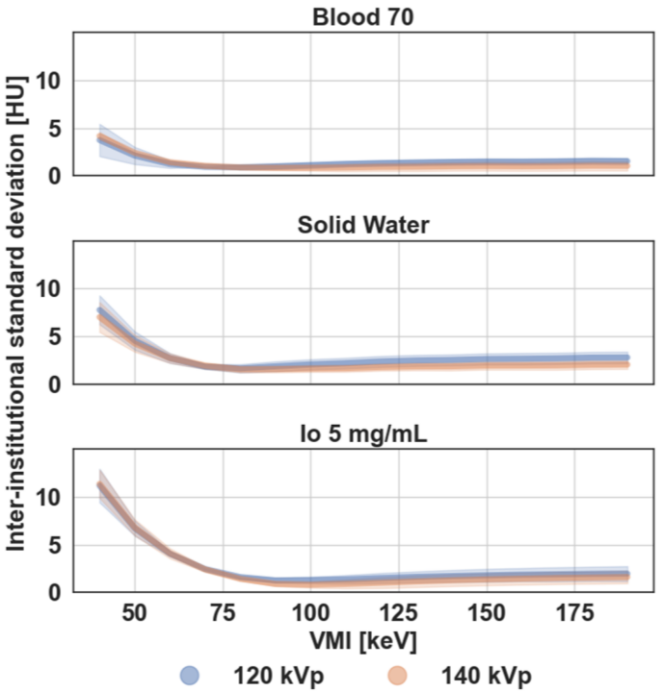


Figure 3. Inter-institutional standard deviation for different inserts and tube voltages. Inter-institutional standard deviations decreased as VMI energy increased and were similar between tube voltages for each insert. Shaded regions correspond to the standard deviation of inter-institutional standard deviations across dose levels for each tube voltage.

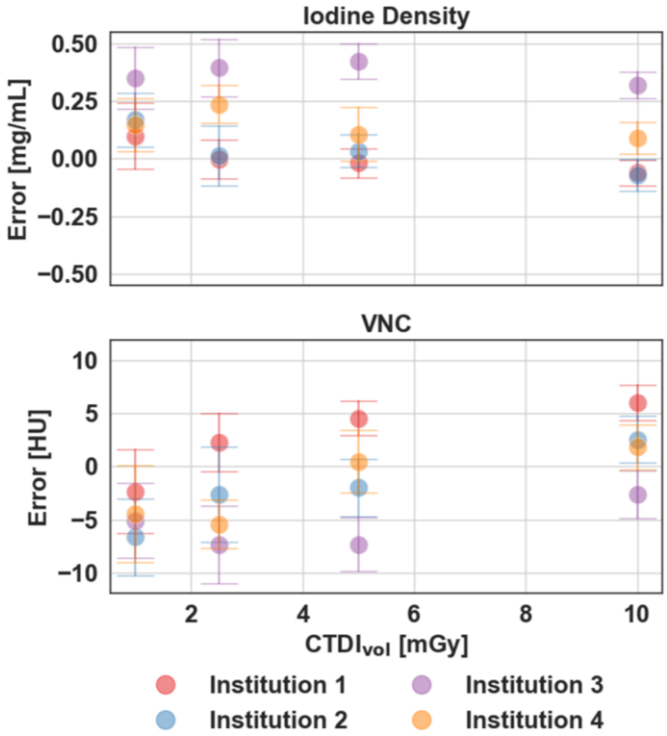


Figure 4. Iodine and VNC error across dose levels for iodine 5 mg/mL insert at a tube voltage of 140 kVp. Iodine (A) and VNC (B) demonstrated quantitative stability at and between institutions from 1 to 10 mGy. Error bars correspond to the standard deviation of measured means across image slices and scan repetitions.

spectral results obtained with different tube voltages and doses. Mean of each ROI was measured across 10 central slices for each repetition (total of 30 slices for each institution, tube voltage, dose combination).

Error for ROI was calculated as the difference between the measured value and expected values for the corresponding institution, insert, and spectral result. Error for each insert, institution, tube voltage, and dose was then averaged across all analyzed slices (10 central slices x 3 repetitions). To quantify the inter-institutional error and standard deviation, the average and standard deviation of error across institutions was calculated for each insert, spectral result, tube voltage, and dose combination

III. Results

Average error for inserts was consistent between institutions across VMIs for each tube voltage. For iodine 5 mg/mL insert, VMI 70 keV error at 10 mGy measured 5 ± 1, 6 ± 1, 2 ± 1, and 8 ± 1 HU for each site at 120 kVp, averaging 6 ± 2 HU (**Figure 2**). At 140 kVp, VMI 70 keV errors were 6 ± 1, 7 ± 1, 3 ± 1, and 9 ± 1 HU, respectively. Similarly, the blood 70 and solid water insert demonstrated an average VMI 70 keV error of 6 ± 1 and 9 ± 2 HU, respectively, for both 120 and 140 kVp. Consistent quantification was maintained at VMIs of high keVs: error of the iodine 5 mg/mL insert averaged 1.0 ± 0.9, -0.7 ± 1.5 and -1.0 ± 1.8 HU at VMI 100, 150, and 190 keV across the four sites at 140 kVp and 10 mGy. Larger deviations between the sites were present at VMIs at low keVs. At VMI 50 keV, errors for each site increased to 12 ± 2, 14 ± 2, 5 ± 1, and 23 ± 2 HU at 140 kVp and 10 mGy, corresponding to an inter-institutional standard deviation of 7 HU (**Figure 3**). A similar inter-institutional standard deviation of 7 HU was observed at 120 kVp. For the remaining inserts, inter-institutional standard deviation for VMI 50 keV ranged from 2 to 9 HU for both 120 and 140 kVp at 10 mGy. As VMI keV decreased, the inter-institutional standard deviation of the inserts followed the same trend as that of the iodine 5 mg/mL, reaching 3 ± 1 and 2.0 ± 0.9 HU for VMI 150 keV at 120 and 140 kVp, respectively.

VMI quantification was preserved across radiation dose between institutions for all inserts. At 140 kVp, errors of the iodine 5 mg/mL insert at VMI 50 keV averaged 17 ± 6 and 14 ± 7 HU across the four institutions for 1 and 10 mGy, respectively (**Figure 2**). The error remained stable between doses as VMI keV increased. For the iodine 5 mg/mL insert, VMI 70 keV errors across institutions measured 6 ± 2 HU while VMI 150 keV errors were -3.9 ± 0.5 and -0.7 ± 1.5 HU for both 1 and 10 mGy, respectively. Similar errors were observed for the blood 70 insert: errors across institutions measured 12 ± 2 and 10 ± 2 HU for VMI 50 keV and -2.2 ± 0.6 and 1 ± 1 HU for VMI 150 keV for 1 and 10 mGy, respectively. Overall, the inter-institutional standard deviation across the inserts averaged 2.1 ± 1.5, 2.1 ± 0.7, 2.0 + 1.1, and 2.0 ± 0.9 HU for 1, 2.5, 5, and 10 mGy, demonstrating the stability of VMIs across the four institutions (**Figure 3**).

Iodine and VNC maps exhibited quantitative stability across institutions for different tube voltages and radiation doses (**Figure 4**). At 10 mGy, the iodine 5 mg/mL insert demonstrated an error of 0.3 ± 0.2 and 0.1 ± 0.2 mg/mL at 120 and 140 kVp, respectively. With decreasing dose, this error measured 0.4 ± 0.2 and 0.4 ± 0.1 mg/mL at 2.5 and 1 mg/mL, respectively. The same minor effect of dose was illustrated in the iodine

quantification for iodine 2 and 10 mg/mL inserts. The corresponding inter-institutional standard deviations for the iodine 2, 5, and 10 mg/mL inserts at 140 kVp and 10 mGy were 0.1, 0.2, and 0.06 mg/mL, respectively. For VNCs, the inter-institutional standard deviations were 3 HU for all three iodine inserts at 140 kVp and 10 mGy.

IV. Discussion

This work highlighted the quantitative stability of PCCT across institutions for different material inserts, spectral results, tube voltages, and dose levels. By utilizing the same phantom model, scanner model, and protocol parameters, scanner-related effects were isolated to extract the associated quantitative variability. With VMIs, inter-institutional standard deviations were greater at keVs less than 70 keV while inter-institutional standard deviations measured less than 4 HU at VMI keVs greater than 70 keV. In comparison, different dose levels and tube voltages had a smaller effect, only modulating the inter-scanner standard deviation by up to 1 HU. The reduced effect of key acquisition parameters, .i.e. tube voltage and dose, across institutions with spectral results further enhances their utility for quantitative imaging applications.

This study has some limitations. While we accounted for known phantom differences by calculating the error relative to the expected value for each institution's phantom, this may not fully capture differences in phantom characteristics resulting from manufacturing. Future studies will include additional scanning of all inserts at a single institution to better extract phantom differences and thereby improving the isolation of scanner-related effects. Additionally, this study only included four institutions. Incorporation of additional institutions may enhance the characterization of inter-institutional quantitative consistency.

V. Conclusion

PCCT demonstrated quantitative consistency across the four sites using the same phantom and scanner model at each tube voltage and radiation dose level. Quantitative values also remained stable across tube voltages and radiation dose, indicating minimal influence of these key protocol parameters. These results build a foundation for multi-institutional clinical studies and support confidence in quantitative measurements across institutions, promoting continued adoption of quantitative imaging. Overall, precise characterization of reproducibility across primary acquisition parameters provides practical guidance to the scientific and clinical community, facilitating the development and clinical adoption of reliable biomarkers for a range of clinical applications.

References

[1] L. Zhu *et al.*, "CT Imaging Biomarkers Predict Clinical Outcomes After Pancreatic Cancer Surgery," *Medicine*, vol. 95, no. 5, 2016, doi: 10.1097/MD.0000000000002664.

[2] P. J. Pickhardt *et al.*, "Automated CT biomarkers for opportunistic prediction of future cardiovascular events and mortality in an asymptomatic screening population: a retrospective cohort study," *Lancet Digit. Health*, vol. 2, no. 4, pp. e192–e200, Apr. 2020, doi: 10.1016/S2589-7500(20)30025-X.

[3] F. Hu *et al.*, "Image harmonization: A review of statistical and deep learning methods for removing batch effects and evaluation metrics for effective harmonization," *Neuroimage*, vol. 274, p. 120125, Jul. 2023, doi: 10.1016/J.NEUROIMAGE.2023.120125.

[4] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293–312, Nov. 2018, doi: 10.1148/RADIOL.2018172656/ASSET/IMAGES/LARGE/RADIOL.2018172656.FIG22.JPEG.

[5] K. Rajendran *et al.*, "First Clinical Photon-counting Detector CT System: Technical Evaluation," *Radiology*, vol. 303, no. 1, pp. 130–138, Dec. 2022, doi: 10.1148/RADIOL.212579.

[6] J. R. Rajagopal *et al.*, "Comparison of Low Dose Performance of Photon-Counting and Energy Integrating CT," *Acad. Radiol.*, vol. 28, no. 12, pp. 1754–1760, Dec. 2021, doi: 10.1016/J.ACRA.2020.07.033.

[7] S. Vrbaski, S. Bache, J. Rajagopal, and E. Samei, "Quantitative performance of photon-counting CT at low dose: Virtual monochromatic imaging and iodine quantification," *Med. Phys.*, vol. 50, no. 9, p. 5421, Sep. 2023, doi: 10.1002/MP.16583.

[8] V. Mergen *et al.*, "Tube voltage-independent coronary calcium scoring on a first-generation dual-source photon-counting CT-a proof-of-principle phantom study," *Int. J. Cardiovasc. Imaging*, vol. 38, no. 4, pp. 905–912, 2021, doi: 10.1007/S10554-021-02466-Y.

[9] L. P. Liu *et al.*, "Assessing the stability of photon-counting CT: insights from a 2-year longitudinal study," *European Radiology 2024*, pp. 1–8, Dec. 2024, doi: 10.1007/S00330-024-11244-Y.

[10] National Institute of Standards and Technology, "XCOM." Accessed: Jan. 12, 2026. [Online]. Available: https://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html

Quantitative Gadolinium K-edge Imaging of Hepatobiliary Excretion Using Clinical PCCT: An In Vivo Study

Martin V. Rybertt[1,2], Leening P. Liu[1], George McClung[1], Pooyan Sahbaee[3], Tristan Nowak[4], Manoj Mathew[1], Peter B. Noël[1], Ali H. Dhanaliwala[1]

***Abstract*—Evaluation of the biliary tree is difficult with computed tomography (CT) due to the small size and poor CT contrast of biliary pathologies such as strictures, gallstones or malignancies and is currently the domain of MRI. Current iodinated intravascular contrast agents enable excellent parenchymal and urinary tract evaluation due to their renal clearance. Hepatobiliary excreted contrast agents are available and provide excellent contrast of the bile ducts, but are gadolinium (Gd) based which limits conspicuity on CT. However, their use in CT has not been evaluated for biliary tree applications. With photon-counting CT (PCCT) Gd based contrast agents (Gd-CA) can be uniquely imaged via K-edge material decomposition, thus improving Gd contrast in CT. By producing gadolinium-specific maps that enhance contrast relative to conventional attenuation images quantitative estimates of gadolinium concentration can be calculated. In this *in vivo* porcine study (n=2), a hepatobiliary Gd-CA (Gadoxetate disodium) was injected intravenously, followed by dynamic PCCT imaging to characterize hepatic arterial and portal venous enhancement and hepatobiliary excretion. Conventional and material-specific images of water, iodine, and gadolinium were generated. The mean attenuation and concentration of Gd-CAs were measured in the biliary system over time. Peak arterial enhancement in the aorta of 110 ± 6 HU and 2.9 ± 0.9 mg/mL occurred 13.2 seconds after injection. Biliary excretion was first detected 10 minutes post-injection. Concentrations plateaued after 15 minutes, averaging 3.3 ± 0.9 mg/mL and 3.0 ± 1.1 mg/mL, respectively. Gadolinium K-edge material maps directly confirmed hepatobiliary clearance, with gadolinium signal localized to the bile ducts and gallbladder over time. These findings support the feasibility of gadolinium-based contrast agents for PCCT imaging of the biliary tree.**

***Index Terms*—Photon-counting CT, K-Edge Imaging, Biliary Imaging, Spectral CT**

I. Introduction

BILIARY tree disease remains a challenge to diagnose with computed tomography (CT). The current gold standard for diagnosis is endoscopic retrograde cholangiopancreatography (ERCP), a moderately invasive technique with risks including duct perforation and pancreatitis [1]. Consequently, non-invasive techniques such as ultrasound (US) and magnetic resonance cholangiopancreatography (MRCP) are preferred first line evaluations. Although CT is a widely used, noninvasive modality, detection of common biliary pathology - isoattenuating gallstones or cholecystitis - can remain challenging [2]. Iodinated contrast agents are often administered to improve lesion conspicuity, but because they are renally cleared, they have limited utility in evaluating the biliary tree. Hepatobiliary gadolinium-based contrast agents (Gd-CAs) are excreted through the biliary system enabling improved visualization of the bile ducts [3]. However, Gd-CAs are predominantly used in magnetic resonance imaging (MRI) as they demonstrate decreased contrast in CT compared to iodine [4]; however, their use in biliary CT has not been explored.

Photon-counting CT (PCCT) represents the next generation of CT imaging, achieving improved spatial resolution, with the potential for reduced radiation dose and advanced spectral capabilities [5], [6]. By measuring the energy of individual photons, PCCT enables advanced material decomposition techniques such as K-edge imaging. This technique leverages the K-edge effect—a sharp increase in attenuation when the incoming photon energy exceeds the K-shell binding energy of a chemical element—to uniquely identify a subset of materials. Conventional material decomposition approaches model the two main X-ray effects, Compton scattering and the photoelectric effect; however, these are unable to differentiate materials with K-edges in the CT diagnostic range (40-150 keV) as these effects cannot describe the increase in attenuation due to the K-edge effect. In contrast, the multi-bin information provided by PCCT facilitates material-specific decomposition of K-edge materials.

Early prototypes of PCCT have already demonstrated the clinical potential of K-edge imaging. Dual-contrast studies can reduce radiation dose and mitigate motion artifacts across sequential acquisitions [7], [8]. K-edge imaging can be used to differentiate between calcifications and endoleaks post-aneurysm [9], or provide targeted imaging of atherosclerotic plaque using gold nanoparticles [10]. Despite these advances, clinical deployment of K-edge imaging remains limited.

We acknowledge support through the National Institutes of Health (R01EB035908) and Siemens Healthineers. M.V. Rybertt is with the Department of Bioengineering and Department of Radiology at the University of Pennsylvania, Philadelphia, PA 19104 USA. L.P. Liu, G. McClung, M. Mathew, P.B. Noël, and A.H. Dhanaliwala are with the Department of Radiology at the University of Pennsylvania, Philadelphia, PA 19104 USA. P. Sahbaee is with Siemens Healthineers, Malvern, PA 19355 USA. T. Nowak is with Siemens Healthineers AG, Forchheim, Germany. (email: mrybertt@seas.upenn.edu)

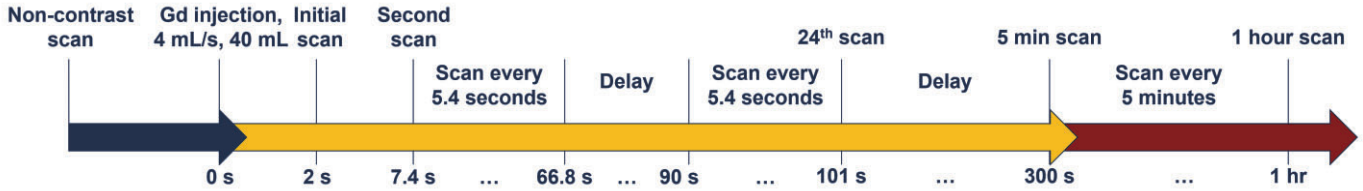


Figure 1. Contrast injection timing and scanning protocol. An injection of a gadolinium-based contrast agent (Gd-CA) was performed at a 4 mL/s rate. Continuous scans on the liver were initiated after a two second delay after injection and repeated 24 times. Scans were then repeated every five minutes for an hour to monitor excretion of Gd-CA through the biliary tree.

Recently, a calibration-based K-edge imaging approach was developed for a clinically available dual-source PCCT [11]. Follow-up work using this approach demonstrated the feasibility of gadolinium K-edge imaging in a controlled phantom setting [12]. To further validate this method, we performed an *in vivo* study using a porcine model to assess a hepatobiliary Gd-CA.

II. Methods

A. Animal model

This study was approved by the local Institutional Animal Care and Use Committee (IACUC). A healthy young Yorkshire-Landrace pig (40 kg) was subjected to the same protocol 48 hours apart to evaluate reproducibility. The pig was sedated using propofol, intubated and anesthetized under isoflurane inhalant throughout the imaging session. A catheter was placed on a right ear vein for intravenous (IV) injection of contrast.

B. Injection and scanning protocol

The imaging session included three sequential stages. First, a full-body non-contrast scan was acquired to establish baseline anatomy for later comparison with post-contrast enhancement. Second, hepatobilliary Gd-CA (0.025 mmol/kg, Eovist, Bayer LLC, Whippany, NJ, United States) was delivered via auricular IV injection (40 mL, 1.0 mL/kg) at a 4 mL/s rate, scans of the liver were initiated 2s after the start of injection and repeated every 5.4s for a total of 24 repetitions to capture the arterial and portal venous phases of Gd-CAs. Third, scans were acquired from the diaphragm to the bottom of the bladder to monitor gadolinium clearance through the biliary and renal systems. These scans were repeated every five minutes, starting at time of injection, for one hour (**Figure 1**). Acquisition parameters were selected by earlier phantom studies. Scans were performed on a clinically approved dual-source PCCT scanner (NAEOTOM Alpha.Peak, Siemens Healthineers, Germany) at a tube voltage of 140 kVp, a volumetric CT dose index ($CTDI_{vol}$) of 12 mGy, and energy bin configuration of 20/55/72/90 keV (**Table 1**).

C. Image reconstruction

CT raw data was exported for all scans and processed using proprietary vendor software (ReconCT, Siemens Healthineers, Forchheim, Germany). Conventional and specialized spectral post-processing (SPP) images were reconstructed with a field of view of 300 mm, square matrix size of 512 px, Qr40 reconstruction kernel, and quantum iterative reconstruction (QIR) level of 3 (**Table 1**). Material-specific water, iodine, and gadolinium maps were reconstructed using a least-squares material decomposition implemented in vendor specific software (eXamine, Siemens Healthineers, Forchheim, Germany), with calibration-derived reference values.

D. Image analysis

Regions of interest (ROIs) were placed in the aorta, liver parenchyma, bile duct, and gallbladder. For each time point, the mean signal and standard deviation (as a noise surrogate) were measured in both conventional attenuation images and the gadolinium material maps. Aortic and hepatic parenchymal time–attenuation and time–concentration curves were used to determine the timing of the hepatic arterial and portal venous phases. Hepatobiliary clearance was assessed by tracking gadolinium signal in the bile ducts and gallbladder qualitatively and quantitatively over time. Scatter plots of mean attenuation and gadolinium signal over time were generated to monitor changes in functional changes in biliary excretion.

Table 1. Scanning and reconstruction parameters

Tube voltage [kVp]	140
$CTDI_{vol}$ [mGy]	12
Bin configuration [keV]	20/55/72/90
Collimation [mm]	96 x 0.4
Rotation time [s]	0.25
Material maps	Water, iodine, gadolinium
Reconstruction kernel/QIR	Qr40/3
Slice thickness [mm]	2
Field of view [mm]	300 x 300
Matrix size [px]	512 x 512

III. Results

Gd-CA enhancement was successfully demonstrated *in vivo* using a clinically available PCCT system. A clear aortic arterial enhancement peak was observed, whereas a distinct portal venous phase was not readily separable under the current acquisition and analysis conditions. Gd-CAs displayed an arterial phase 13.2 seconds after the start of injection with enhancement of 110 ± 6 and 89 ± 7 HU for the first and second pig, respectively, while in gadolinium-specific images concentration measured 2.8 ± 0.9 and 2.3 ± 0.9 mg/mL, respectively. Furthermore, liver parenchyma did not present an increase of more than 15 HU in attenuation or 0.5 mg/mL in concentration compared to non-contrast.

On baseline non-contrast images, no measurable signal was present in the bile duct or gallbladder (**Figure 2A**). Gadolinium excretion into the bile duct first became apparent at 10 minutes post-injection on conventional images (**Figure 2B**) and was corroborated by a concordant increase on the gadolinium material map. By the 30-minute time point

(**Figure 2C**), the full excretory pathway was visualized, with concurrent gadolinium signal in the bile duct, gallbladder, and bowel.

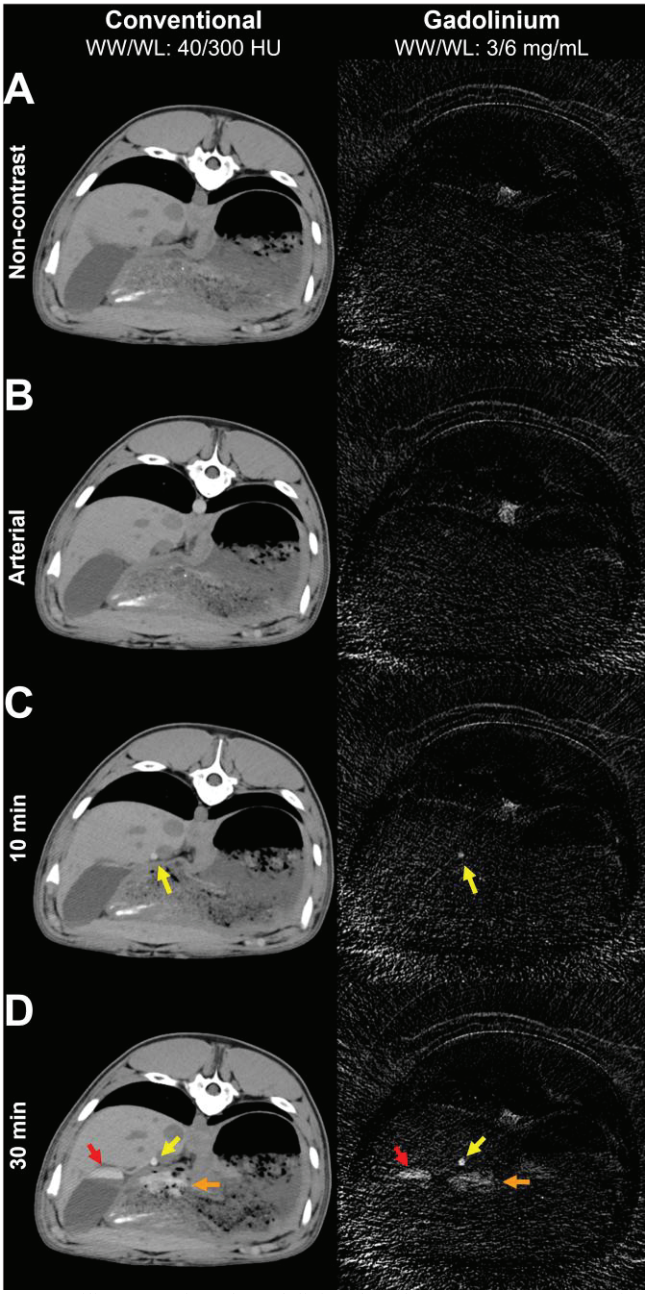


Figure 2. Biliary excretion of a gadolinium-based contrast agent on a porcine model. Non-contrast (A) conventional (left) and gadolinium (right) images displayed no enhancement of the bile ducts. The complete Gd-CA excretory pathway is shown. Gadolinium signal was first measured at the aorta (B) 13.2 seconds after injection. Subsequently, biliary excretion was observed at the 10-minute timepoint (C) as indicated by the yellow arrows. Notably, gadolinium clearance into the gallbladder (red arrow), bowel (orange arrow), and bile duct (yellow arrow) were captured simultaneously at 30 minutes (D).

Gadolinium excretion remained stable following its initial detection at ten minutes after the start of injection. At this time, attenuation within the bile duct measured 100 ± 9 and 98 ± 8 HU for session 1 and 2, respectively, corresponding to gadolinium concentrations of 3.3 ± 0.8 and 2.7 ± 1.0 mg/mL on gadolinium maps (**Figure 3**). By 15 minutes, biliary clearance of the Gd-CA stabilized, with mean bile duct attenuation of 123 ± 10 and 110 ± 10 HU and corresponding concentrations of 3.3 ± 0.9 and 2.5 ± 0.9 mg/mL for session 1 and 2, respectively. Concurrently, gallbladder enhancement stabilized at 104 ± 7 and 87 ± 5 HU, corresponding to gadolinium concentrations of 3.2 ± 0.9 and 2.5 ± 0.9 mg/mL (**Table 2**).

Table 2. Measurements of biliary excretion of Gd

Location	Attenuation [HU]		Gd Concentration [mg/mL]	
	Gd signal 10 minutes post-injection			
	Session 1	Session 2	Session 1	Session 2
Bile ducts	100 ± 9	98 ± 8	3.3 ± 0.8	2.7 ± 1.0
Gallbladder	42 ± 11	51 ± 10	0.7 ± 1.0	2.2 ± 0.8
	Average 15 minutes post-injection			
Bile ducts	123 ± 9	110 ± 10	3.3 ± 0.9	3.0 ± 1.1
Gallbladder	104 ± 7	87 ± 5	3.2 ± 0.9	2.5 ± 0.9

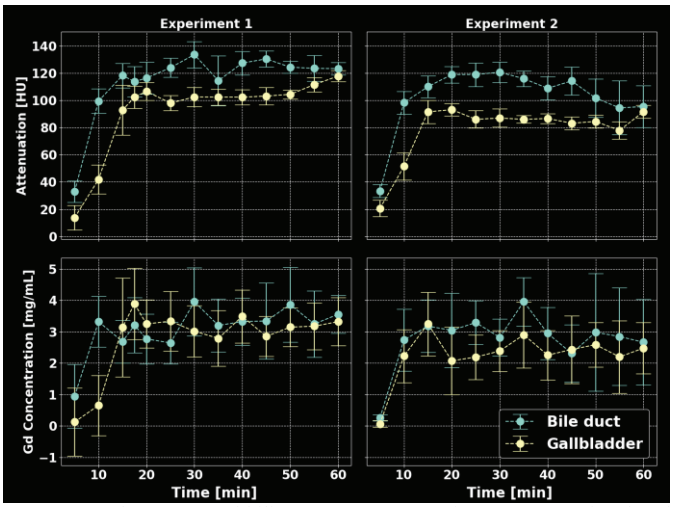


Figure 3. Enhancement of biliary structures over time on conventional and gadolinium images. Conventional images (A) and gadolinium maps (B) display an increase in gadolinium concentration in the bile ducts and gallbladder ten minutes after the start of injection. Error bars describe the standard deviation of the region of interest.

IV. Conclusion

PCCT imaging of a hepatobiliary Gd-CA in an *in vivo* porcine model demonstrated appropriate hepatobiliary excretion, with gadolinium signal correctly localized to the biliary tree. A clear hepatic arterial enhancement peak was identified, whereas a distinct portal venous phase was not readily separable under the current acquisition and analysis conditions such as contrast volume. PCCT with gadolinium K-edge material mapping enabled visualization of the excretory pathway beyond the liver, including the bile duct, gallbladder, and bowel. Quantitatively, bile duct and gallbladder gadolinium concentrations reached a stable plateau approximately 15 minutes post-injection, supporting the feasibility of extracting functional biomarkers of biliary excretion (e.g., time-to-appearance, time-to-plateau, and plateau concentration) or collecting diagnostic information such as lack of reflux into the gallbladder due to cholecystitis.

Collectively, these findings support hepatobiliary gadolinium-based contrast agents as a viable alternative to iodinated agents for PCCT-based biliary imaging. With

further optimization of dosing and acquisition timing, gadolinium-enhanced PCCT has the potential to provide combined anatomical depiction and quantitative functional assessment of the biliary system via K-edge imaging.

REFERENCES

[1] K. M. Sundaram, M. A. Morgan, M. Itani, and W. Thompson, "Imaging of benign biliary pathologies," *Abdom. Radiol.*, vol. 48, no. 1, pp. 106–126, Jan. 2023, doi: 10.1007/s00261-022-03440-5.

[2] J. S. Shakespear, A. M. Shaaban, and M. Rezvani, "CT Findings of Acute Cholecystitis and Its Complications," *Am. J. Roentgenol.*, vol. 194, no. 6, pp. 1523–1529, Jun. 2010, doi: 10.2214/AJR.09.3640.

[3] C. L. Vendrami *et al.*, "Imaging of Biliary Tree Abnormalities," *RadioGraphics*, Jul. 2024, doi: 10.1148/rg.230174.

[4] J. W. Henson, R. G. Nogueira, D. J. Covarrubias, R. G. Gonzalez, and M. H. Lev, "Gadolinium-Enhanced CT Angiography of the Circle of Willis and Neck," *Am. J. Neuroradiol.*, vol. 25, no. 6, pp. 969–972, Jun. 2004.

[5] T. Flohr, M. Petersilka, A. Henning, S. Ulzheimer, J. Ferda, and B. Schmidt, "Photon-counting CT review," *Phys. Med.*, vol. 79, pp. 126–136, Nov. 2020, doi: 10.1016/j.ejmp.2020.10.030.

[6] C. H. McCollough, S. Leng, L. Yu, and J. G. Fletcher, "Dual- and Multi-Energy CT: Principles, Technical Approaches, and Clinical Applications," *Radiology*, vol. 276, no. 3, pp. 637–653, Sep. 2015, doi: 10.1148/radiol.2015142631.

[7] S. Si-Mohamed *et al.*, "Spectral Photon-Counting Computed Tomography (SPCCT): in-vivo single-acquisition multi-phase liver imaging with a dual contrast agent protocol," *Sci. Rep.*, vol. 9, no. 1, p. 8458, Jun. 2019, doi: 10.1038/s41598-019-44821-z.

[8] R. Symons *et al.*, "Photon-counting CT for simultaneous imaging of multiple contrast agents in the abdomen: An in vivo study," *Med. Phys.*, vol. 44, no. 10, pp. 5120–5127, Oct. 2017, doi: 10.1002/mp.12301.

[9] J. Dangelmaier *et al.*, "Experimental feasibility of spectral photon-counting computed tomography with two contrast agents for the detection of endoleaks following endovascular aortic repair," *Eur. Radiol.*, vol. 28, no. 8, p. 3318, Feb. 2018, doi: 10.1007/s00330-017-5252-7.

[10] D. P. Cormode *et al.*, "Atherosclerotic Plaque Composition: Analysis with Multicolor CT and Targeted Gold Nanoparticles," *Radiology*, vol. 256, no. 3, pp. 774–782, Sep. 2010, doi: 10.1148/radiol.10092473.

[11] J. Neumann, T. Nowak, B. Schmidt, and J. von Zanthier, "An Image-Based Prior Knowledge-Free Approach for a Multi-Material Decomposition in Photon-Counting Computed Tomography," *Diagnostics*, vol. 14, no. 12, Art. no. 12, Jan. 2024, doi: 10.3390/diagnostics14121262.

[12] M. V. Rybertt *et al.*, "K-Edge Imaging Using a Clinical Dual-Source Photon-Counting CT System," Aug. 24, 2025, *medRxiv*. doi: 10.1101/2025.08.21.25333798.

Water Beam Hardening and Pulse Pileup Correction with Small Neural Networks

Jannik Dettmer*†, Julien Erath*, Matthias Baer-Beck*, Juliane Szkitsak†‡, Karl Stierstorfer*, Christoph Bert†‡

***Abstract*—Due to pulse pileup in photon counting CT and water beam hardening, there is a non-linear relation between measured x-ray attenuation and object path length. This leads to artifacts in reconstructed images, e.g. cupping artifacts. A common approach to correct for non-linearities in the measured data are polynomials, whose optimal coefficients are determined through optimization algorithms. Although they provide good results, issues like limited generalization remain. This work investigates neural networks as an alternative to perform these corrections, with the goal of determining whether neural networks can handle the corrections, generalize well, and potentially combine multiple corrections within a single model. As of now, training and validation are performed solely on simulated attenuation data. We found that even very small neural networks with only one hidden layer and a small number of neurons can perform water beam-hardening correction and pulse-pileup correction effectively. The network also successfully includes bowtie correction. Their ability to generalize to unseen data is promising but requires further optimization. Using neural networks for these types of corrections could streamline current pre-correction pipelines while offering greater flexibility and potential for integrating additional correction types.**

***Index Terms*—Computer Tomography, Water Beam Hardening Correction, Non-Linearity Correction, Photon Counting, Neural Network**

I. Introduction

CT imaging requires a linear relationship between the measured attenuation $q = \ln\frac{N}{N_0}$ and the line integral to reconstruct artifact free images:

$$-q = \int d\lambda\, \mu(\vec{s} + \lambda\vec{\Theta}, E) \tag{1}$$

where N_0 is the number of photons emitted by the x-ray source, N is the number of photons detected and $\mu(\vec{s}+\lambda\vec{\Theta}, E)$ is the attenuation coefficient dependent on the photon energy and the position along an x-ray path $\vec{s}+\lambda\vec{\Theta}$, with $\vec{s}$ the origin of the ray, $\vec{\Theta}$ as direction and λ the position parameter.
In theory, this linear relation is achievable, if the x-ray source emits only photons of a single energy E and the detector response is perfectly linear. Neither assumption holds in real CT systems. X-ray sources emit broadband spectra, and detectors - especially photon-counting detectors - exhibit non-linear responses at high count rates due to pulse pileup (PP). [1].

* Siemens Healthineers, Forchheim, Germany
† Friedrich-Alexander-University Erlangen-Nuernberg
‡ University Hospital Erlangen, Radiation Clinic

A. Water Beam Hardening

The non-linearity introduced by a polychromatic spectrum leads to the following change of Equation 1:

$$q = \ln \int dE\; w(E)\; e^{-\int d\lambda \mu(\vec{s}+\lambda\vec{\Theta}, E)} \tag{2}$$

where $w(E)$ is the normalized spectrum. Because different energies are attenuated differently, low-energy photons are attenuated more strongly than high-energy photons, causing spectral beam hardening. If uncorrected, this results in image artifacts such as cupping (see Figure 5 a)) [2] and shadows behind dense materials like bone. First order corrections typically target water, as it is the dominant material in the human body, hence water beam hardening correction (WBHC). Higher order, usually iterative, corrections are required for materials like bone [3]. This work focuses solely on WBHC. To perform WBHC, the goal is to find a function P that maps the measured attenuation q to an equivalent line integral at a monochromatic reference energy E:

$$P(q, b) \stackrel{!}{=} \int d\lambda\; \mu(\vec{s} + \lambda\vec{\Theta}, E) \tag{3}$$

Polynomials with optimized coefficients are commonly used to model P [4] [5] [6] [7]. This type of WBHC has been very successful but comes with some problems, e.g. the polynomials used for correction generalize poorly [8] and overfitting.
In this work, we instead train small neural networks for WBHC and evaluate them using simulated CT data.

B. Non-Linear Detector Response / Pulse Pileup

Another reason for a non-linear relation between measured attenuation data and the line integral is the non-linear response of photon-counting detectors. These detectors count individual incoming photons, but if two photons arrive within too short a time window, the detector may interpret them as one photon of higher energy (PP). The probability of this effect increases with higher photon rates (larger x-ray tube currents). Adding this effect to Equation 2 results in:

$$q = \ln \int dE\; w(E)\; D(E, n(E))\; e^{-\int d\lambda \mu(\vec{s}+\lambda\vec{\Theta}, E)} \tag{4}$$

where $D(E, n(E))$ describes the non-linear detector response, depending on the photon energy E and the initial photon rate $n(E)$. In reality, the detector response also depends on the path through the phantom, since the spectrum is modified as it traverses the material, but this is challenging to express analytically.

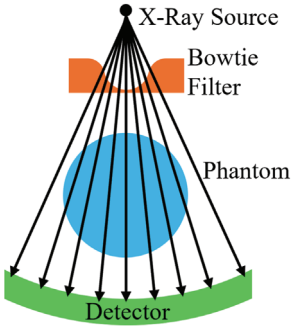


Fig. 1. Experimental setup in the simulation software.

Commonly, WBHC and pulse pileup correction (PPC) are performed in separate steps. However, due to the flexibility and representational capacity of neural networks, this work aims to combine both corrections within a single neural network.

C. Bowtie Filter

Typically, CT systems are equipped with a bowtie filter that optimizes the pre-patient x-ray intensity distribution according to a typical patient shape. The approximate bowtie profile is shown in Figure 1 in orange. These filters pre-attenuate the x-rays depending on their fan angle, compensating for the fact that x-rays passing through a patient's center undergo stronger attenuation than those passing through the periphery (see Figure 1). This is done to decrease the dose in the peripheral regions of the body and yields a more uniformly distributed beam intensity across all ray paths.
For image reconstruction, the bowtie attenuation must be removed from the measured detector data. However, a simple subtraction is insufficient (although it would be valid for a monochromatic spectrum).
Depending on the fan angle, the bowtie-induced pre-filtration modifies the spectrum differently. Therefore, the beam hardening correction must also account for the bowtie filter attenuation. A neural network is well suited for this task, as it can naturally incorporate the bowtie's influence.

To jointly correct for water beam hardening, pulse pileup and the bowtie filter, the correction function P must depend not only on the measured attenuation q, but also on the photon rate n and the bowtie attenuation b:

$$P(q,n,b) \stackrel{!}{=} \int d\lambda \; \mu(\vec{s}+\lambda\vec{\Theta},E) \tag{5}$$

II. Methods

This work is completely based on simulations. Figure 1 shows the experimental setup for the model in the simulation. Following the direction of the ray path, the setup consists of: an x-ray source, a bowtie filter (in some cases), water phantoms of varying size, and a detector with the geometry of the Siemens Healthineers NAEOTOM® Alpha CT-scanner.

A. Training Data

Using the Deterministic Radiological Simulation (DRASIM) software from Siemens Healthineers [9], the attenuation of the x-rays from the source is calculated deterministically by computing path lengths through objects. The attenuation can be obtained either for a polychromatic spectrum (spectrum of the Siemens Healthineers NAEOTOM® Alpha CT-scanner, here always at 120 kV) or a single photon energy (monochromatic, here always at 60.67 keV (mean of the 120 kV spectrum)).
The monochromatic attenuation without the bowtie filter is always used as the target output of the neural network. The inputs to the neural network depend on the type of corrections the network is supposed to learn. The three cases are:

1) WBHC only
2) WBHC and bowtie filter correction
3) WBHC, bowtie filter correction and PPC

Case 1: The network has only one input, which is the polychromatic attenuation of the phantom (no bowtie filter).
Case 2: The network receives two inputs, the combined attenuation of the bowtie filter and the phantom and as second input the bowtie attenuation only.
Case 3: The network receives three inputs, the combined attenuation of the bowtie filter and the phantom (with PP non-linearity), as second input the bowtie attenuation only (with PP non-linearity) and as third input a parameter for the photon rate or tube current.

For training, projections of four different size water cylinders are used with diameters: 10 cm, 20 cm, 30 cm and 40 cm. For each phantom size, one projection of the phantom centered at the isocenter and 18 projections (uniformly distributed over 360°) with the phantoms moved off-center by 6 cm in vertical direction are used.
Figure 2 shows part of the training data: the attenuation from the center detector row for one projection of the 40 cm water cylinder. The attenuation is plotted against the path length through the phantom. The black curve represents the monochromatic data, the ideal case in which no correction is necessary. Here, the relation between acquired data and phantom path length is linear and can be used for artifact free reconstruction. The cyan curve shows the polychromatic attenuation data of the phantom (no bowtie filter). For small path lengths its attenuation is larger than the monochromatic attenuation and for large path lengths, it is smaller than the monochromatic attenuation. Here only beam hardening leads to a non-linear relationship between path length and data.
The red curve shows the combined polychromatic attenuation of the bowtie and the water phantom. Because of the additional attenuation of the bowtie, this curve lies above the black line (monochromatic attenuation).
The blue and purple curves include non-linearity due to pulse pileup. The behavior is not fully identical to that of a real photon-counting detector but follows a simplified model to evaluate whether the neural network can learn and correct this effect. The non-linearity is introduced using the following expression (based on the non-paralyzable counter $n = \frac{m}{1+m\tau}$

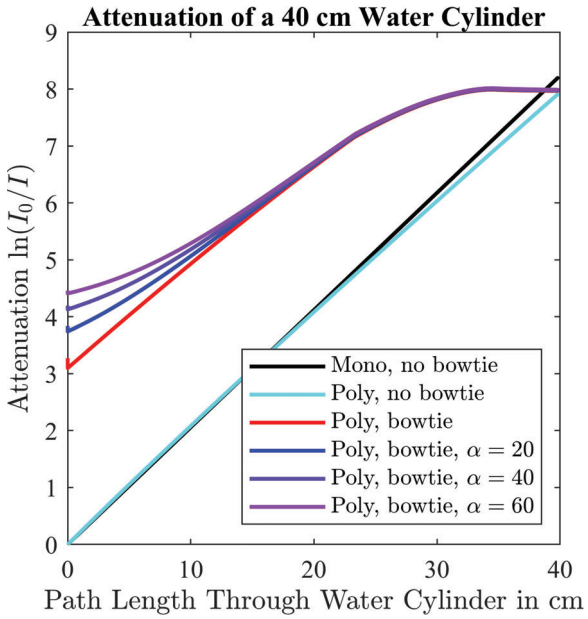


Fig. 2. Attenuation from one projection of monochromatic x-rays of 60.67 keV (black) and a polychromatic spectrum at 120 kV by a 40 cm cylindrical water phantom over the path length traveled in the phantom. For some projections a bowtie filter was part of the simulation (red, blue/purple). For the blue/purple curves an additional non-linearity is introduced which follows $f(x) = x + \ln(1 + \alpha \cdot e^{-x})$. This is a model for the non-linearity introduced in photon counting systems due to pulse pileup, which depends on the tube current. α represents the degree of freedom of the tube current.

[10], where n is the true countrate, m is the measured countrate and τ is the recovery time of the detector):

$$X' = X + \ln(1 + \alpha \cdot e^{-X}) \tag{6}$$

where X is the attenuation data and α controls the degree of non-linearity. α can be interpreted as a tube-current parameter. During training, 10 different tube-current settings are used: $\alpha = 0.1,\ 0.2,\ ...,\ 0.9,\ 1.$

In essence, the goal of the neural network is to transform the input data from case 1, 2 or 3 such that resulting corrected curves match the monochromatic (black) curve exactly.

B. Network Architecture

The general structure of the neural networks is always the same. There is a single output and, depending on the correction case, either one, two or three inputs. The architecture is illustrated for a single layer for case 3 in Figure 3.

We investigated various architectures for the hidden layers, ranging from one up to three hidden layers. For each layer, neuron counts between 1 and 12 were examined closely, as well as some larger neuron numbers (50, 100, 200 neurons per layer).

The network is fully connected, using softplus activations:

$$f(x) = \ln(1 + e^{x}) \tag{7}$$

in all hidden layers. Only the output layer uses a linear activation.

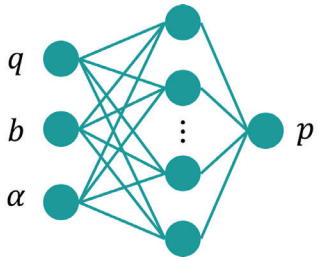


Fig. 3. Schematic of the network architecture for the case with three inputs: the combined polychromatic attenuation of the bowtie and the phantom q with PP, the polychromatic attenuation of the bowtie only b with PP and the current parameter for the pulse pileup non-linearity α. The one output p is the corrected attenuation data.

C. Training Procedure

Training is performed through stochastic gradient descent using JAX [11] with Adam [12] as optimizer and a starting learning rate of 0.001. A batch size of 10 is used. Early stopping is implemented with patience of 1000 epochs. The standard mean square error (MSE) loss function is used.

III. Results

A. Network Architecture

We evaluated different activation functions. For WBHC and PPC, the softplus activation function (see Equation 7) proved to be highly suitable. It behaves almost linearly for large values (good generalization) and is fully differentiable. We also tested ReLu, but the non-differential point at zero required significantly more neurons to avoid visible kinks in the corrected attenuation curves, making it less efficient for our purposes.

We found that for case 1 (only WBHC), a single hidden layer with seven neurons provided good results while keeping the network architecture simple. For case 2 (WBHC and bowtie correction) a single hidden layer with eleven neurons offered the best balance between simplicity and accuracy. For case 3 (WBHC, bowtie correction and PPC), we likewise used the one-layer eleven neuron architecture, however, no further architecture optimization has been conducted yet, and will be addressed in future work. The networks converge quickly, and training takes only seconds to minutes, depending on the case.

B. Corrections

The results obtained with these networks are very promising. Figure 4 shows the difference of the corrected attenuation and the monochromatic attenuation for one projection of a 50 cm water cylinder. As in the previous section, the three correction cases are shown: correction of polychromatic attenuation without a bowtie filter (cyan), correction of polychromatic attenuation with a bowtie filter (red) and correction of polychromatic attenuation with a bowtie filter and PP non-linearity (blue and purple curves). The deviation remains very small up to path lengths of 40 cm, which corresponds to the maximum path length included in the training data. In all cases, except case 2 (red), the curves start to deviate strongly from the monochromatic attenuation at this point. Interestingly

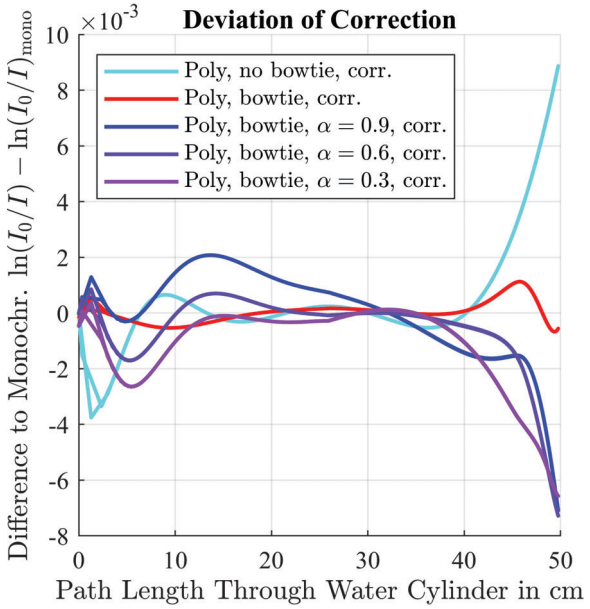


Fig. 4. Each trace shows the difference of the corrected data to the monochromatic data. For the cases: correction of polychromatic attenuation without bowtie (cyan), polychromatic attenuation with bowtie (red) and polychromatic attenuation with bowtie and pulse pileup (blue and purple hues). The neural network responsible for correction was trained only on water path lengths of 40 cm and below.

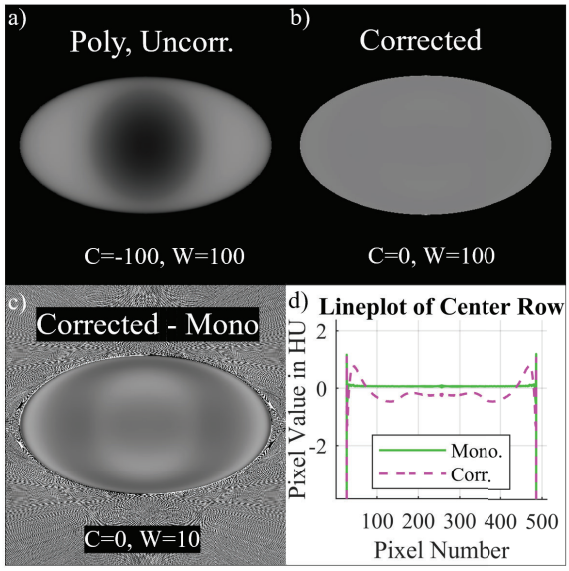


Fig. 5. Reconstructions of a 25 cm/45 cm elliptical water cylinder: a) reconstruction of the uncorrected polychromatic attenuation with bowtie filter and pulse pileup non-linearity (bowtie attenuation subtracted before reconstruction), b) reconstruction of the data corrected by the neural network, c) difference between the reconstruction from the corrected data and reconstruction of the monochromatic data (without bowtie filter and without pulse pileup non-linearity), d) line plot of the center row of the reconstructed image of the monochromatic data and the corrected data.

the red curve continues to oscillate around the zero line, and therefore generalizes well to unseen data.

Overall, the deviation in percent remains mostly below 0.25%.

Figure 5 shows the reconstruction of the data corresponding to the blue curves, so the combined attenuation of the phantom and the bowtie with strong PP non-linearity ($\alpha = 0.9$). The phantom is an elliptical water cylinder with axis lengths 25 cm and 45 cm. The attenuation data thus includes path lengths exceeding those in training.

Figure 5 a) shows the reconstruction of the uncorrected data (bowtie attenuation subtracted prior to reconstruction). The combined water beam hardening and bowtie-induced beam hardening lead to a darkening toward the phantom center. The window center is at −100 HU and the width is 100 HU. Figure 5 b) shows the reconstruction of the corrected data (network with three inputs, one hidden layer, 11 neurons). The window center is at 0 HU and the window width is at 100 HU. No visible artifacts remain. Figure 5 c) shows the difference between the corrected reconstruction and the monochromatic reconstruction (ground truth). The window center is at 0 HU and the window width is at 10 HU. A small residual deviation remains, with a maximum absolute error of only 0.6 HU, Figure 5 d) shows a line plot through the central row of the corrected reconstruction (magenta) and the monochromatic reference (green).

IV. Discussion

Based on these findings, the correction of water beam hardening, bowtie filter effects, and pulse pileup using small neural networks is feasible for simulated data. The networks successfully learn the correction functions and exhibit a certain degree of generalization. Improving this generalization will be part of future research. Additionally, extending the approach to other corrections, such as scatter, and incorporating detector pixel-dependent corrections are planned future steps.

A key next step is applying the neural networks to real measurement data from a CT scanner and developing methods to train the models directly on measured data by using scans of a set of calibration phantoms with know geometry and material decomposition.

References

[1] S. Usman and A. Patil, "Radiation detector deadtime and pile up: A review of the status of science," *Nuclear Engineering and Technology*, vol. 50, no. 7, pp. 1006–1016, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1738573318302596

[2] J. F. Barrett and N. Keat, "Artifacts in ct: Recognition and avoidance," *RadioGraphics*, vol. 24, no. 6, pp. 1679–1691, 2004, pMID: 15537976. [Online]. Available: https://doi.org/10.1148/rg.246045065

[3] Y. Zhao and M. Li, "Iterative beam hardening correction for multi-material objects," *PloS one*, vol. 10, no. 12, p. e0144607, 2015.

[4] G. T. Herman, "Correction for beam hardening in computed tomography," *Physics in Medicine & Biology*, vol. 24, no. 1, p. 81, 1979.

[5] A. Coleman and M. Sinclair, "A beam-hardening correction using dual-energy computed tomography," *Physics in medicine & biology*, vol. 30, no. 11, p. 1251, 1985.

[6] M. Kachelrieß, K. Sourbelle, and W. A. Kalender, "Empirical cupping correction: a first-order raw data precorrection for cone-beam computed tomography," *Medical physics*, vol. 33, no. 5, pp. 1269–1274, 2006.

[7] Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelrieß, "Empirical beam hardening correction (ebhc) for ct," *Medical Physics*, vol. 37, no. 10, pp. 5179–5187, 2010. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3477088

[8] C. Martinez, J. A. Fessler, M. Desco, and M. Abella, "Simple beam hardening correction method (2dcalbh) based on 2d linearization," *Physics in Medicine & Biology*, vol. 67, no. 11, p. 115005, 2022.

[9] S. Kappler, D. Niederlöhner, S. Wirth, and K. Stierstorfer, "A full-system simulation chain for computed tomography scanners," in *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*. IEEE, 2009, pp. 3433–3436.

[10] R. D. Evans and R. Evans, *The atomic nucleus*. McGraw-Hill New York, 1955, vol. 582.
[11] R. Frostig, M. J. Johnson, and C. Leary, "Compiling Machine Learning Programs via High-Level Tracing," in *SysML Conference 2018*, Stanford, United States, Mar. 2019. [Online]. Available: https://hal.science/hal-05188750
[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017. [Online]. Available: https://arxiv.org/abs/1412.6980

Comparison of SIRT and FBP Reconstruction Methods for Material Characterization in Photon-Counting CT: a Monte Carlo Study

Gustavo F. P. da Silva, Renan F. dos Santos, Alessandra Tomal, Gisell R. Boiset and Paulo R. Costa

Abstract— Photon-counting computed tomography exhibits the capability of considering spectral information for data analysis and processing. This capability enables the quantitative characterization of materials in an image. For this purpose, the impact of reconstruction algorithms was evaluated by comparing Filtered Backprojection with Simultaneous Iterative Reconstruction Technique to estimate electron density relative to water (ρ_e) and effective atomic number (Z_{eff}). The analysis used dual-energy CT images obtained from Monte Carlo simulations of a tabletop PCCT system, via GEANT4. To estimate these quantities from the CT images, the DEEDZ methodology was applied, which incorporates linear fitting for ρ_e and Z_{eff} estimation. Two phantoms containing water and embedded with material inserts were simulated, and the nominal values were obtained by weighting the chemical composition. Using the SIRT algorithm, an improvement of 0.55% in the R^2 of the linear fit was observed for ρ_e, while no significant difference was found for Z_{eff}. The mean absolute percentage error (MAPE) in material characterization decreased by 23% and 25% for ρ_e and Z_{eff}, respectively. In conclusion, when applied within the DEEDZ methodology, the SIRT algorithm represents a consistent improvement in material characterization compared with FBP, as evaluated by the MAPE metric.

***Index Terms*—: Effective Atomic Number, Electron Density, PCCT, Monte Carlo Simulation, Image Reconstruction.**

I. Introduction

Photon-counting detectors (PCDs) have the capability of accounting individual photon energies, allowing the acquisition of spectral images. They can be used for multi-energy separation in computed tomography image reconstruction (PCCT) [1, 2]. For material characterization, physical quantities such as electron density relative to water (ρ_e) and effective atomic number (Z_{eff}) play important roles and are essential for dose calculations [3]. Methods for estimating these quantities using PCCT have been proposed in the literature [4, 5].

This study aimed to evaluate the improvements offered by the Simultaneous Iterative Reconstruction Technique (SIRT) over Filtered Backprojection (FBP), utilizing raw data from an ideal simulated PCCT system. The methodology developed by Saito and Sagara [6, 7], was applied to determine the ρ_e, and Z_{eff} of known materials through Monte Carlo simulations, based on GEANT4 [8- 11].

II. Methods

A. PCCT Simulation

A PCCT system was simulated using GEANT4 toolkit (version 11.1.2). The experimental setup is shown in Fig. 1 and consisted of a tungsten-anode X-ray linear source, emitting parallel beams, with a focal spot-to-detector distance of 1045 mm. A sample was positioned 225 mm from the focal spot. The transmitted photons were collected by a virtual 13 mm^2 effective area PCD with ideal efficiency. This PCD produced signal matrices of 128 × 128 elements. Simulated sinograms were acquired considering sample rotation from 0° to 180°, in 1° step. Each projection assumed that 3 million primary photons were generated by the X-ray source per angle. An experimental 50 kV tungsten target X-ray spectrum measured using a MCN 421 X-ray system (Philips Inc., Netherlands) and a XR-100T CdTe spectrometer (Amptek Inc., USA) was incorporated into the simulation [12].

Two cylindrical water phantoms with a 6 mm radius were simulated as samples, Fig. 2. One phantom was used for calibration and contained 1 mm radius inserts of polypropylene, nylon 6/6, silicon dioxide, polycarbonate, and Teflon. The second phantom was used for testing, containing polyethylene and paraffin inserts with radii of 1.5 mm and 2.5 mm, respectively, to evaluate materials with different contrast properties.

Manuscript received January 27, 2026. This work was supported in part by the São Paulo Research Foundation (FAPESP), Brazil (grant numbers 2018/05982–0, 2020/04867-2, 2021/14688-0, 2022/11457-0, 2023/03535-4, and 2023/13993-0), by the National Council for Scientific and Technological Development (CNPq), Brazil (fellowships 302986/2023-5, 311657/2021-4, 313600/2021-0, and 313342/2025-3), and by the University of Campinas Research, Teaching and Extension Support Fund (FAEPEX) (grant 2581/23).

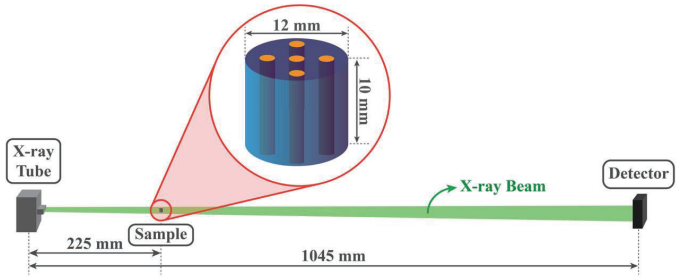
Gustavo F. P. da Silva is with Institute of Physics of the University of São Paulo, São Paulo, SP, 05508-090, Brazil (e-mail: freiregustavo@usp.br).

Renan F. dos Santos is with Institute of Physics of the University of São Paulo, São Paulo, SP, 05508-090, Brazil (e-mail: rsantos@if.usp.br).

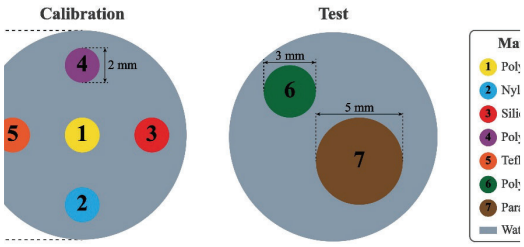
Alessandra Tomal is with Institute of Physics Gleb Wataghin, University of Campinas, Campinas, SP, 13083859, Brazil (e-mail: atomal@ifi.unicamp.br).

Gisell R. Boiset was with Institute of Physics of the University of São Paulo, São Paulo, SP, 05508-090, Brazil. She is now with Institute of Physics, Federal University of Rio de Janeiro, Rio de Janeiro, RJ, 21941909, Brazil (e-mail: gruizboiset@gmail.com).

Paulo R. Costa is with is with Institute of Physics of the University of São Paulo, São Paulo, SP, 05508-090, Brazil (e-mail: pcosta@if.usp.br).



tube emitting parallel beams. Center: sample rotating from 0° to 180° in 1°



polypropylene, nylon 6/6, silicon dioxide, polycarbonate, and Teflon, and

B. *Image Reconstruction*

From the projection data, sinograms were generated and two image reconstruction algorithms were implemented in a Python environment. Filtered Backprojection (FBP) reconstructions were obtained using the scikit-image library [13], implementing the inverse Radon transform with a Hamming filter [14]. The Simultaneous Iterative Reconstruction Technique (SIRT) was implemented using the TomoPy toolkit [15], with a FBP reconstruction as the initial image and multiple iterations performed for each dataset. For each algorithm, a rectangular region of interest (ROI), with size 25×25 pixels, was defined in a homogeneous section of the image, and the signal-to-noise ratio (SNR) was measured according to the equation (1)

$$\mathrm{SNR} = \frac{\mu}{\sigma}, \tag{1}$$

where μ and σ are the signal mean and standard deviation, respectively [14]. Although SIRT is an iterative method, no explicit regularization term was employed, as the optimal number of iterations was selected to maximize the SNR, thereby avoiding noise overfitting. Since SIRT was initialized with the FBP image, maximizing the SNR corresponds to selecting reconstructions with a signal-to-noise ratio equal to or greater than that of the initial image.

C. *Material Characterization*

Equations (2) and (3) present the formulas used to estimate the electron densities and the effective atomic numbers of known materials using their chemical composition [16]:

$$\rho_e = \frac{\rho N_A \sum_i \omega_i \frac{Z_i}{A_i}}{\rho_e^W}, \tag{2}$$

and

$$Z_{\mathrm{eff}} = \left(\frac{\sum_i \omega_i \frac{Z_i}{A_i} Z_i^m}{\sum_i \omega_i \frac{Z_i}{A_i}} \right)^{1/m}. \tag{3}$$

In these equations, ω_i, A_i and Z_i correspond to the elemental mass fraction, atomic mass, and atomic number of the i-th element of the material, respectively. ρ denotes the mass density, ρ_e^W the electron density of water and $m = 3.3$ is a numerical parameter that represents the contribution of Compton scattering and the photoelectric effect to the total photon attenuation [4].

Using the energy discrimination capability of the PCD, the 50 kV spectrum was split into two energy ranges, generating images for the low- and high-energy bins, enabling dual-energy image analysis. Saito and Sagara [6] proposed a methodology for ρ_e and Z_{eff} estimation in dual-energy CT, which adopts the quantity $\mathrm{DHU} = (\mathbf{1} + \alpha)\mathrm{CTN}_{\mathrm{high}} - \alpha\mathrm{CTN}_{\mathrm{low}}$, where $\mathrm{CTN}_{\mathrm{low}}$ and $\mathrm{CTN}_{\mathrm{high}}$, are mean CT numbers of a region of interest measured in the target material low and high energy images, respectively, and α will be defined in the next paragraph. DHU is related to ρ_e by the linear relation expressed by

$$\rho_e = a\left(\frac{DHU}{1000}\right) + b. \tag{4}$$

In equation (4) a and b are fit parameters and α a weighting factor, numerically estimated to maximize the R^2 value of the fit. Z_{eff} is determined using the equation

$$\left(\frac{Z_{\mathrm{eff}}}{Z_{\mathrm{eff,W}}}\right)^m - 1 = \gamma_{low}\left(\frac{\eta_{low}}{\rho_e} - 1\right). \tag{5}$$

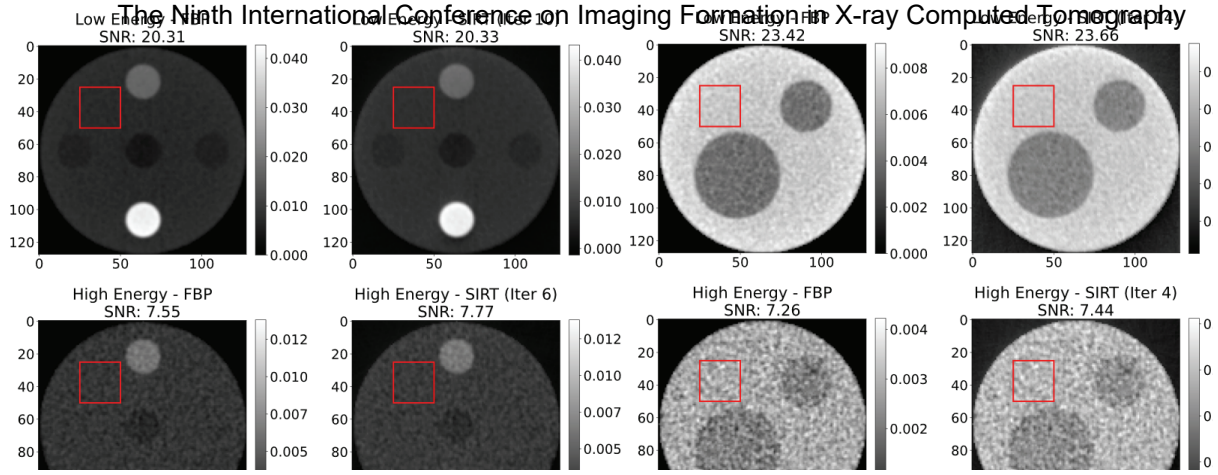
In this equation, the reduced CT number (η_{low}) for the low energy image is modeled as

$$\eta_{low} = \frac{CTN_{low}}{1000} + 1. \tag{6}$$

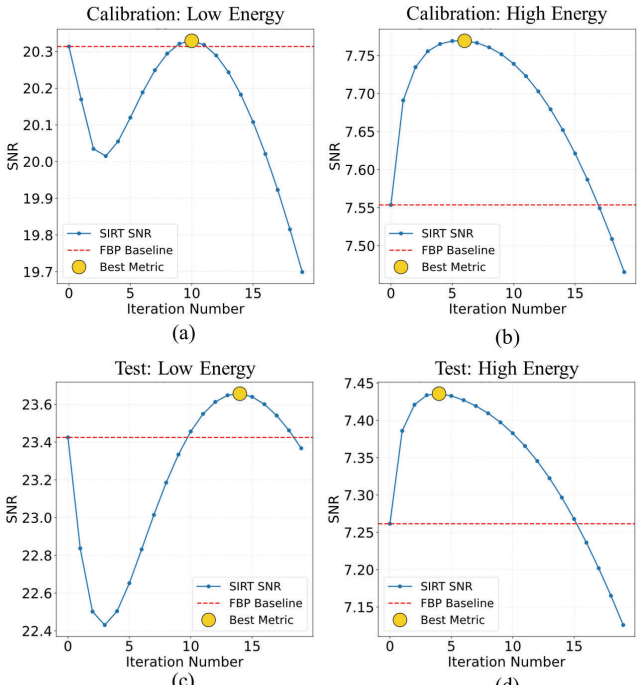
Finally, γ_{low} is a proportionality parameter, and $Z_{\mathrm{eff,W}}$ is the effective atomic number of the water. A linear fit between the left- and right-hand sides of the equation can be performed, estimating γ_{low} as the slope and allowing the intercept to vary as a free parameter. m was numerically determined to maximize the R^2 value of the fit.

III. Results

The projection data from the simulations were reconstructed by applying the FBP and SIRT algorithms for the calibration (left) and test (right) phantoms, shown in Fig. 3. An increase in SNR is observed for each image when using the iterative method. For the test phantom, a relative



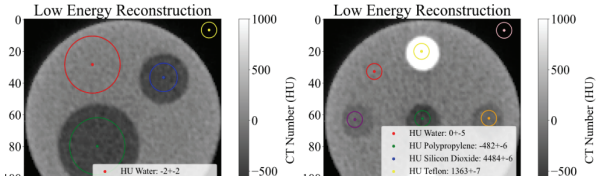
smoothing of the noise can be qualitatively observed. The determination of the optimal number of iterations is shown in Fig. 4 for each phantom; the SNR increases, reaches a maximum, and then decreases. Thus, the optimal number of iterations was selected by maximizing the SNR metric.



Evolution of the signal-to-noise ratio as a function of the num ions for the calibration phantom in (a) low-energy and (b) images; and for the test phantom in (c) low-energy and (d) images.

Usi ration was carried out, and circular ROIs were established for each material in each phantom, as illustrated for the low-energy band image in Fig. 5. For the calibration phantom, linear fits corresponding to equations (4) and (5) were compared for both reconstruction algorithms, as shown in Fig. 6. Using the SIRT algorithm, a 0.55% improvement in R^2 values of the fits for ρ_e and a decrease of 0.01% for Z_{eff} were achieved, when compared to FBP.

Applying this linear fit to the low- and high-energy images of the test phantom on a pixel-by-pixel basis, maps of ρ_e and Z_{eff} can be obtained (Fig. 7). By defining ROIs for each material, the ρ_e and Z_{eff} values were calculated and compared with the nominal values using the mean absolute percentage error (MAPE) metric, as presented in Table I. For both physical quantities and materials, the SIRT algorithm showed lower MAPE values, and an evaluation of the average MAPE shows a reduction of 23% for ρ_e and 25% for Z_{eff} estimation compared to the FBP algorithm.

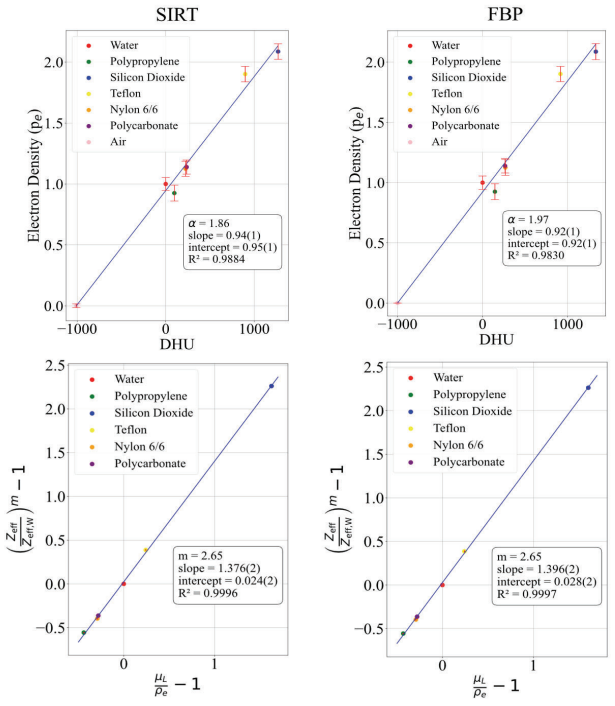


ig. 5 Calibrated CT number for the low-energy band for the test phantom eft) and the calibration phantom (right). The colored circles indicate the gions of interest where the CT number of each material composing the mple was estimated.

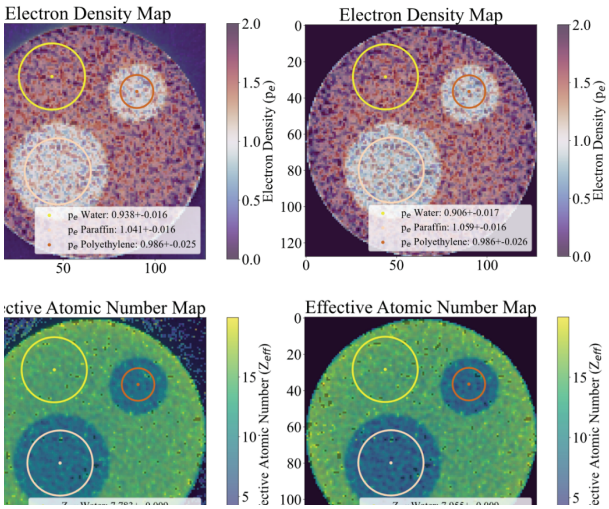
TABLE I

ALUES OF THE MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) OF ρ_E, AND FF FOR EACH MATERIAL AND RECONSTRUCTION METHOD. THE GLOBAL /ERAGE MAPE IS ALSO REPORTED.

Material	Method		
		$_e$	Z_{eff}
Water	FBP	9.41	6.38
	SIRT	6.18	4.08
Polyethylene	FBP	2.09	1.99
	SIRT	2.08	1.94
Paraffin	FBP	10.36	8.47
	SIRT	8.47	6.58
Average	FBP	7.29	5.61
	SIRT	5.58	4.20



the left, the SIRT fits; on the right, the FBP fits. The top row shows the fits for the electron density relative to water, and the bottom row shows the fits



Electron density maps (top row) and effective atomic number maps (bottom row) reconstructed using SIRT (left) and FBP (right) for the test phantom. Regions of interest, represented by the colored circles, were used to estimate the values of ρ_e and Z_{eff} of each material.

IV. CONCLUSIONS

Through Monte Carlo simulations, the feasibility of the tabletop PCCT system for material characterization based on ρ_e and Z_{eff} was demonstrated. The SIRT reconstruction algorithm yielded an increase of 0.55% in R^2 for ρ_e and a negligible difference of 0.01% for Z_{eff}, compared with FBP.

For ρ_e and Z_{eff} values of each material in the test phantom, an MAPE average reduction of 23% for ρ_e and 25% for Z_{eff} was obtained when adopting SIRT. Thus, based on these metrics, it was shown that the SIRT algorithm provides superior material characterization in this setup. This indicates that even a modest increase in the signal-to-noise ratio provided by SIRT, compared to FBP, results in an improvement in the accuracy of ρ_e and Z_{eff} estimation. Post-processing approaches aimed at reducing image noise and improving the SNR are being evaluate and should be the focus of future work. The impact of filtering in the FBP algorithm, since the Hamming window was selected, which smooths the image and directly affects the obtained SNR, as well as incorporating regularization methods for the SIRT algorithm, such as Tikhonov or total variation to minimize the effects of noise, given that we are working with low photon counts compared to experimental setups, will also be considered in future works. Additionally, the authors are working on an experimental approach to validate the results using data acquired from a physical PCCT system, assessing the applicability of the proposed methodology and the obtained results under realistic conditions.

REFERENCES

] K. Taguchi, I. Blevis, and K. Iniewski, *Spectral, Photon Counting Computed Tomography: Technology and Applications* (Devices, Circuits, and Systems). CRC Press, 2020.

[2] S. Leng *et al.*, "Photon-counting detector CT: system design and clinical applications of an emerging technology," (in eng), *Radiographics,* vol. 39, no. 3, pp. 729-743, 2019, doi: 10.1148/rg.2019180115.

[3] M. Torikoshi *et al.*, "Electron density measurement with dual-energy x-ray CT using synchrotron radiation," (in eng), *Phys Med Biol,* vol. 48, no. 5, pp. 673-85, Mar 07 2003, doi: 10.1088/0031-9155/48/5/308.

[4] G. Landry *et al.*, "Extracting atomic numbers and electron densities from a dual source dual energy CT scanner: experiments and a simulation model," (in eng), *Radiother Oncol,* vol. 100, no. 3, pp. 375-9, Sep 2011, doi: 10.1016/j.radonc.2011.08.029.

[5] M. Saito, "Simulation of photon-counting detectors for conversion of dual-energy-subtracted computed tomography number to electron density," (in eng), *Radiol Phys Technol,* vol. 12, no. 1, pp. 105-117, Mar 2019, doi: 10.1007/s12194-018-00497-0.

[6] M. Saito and S. Sagara, "A simple formulation for deriving effective atomic numbers via electron density calibration from dual-energy CT data in the human body," (in eng), *Med Phys,* vol. 44, no. 6, pp. 2293-2303, Jun 2017, doi: 10.1002/mp.12176.

] M. Saito and T. Saito, "A simple algorithm to derive virtual non-contrast electron density from dual-energy computed tomography data for radiotherapy treatment planning," (in eng), *Med Phys,* vol. 52, no. 5, pp. 3107-3116, May 2025, doi: 10.1002/mp.17648.

] S. Agostinelli *et al.*, "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 506, no. 3, pp. 250-303, 2003/07/01/ 2003, doi: https://doi.org/10.1016/S0168-9002(03)01368-8.

[9] P. Arce *et al.*, "Results of a Geant4 benchmarking study for bio-medical applications, performed with the G4-Med system," *Medical Physics,* vol. 52, no. 5, pp. 2707-2761, 2025/05/01 2025, doi: https://doi.org/10.1002/mp.17678.

[10] J. Allison *et al*, "Geant4 developments and applications," *IEEE Transactions on Nuclear Science,* vol. 53, pp. 270-278, 2006.

[11] J. Allison *et al.*, "Recent developments in Geant4," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* vol. 835, pp. 186-225, 2016/11/01/ 2016, doi: https://doi.org/10.1016/j.nima.2016.06.125.

[12] A. Tomal, J. C. Santos, P. R. Costa, A. H. Lopez Gonzales, and M. E. Poletti, "Monte Carlo simulation of the response functions of CdTe detectors to be applied in x-ray spectroscopy," *Applied Radiation and Isotopes,* vol. 100, pp. 32-37, 2015/06/01/ 2015, doi: https://doi.org/10.1016/j.apradiso.2015.01.008.

[13] S. v. d. Walt *et al.*, "scikit-image: Image processing in Python," *CoRR,* vol. abs/1407.6245, / 2014.

[14] T. M. Buzug, *Computed tomography : from photon statistics to modern cone-beam CT*. Berlin: Springer, 2008, pp. xiii, 521 p.

[15] D. Gürsoy, F. De Carlo, X. Xiao, and C. Jacobsen, "TomoPy: a framework for the analysis of synchrotron tomographic data," (in eng), *J Synchrotron Radiat,* vol. 21, no. Pt 5, pp. 1188-93, Sep 2014, doi: 10.1107/s1600577514013939.

[16] N. Hünemohr, H. Paganetti, S. Greilich, O. Jäkel, and J. Seco, "Tissue decomposition from dual energy CT data for MC based dose calculation in particle therapy," (in eng), *Med Phys,* vol. 41, no. 6, p. 061714, Jun 2014, doi: 10.1118/1.4875976.

Physics-Augmented Simulation of Deep Silicon Photon Counting Detector

Haomin Hou, Jiabing Sheng, Yankun Wang, Yule Li, Junhao Wu, Deyang Kong, Jingyi Liao
Hao Wang*, Zhaoying Bian*, Dong Zeng*

***Abstract*—Edge-on deep silicon photonic counting detectors (PCDs) show great promise in spectral computed tomography (CT). However, accurately simulating such detectors remains a significant challenge. Their fundamental response is primarily governed by Compton scattering. This shift in the underlying interaction mechanism, coupled with the use of inter-pixel tungsten foils for collimation and crosstalk suppression, results in severe geometric anisotropy and complex localized energy deposition events. Existing simulation toolkits lack physics models tailored to this unique architecture. Consequently, significant gaps exist in the systematic modeling and performance analysis of deep silicon PCDs, particularly regarding the fine effects of Compton recoil spectra and their interactions with anisotropic detector structures. To address this, this study proposes an extended, physics-driven simulation framework specifically designed for deep silicon detectors. We reconfigured the core physics engine by replacing the traditional discrete fluorescence model with a continuous differential cross-section model based on the Klein-Ninka form. Furthermore, the introduction of a geometric masking kernel within the charge-sharing induction model explicitly implements the physical isolation requirement of the tungsten collimator from the silicon channel. This extended framework provides a tool that combines physical rigor with computational efficiency for optimizing deep silicon detector geometries, which is crucial for fully realizing the potential of silicon-based photon-counting CT.**

***Index Terms*—Deep Silicon, Photon Counting Detector, Compton Scattering, Tungsten Collimator**

I. Introduction

PHOTON-COUNTING computed tomography (PCCT) which utilizes semiconductor-based photon-counting detectors (PCDs) has emerged as a paradigm-shifting modality in medical imaging. Unlike conventional energy-integrating detectors (EIDs), which suffer from signal degradation through multi-stage light to electricity to light conversion processes, PCD directly convert X-ray photons into discrete electrical pulses. In PCD, the charge generated by each incident photon is proportional to its energy, allowing precise energy resolution via pulse height analysis. This mechanism not only avoids spectral distortion caused by EID signal averaging but also unlocks unprecedented performance: (1) Intrinsic spectral differentiation for multi-material decomposition; (2) Dose-efficient imaging with reported radiation exposure reductions of 20-40% in clinical protocols; (3) Suppressing bean hardening artifacts via PCD uniformly weights photons of all energies, effectively reducing beam hardening artifacts [1,2].

The widely studied detector materials for PCCT are cadmium telluride (CdTe) and cadmium zinc telluride (CZT) [3]. Their primary advantage lies in their high atomic numbers, which confer a high linear attenuation coefficient and a dominant photoelectric absorption fraction within the diagnostic X-ray energy range (30–150 keV). This results in excellent quantum efficiency with relatively thin detector geometries, making them strong candidates for capturing the full spectrum of incident photons.

However, the performance of these high-Z semiconductors in clinical environments is challenged by the extreme flux conditions of a CT scan, where local photon fluence rates can reach 100–1000 Mcps mm^{-2} at the detector surface [4]. Such high fluxes place severe demands on detector and readout electronics speed. A critical limitation of CdTe and CZT is its susceptibility to pulse pile-up at flux rates significantly lower than those encountered in clinical practice, as demonstrated in prior studies [5]. Any residual pile-up degrades the energy resolution of the detector, thereby eroding a substantial portion of the potential image quality improvements offered by photon counting over conventional energy integrating systems. Furthermore, a notable intrinsic drawback of CdTe and CZT is the high probability of K-shell fluorescence emission following a photoelectric interaction. In this process, a portion of the deposited energy is re-emitted as a secondary fluorescence photon [6]. This photon may escape the detector entirely, leading to an incomplete charge collection and spectral distortion, or it may be reabsorbed in a neighboring

This work was supported in part by the National Key R&D Program of China under Grant 2024YFA1012000 and 2024YFC2417800, the National Natural Science Foundation of China under Grant 62571228, Beijing Natural Science Foundation under Grant Z250002, and Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515140146. (Corresponding author: Zhaoying Bian, Hao Wang, Dong Zeng.)

Haomin Hou, Yankun Wang, Yule Li, Junhao Wu, Deyang Kong, Jingyi Liao, Hao Wang, Zhaoying Bian and Dong Zeng is with the School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China. (e-mail: houhm@smu.edu.cn, ykwang@smu.edu.cn, ylli@smu.edu.cn, wuiunhao@smu.edu.cn, kongdy@smu.edu.cn, 1jy@smu.edu.cn, wang_hao@smu.edu.cn, zybian@smu.edu.cn and zd1989@smu.edu.cn)

Jiabing Sheng is with the Sun Yat-sen University Cancer Center, Guangzhou 510515, China. (e-mail: 1606941702@qq.com)

pixel. The latter case results in charge sharing and correlated noise, manifesting as spatial blurring and a degradation of both the signal to noise ratio and the spectral resolution [7]. These effects collectively offset some of the inherent signal advantages provided by the high photoelectric cross section of these materials. Silicon (Si) offers several advantages as a material for x ray detectors in PCCT. In addition to its low cost and manufacturability, silicon benefits from short charge carrier collection times, which reduce intrinsic signal pile-up and enable operation at high fluence rates [8,9]. Compared to high-Z alternatives, silicon is also less susceptible to K-edge fluorescence, a source of spectral distortion and spatial blurring. The primary limitation of silicon is its relatively low linear attenuation coefficient in the diagnostic energy range, which necessitates considerable detector thickness to achieve sufficient quantum efficiency. Moreover, Compton scattering dominates over photoelectric absorption in silicon at these energies, leading to partial energy deposition events. In such events, the incident photon deposits only a fraction of its energy locally before scattering; the scattered photon may then escape the detector or cause secondary interactions, resulting in signal loss, cross-talk, and degraded spectral performance.

To overcome silicon's inherent low attenuation, an edge-on detector geometry has been introduced, notably in developments by GE HealthCare. In this design, x-rays enter parallel to the wafer plane, traversing a long absorption path through the silicon bulk [10]. The high purity and excellent charge-transport properties of silicon minimize polarization and charge tailing effects in such detectors. To mitigate the adverse effects of Compton scattered photons, thin tungsten (W) foils are often integrated between silicon layers, serving as both collimators and cross-talk barriers. Consequently, the response of a deep silicon detector is governed by distinct physics: Compton scattering dominates the initial interaction, recoil electrons deposit a continuous spectrum of energy, and the anisotropic geometry of the tungsten foils constrains subsequent photon transport.

Existing simulation methodologies, particularly Monte Carlo (MC) based tools such as the photon counting toolkit (PcTK) [10], are predominantly optimized for conventional photoelectric dominant materials like CdTe. These tools emphasize K-fluorescence and isotropic charge sharing models but lack dedicated treatments for the Compton recoil energy spectrum and the directional masking effects of embedded tungsten foils. As a result, current modeling of deep silicon detectors relies heavily on computationally intensive MC simulations [10,11].

The current development of deep silicon Geometry's photon counting detector is centered on MC simulations. In this article, we present an analytical modeling framework for deep silicon spectroscopic x-ray detectors. We extend the model by incorporating a Compton-scattering branch based on the Klein-Nishina cross section [12] and introducing a geometric masking matrix to accurately model the influence of tungsten foils on charge sharing and photon re-absorption. This model provides a fast, accurate, and physically transparent alternative to full MC simulations for the design and performance analysis of deep silicon PCCT detectors.

II. Method

Our silicon strip detector model is based on the MC model constructed in Reference [9]. The detector consists of a silicon wafer shown in Fig.1. These foils are designed to block scattered X-ray photons, thereby reducing the probability of repeated counting of incident photons. Unlike the actual detector design, the simulated detector is not segmented along the depth direction. This simplified design stems from the omission of stacking effects in low count rate studies, where optimizing the number of depth segments exceeds the scope of this research. By disregarding charge sharing effects between depth segments, this approach maintains manageability in detector simulation calculations.

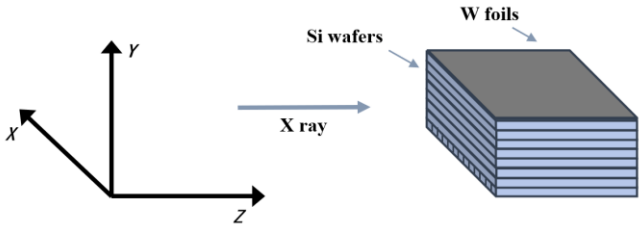


Fig.1. Simulation geometry of with tungsten foils.

A. Physics Interaction Modeling

For an incident photon of energy E_1, the probability of detection is governed by the total linear attenuation coefficient $\mu_{tot}(E_1)$. In the CdTe model, the interaction logic branches into three discrete K-shell fluorescence states ($q = 0, 1, 2$). For deep silicon geometry, the K-shell fluorescence yield is negligible and locally absorbed. Therefore, we simplified the photoelectric branch ($p = 1$) to a single full-energy deposition state ($q = 0$) and structurally activated the Compton scattering branch.

We redefine the conditional probability space $Pr(p|E_1)$ based on the interaction cross-sections of silicon. $p = 1$ represents energy absorption. The deposited energy is $E_2 = E_1$. $p = 2$ represents inelastic scattering. Unlike the discrete energy loss in fluorescence, this branch generates a continuous spectrum of deposited energy $E_2 = T_{recoil}$.

To model the Compton branch accurately, we replaced the deterministic energy transfer logic with a probabilistic recoil model. The probability density function (PDF) of the scattered photon angle θ is derived from the differential Klein-Nishina cross-section:

$$\frac{d\sigma_{KN}}{d\Omega} \propto \left(\frac{E'}{E_1}\right)^2 \left(\frac{E'}{E_1} + \frac{E_1}{E'} - \sin^2\theta\right), \quad (1)$$

where $E'(\theta)$ is the scattered photon energy. The energy deposited into the pixel grid, E_2, is equal to the kinetic energy of the recoil electron:

$$E_2(\theta) = E_1 - \frac{E_1}{1 + \alpha(1 - \cos\theta)}, \quad (2)$$

where $\alpha = \frac{E_1}{m_e c^2}$. This reformulation allows the simulation to generate the characteristic Compton shelf observed in silicon, contrasting with the discrete the escape peaks.

B. Electronic Noise Generating

Following the energy deposition E_2, electronic noise E_3 is introduced as an additive Gaussian process. We model this stage using a zero-mean normal distribution $N(0, \sigma_e^2)$, where σ_e represents the equivalent noise charge. This stage remains consistent with the PcTK to ensure comparability.

C. Anisotropic Charge Summating

We simulate the charge sharing by integrating a 2D Gaussian charge cloud $\rho(x, y)$ over the pixel array. The signal induced in a neighbor pixel (i, j) is traditionally calculated as:

$$w_{i,j} = \iint_{Pixel_{i,j}} \rho(x - x_0, y - y_0)\, dxdy. \quad (3)$$

This assumes an isotropic medium where charge carriers can diffuse freely to any of the 8 neighbors in a 3×3 cluster.

The deep silicon detector architecture employs a stacked geometry where silicon wafers are interleaved with tungsten foils along the y axis. These foils act as physical barriers, strictly prohibiting charge diffusion between layers.

To model this, we introduced a geometric masking operator M into the summation. The effective charge induction matrix W_{eff} is defined as the Hadamard product of the isotropic induction matrix W_{iso} and the mask M:

$$W_{eff} = W_{iso} \circ M. \quad (4)$$

The Tungsten masking kernel M for a 3×3 neighborhood centered at pixel 5 is defined as:

$$M = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad (5)$$

where $M_{i,j} = 1$ allows charge sharing laterally within the silicon wafer along the x axis, while $M_{i,j} = 0$ enforces the isolation imposed by the tungsten foils in the vertical and diagonal directions. This modification fundamentally alters the covariance structure of the output, nullifying the cross-talk terms $\mathrm{Cov}(P_{center}, P_{vertical})$ and ensuring the simulation reflects the edge-on collimated geometry.

D. Covariance Matrix Synthesizing

The final detector response is synthesized by aggregating the moments from the redefined branches. The total covariance matrix $K_{total}(E_1)$ is computed by the weighted summation of $p = 1$ and $p = 2$ contributions:

$$K_{total}(E_1) = \sum_{p \in \{PE, CS\}} Pr(p|E_1) \cdot \left[K_p(E_1) + \Delta_p(E_1)\right], \quad (6)$$

where K_p is the covariance within branch p subject to the masking operator M, and Δ_p accounts for the inter-branch variance caused by the difference between the mean recoil energy and the full energy peak. This unified formula enables us to more accurately simulate low-atomic number Compton-dominated detectors with complex internal geometries based on simulations of high-atomic number fluorescence-dominated detectors.

III. Experiments and Results

A. Simulation Methods

The simulation modeled the stochastic transport and energy deposition of monochromatic X-ray pencil beams at three diagnostic energies. These energies were selected to probe the distinct interaction regimes: photoelectric dominance, transitional mixing, and Compton dominance.

For the deep silicon model, the detector geometry was defined as an edge-on stack. The active sensor volume comprised silicon wafers with a pixel pitch of 0.5 mm to ensure sufficient quantum efficiency for high energy photons. Crucially, the geometric masking kernel was parameterized to represent 20 μm tungsten foils interleaved between 480 μm silicon layers. The physics engine was configured to enable the Klein-Nishina Compton branch (q_{CS}) while suppressing the K-shell fluorescence branch (q_{fl}). For the baseline comparison, a standard CdTe detector was modeled using the PcTK physics settings, enabling K-shell fluorescence escape and reabsorption ($q = 0, 1, 2$). The geometry was set to a conventional face-on configuration with a thickness of 1.6 mm, typical for clinical photon counting systems. Electronic noise was modeled identically for both sensors as additive white Gaussian noise. A summary of the key physical and geometric parameters used in the simulation is provided in Table I.

TABLE I
THE RELEVANT DETECTOR AND EXPERIMENTAL PARAMETER SETTINGS

Explain	Data Experiment Parameter
Density	2.33 g/cm^3
Detector size	100 mm × 100.5 mm
Sensor depth	31.0 mm
Pixel pitch	0.5 mm × 0.48 mm
Anti-scatter barrier	20 μm, Tungsten Foil
Electronic noise	2 keV
Voltage	120 keV
Energy threshold	20, 45, 62 keV
Diagnostic energies	20, 60, 100 keV

B. *Results on Spectra Analysis*

The primary validation of the extended framework relies on the comparative analysis. Fig.2(a) presents the simulated energy response profiles for the deep silicon detector and the standard CdTe detector (Fig.2(b)). As shown in Fig.2(a), the silicon response at high energies (100 keV) exhibits a distinct spectral morphology characterized by a reduced full energy peak and a broad, continuous low energy plateau. This continuum extends from 0 keV up to the Compton edge. Notably, at 20 keV, the response is nearly Gaussian. This confirms that at energies below the threshold for significant scattering, the model correctly reverts to photoelectric dominated behavior, accurately preserving the full energy of the incident photon. Furthermore, the absence of discrete peaks in the low energy tail validates the assumption that silicon's low energy fluorescence is locally absorbed and does not contribute to spectral crosstalk. In contrast, the CdTe response displays characteristic discrete features associated with high-Z materials. For incident energies of 60 keV and 100 keV, distinct escape peaks are visible at energies $E_{in} - E_{fluoro}$.

In the absence of the tungsten mask, isotropic charge sharing would result in significant low energy tailing due to charge loss to vertical neighbors. The observed response indicates that the masking algorithm successfully confined signal collection to the lateral silicon wafer, effectively suppressing vertical crosstalk and preserving spectral fidelity. The direct comparison demonstrates that the framework captures the transition from line feature dominated spectra to continuum feature dominated spectra.

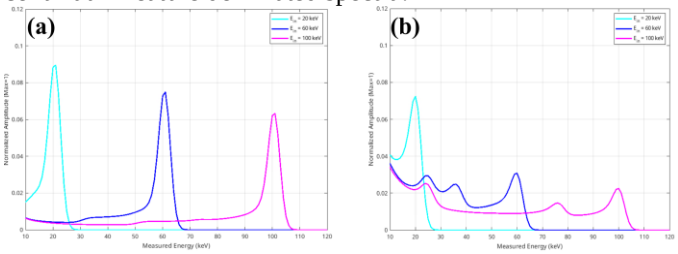


Fig.2. (a) Energy response curves of the deep silicon analysis model, (b) Response curves for the CdTe model in PCTK, corresponding to 20 keV, 60 keV, and 100 keV.

IV. Discussion and Summary

The transition to silicon necessitates a re-evaluation of detector response. Fig.3 illustrates the simulated response to a 120 kVp spectrum. The comparison between the incident flux and the detected spectrum reveals the continuous elevation in low energy counts. This validates our physics engine, confirming the implementation of model partial energy deposition from recoil electrons. Furthermore, the spectral attenuation pattern confirms that the tungsten masking kernel successfully treats collimator interactions as energy loss events rather than erroneous crosstalk, preventing the overestimation of spatial correlation inherent in isotropic models.

The primary contribution of this work is the establishment of a high-fidelity simulation environment that bridges the gap between standard tools optimized for CdTe and the emerging architecture of deep silicon PCD. We have introduced two novel capabilities: (1) The replacement of discrete fluorescence logic with a differential cross section model allows for the precise prediction of the continuous recoil electron spectrum, a prerequisite for understanding silicon's energy response; (2) The introduction of the geometric masking operator provides a mathematical method to simulate anisotropic charge transport in stacked detector arrays without the computational cost of full MC particle tracking.

While this framework successfully captures the physics of deep silicon, it also highlights a critical limitation that the Compton shelf represents a significant degradation of spectral fidelity compared to ideal photon peak responses. If uncorrected, this continuum of partial energy deposition events complicates photon pile-up correction and reduces the separability of basic materials in spectral CT. We aim to develop Compton correction algorithms utilizing the precise detector response matrix generated by this simulation. By treating the simulated shelf as a known prior, we will implement regularized unfolding methods to mathematically recover the incident photon spectrum from the distorted pulse-height measurements. The ultimate goal is to transform the complex physics of silicon into clinical value. We will explore optimized energy weighting schemes that specifically account for the Compton variance, aiming to maximize the Contrast-to-Noise Ratio for iodine and gadolinium maps, thereby mitigating the inherent low-Z disadvantage of silicon.

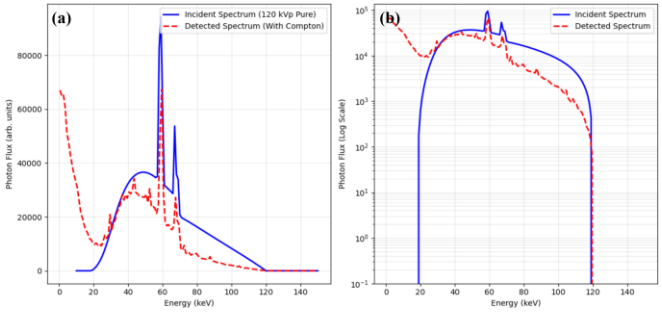


Fig.3. Comparison of the output spectrum and the spectrum detected by the detector. (a) Shows macroscopic intensity and main peak shape. (b) Shows dynamic range.

References

[1] A. Byl, L. Klein, S. Sawall, et al., "Photon-counting normalized metal artifact reduction (NMAR) in diagnostic CT," Medical Physics, vol. 48, no. 7, pp. 3572–3582, 2021.

[2] J. Schreck, K. R. Laukamp, J. H. Niehoff, et al., "Metal artifact reduction in patients with total hip replacements: Evaluation of clinical photon counting CT using virtual monoenergetic images," European Radiology, vol. 33, no. 12, pp. 9286–9295, 2023.

[3] Schlomka, JPea, et al. "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography." Physics in Medicine & Biology 53.15 (2008): 4031.
[4] Roessl E, Schlomka J-P and Proksa R 2008 Edge-on semiconductor x-ray detectors—towards high-rate counting computed tomography IEEE Nucl. Sci. Symp. Conf. Record pp 1748–51.
[5] Barber W C et al 2009 Characterization of a novel photon counting detector for clinical CT; count rate, energy resolution, and noise performance Proc. SPIE 7258 725824
[6] C. Xu, M. Danielsson, and H. Bornefalk, “Evaluation of energy loss and charge sharing in cadmium telluride detectors for photon-counting computed tomography,” IEEE Trans. Nucl. Sci. 58, 614–625 (2011).
[7] Taguchi, Katsuyuki, et al. "Spatio-energetic cross talk in photon counting detectors: detector model and correlated Poisson data generator." Medical physics 43.12 (2016): 6386-6404.
[8] Liu, Xuejin, et al. "A silicon-strip detector for photon-counting spectral CT: energy resolution from 40 keV to 120 keV." IEEE Transactions on Nuclear Science 61.3 (2014): 1099-1105.
[9] Persson, Mats, Adam Wang, and Norbert J. Pelc. "Detective quantum efficiency of photon-counting CdTe and Si detectors for computed tomography: a simulation study." Journal of Medical Imaging 7.4 (2020): 043501-043501.
[10] Taguchi, Katsuyuki, et al. "Spatio-energetic cross-talk in photon counting detectors: Numerical detector model (Pc TK) and workflow for CT image quality assessment." Medical physics 45.5 (2018): 1985-1998.
[11] Yang, Shiyan, Mengjia Xue, and Tianwu Xie. "Development of a Monte Carlo simulation platform for the systematic evaluation of photon-counting detector-based micro-CT." Physica Medica 126 (2024): 104824.
[12] Klein, Oskar, and Yoshio Nishina. "Über die Streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac."Zeitschrift für Physik 52.11 (1929): 853-868.

Bias and Noise Problems in Spectral Photon-counting CT Imaging Task at Low-count Conditions

Jia Wei, Karl Stierstorfer, George S. K. Fung, Christoph Polster, and Katsuyuki Taguchi

Abstract— **Photon counting detector (PCD)-based computed tomography (CT) has demonstrated significant clinical potentials, especially for low-dose and spectral imaging applications. However, its effectiveness is compromised at extremely low photon counts as the detector pixel area decreases significantly compared to conventional energy-integrating detectors (EIDs). The thresholding process further reduces count per bin. The logarithmic operation from counts to line integrals introduces biases and artifacts. In PCD spectral CT imaging, photon counts are unevenly distributed across energy bins, and each energy bins have fewer counts. This results in bin-dependent biases, which, in turn, lead to biases in material decomposition (MaD). In this study, we present a novel method that optimizes the trade-off between bias and noise in post-log data, enhancing accuracy and repeatability of low-dose spectral CT imaging. The proposed method introduces count-dependent terms to independently address bias and variance. Computer simulations were conducted to assess the performance of the proposed method. Results demonstrated that the proposed method showed overall best performance in reducing the bias and variance in line integral, reconstructed images, and in MaD results. By mitigating issues associated with low photon counts, the proposed method enhances clinical viability of low-dose PCD CT, with the potential of facilitating more accurate quantification and diagnosis.**

Index Terms— **Computed tomography, photon counting, bias, variance, spectral imaging.**

I. Introduction

In the past few decades, photon counting detector (PCD) computed tomography (CT) has demonstrated significant clinical potentials, especially in spectral imaging applications [1-3]. Different from energy-integrating detectors (EIDs) where spectral information is generally obtained from multiple measurements with different incident spectrums, PCDs acquire spectral information through sorting each incoming photons to different energy bins based on the pre-defined energy thresholds. This classification process provides better energy separation which is required for accurate material decomposition (MaD) [1, 4]. The drawback is that this process results in fewer number of photons in each energy bin. In spectral PCD CT imaging, photon counts are unevenly distributed among energy bins. The count distribution is strongly influenced by the energy thresholds, imaging system configuration, and charge-sharing (CS) effect which contaminates the lower energy bins with incorrect counts and leads to count loss in the higher energy bins [1]. The bias in projection domain material decomposition (pMaD) with fewer counts is well known [5-7]. Maximum likelihood (ML) is usually used for the mapping process; however, it was shown that even ML had biases when the count is low as the MaD process inherently incorporates similar domain conversion [8].

Besides the effect on bias in MaD results, the low photon count also results in higher variance and bias in the reconstructed image after the logarithmic process. This amplifies the effect of quantum noise in the measurements and introduces bias at extremely low photon counts [9, 10]. Low-count bias can be decoupled into two different processes. First is the statistical bias corresponding to the non-linear logarithmic conversion process used for computing line integrals for CT image reconstruction: $E[\ln(N)] \neq \ln(E[N])$, where $E[\cdot]$ is the expectation and N is a Poisson variable [10]. Another source of bias is the handling of "zero count" (i.e., the number of recorded events being zero) in Poisson data, as $\ln(0)$ is undefined [9]. This issue occurs more frequently as the detector pixel area decreases by a factor of ten from EIDs [1]. A common solution for zero presence is replacing zero with a small positive value so the logarithm can be solved mathematically [9]. Unfortunately, the additional small positive value skews Poisson statistics and introduces additional bias since $E[N'] > E[N]$, where N' is the modified Poisson random variable after replacing zero with another value. In addition, average counts vary significantly across energy bins of multi-bin spectral CT. The bin-dependent bias and noise make MaD even more complicated. Bias correction may be effective during the MaD process [5,6] but in this

This work was partially supported by Siemens Healthineers.

J. Wei is with Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD 21205 USA.

K. Stierstorfer and C. Polster are with Computed Tomography, Siemens Healthineers, Forchheim 91301, Germany.

G. S. K. Fung is with Siemens Healthineers, Malvern, PA 19355 USA.

K. Taguchi is with the Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD 21287 USA (e-mail: ktaguchi@jhmi.edu).

study, we aim to address the bias at the very source of the issue.

In our previous work, we proposed a data processing method that simultaneously reduces bias and noise of post-log data [11]. The proposed method aims to enhance the accuracy and repeatability of CT imaging, particularly at an ultra-low-dose level, where expected count (λ) is less than 1. At this count level, the zero-count problem is prevalent. In this study, we extend this approach to spectral PCD CT by applying the method to each energy bin and evaluating its improvement on spectral imaging tasks. In addition, we investigate how count-dependent bias and CS effect influence MaD performance.

II. Methods and Materials

We outline four algorithms (the proposed method and three other competing methods) in *Sec.A* and describe the computer simulation study in *Sec.B*.

A. Data processing algorithms including the proposed Low-count Bias–noise reduction method (LBN)

Overview: All methods take count measurements at input. For each sinogram point, there are n_k measurements, where n_k is the number of energy bins, and each measurement corresponds to the number of photons recorded in the k-th energy bin, N_k. Then, one of the four methods outlined below is applied to perform logarithmic conversion of N_k and outputs y_k in Step 1; the line integral is calculated based on y_k and CT image for bin k is reconstructed using filtered backprojection in Step 2; and an image-domain material decomposition (iMaD) is performed in Step 3.

A.1 Step 1: Logarithm conversion

Method 0: Let N be the Poisson-distributed random variable with mean λ, $N \sim \text{Poisson}(\lambda)$. N' represents the variable after replacing zero with a small positive value, ϵ. The post-log result of the new variable is shown below:

$$y_{method0} = \ln(N'). \quad (1)$$

Method 1: Based on the Taylor series, Li et al. showed that the statistical bias could be modeled, hence corrected, using Laurent polynomials of the random variable [9]:

$$y_{method1} = y' + \sum_{k=1}^{K} C_k N'^{-k}, \quad (2)$$

where C_k are the polynomial coefficients. Method 1 computes C_k analytically, while ignoring the effect of zero-replacement.

Gaussian Filtering (Smoothing): Gaussian filter with width (σ) is applied to the count data among the detector plane for each view. When the filtered count value $\widehat{N}$ is 0, it will be replaced by ϵ. The zero-replaced new data, $\widehat{N}'$, will be converted into $\ln(\widehat{N}')$.

Proposed Low-count Bias-Noise reduction (LBN) method:

The proposed method is motivated by two key considerations. First, we aimed to decouple bias and noise handling. The bias–noise tradeoff stems from the Laurent polynomial formulation, which models both post-log bias and variance; consequently, the correction coefficients jointly affect bias and noise, so improving one degrades the other [12]. Second, the LBN method leverages local optimization for each individual λ. To address this, we introduced λ-specificity to the data processing framework, where coefficients are optimized for each expected count individually:

$$\hat{y}_{LBN} = \ln(N') + C_0(\hat{\lambda}) + \sum_{k=1}^{K} C_k(\hat{\lambda}) N'^{-k}. \quad (3)$$

The first step in the LBN optimization process is to estimate the coefficients that minimize variance:

$$\hat{C}_k(\lambda) = \arg\min_{C} \text{Var}\left[\ln(N_i') + \sum_{k=1}^{K} C_k(\lambda)\ (N_i')^{-k}\right], \quad (4)$$

where Var[·] represents the oprerator to compute the variance, i represents noise realizations for each λ, and k is the power of the polynomial terms. The coefficients $C_{k>0}$ and their associated terms influence both bias and variance. When fine-tuned to reduce variance, bias will increase, and vice versa. C_1 was implemented in this study.

The constant term, C_0, targets the statistical bias caused by the curvature of logarithm, zero replacement, and the bias introduced by non-zero-order polynomial terms:

$$\hat{C}_0(\lambda) = \text{argmin}_{C_0} \sum_{i=0}^{n} \left(\ln(N_i') + C_0(\lambda) + \sum_{k=1}^{K} \hat{C}_k(\lambda) N_i'^{-k} - \ln(\lambda)\right)^2. \quad (5)$$

The constant C_0 does not contribute to the variance of y_{LBN}. Therefore, $C_{k>0}$ terms are constructed to solely control variance while C_0 controls bias.

When actual count data comes, the initial estimation for each pixel in the detector plane for each projection view is calculated:

$$\hat{\lambda} = G_\sigma(i) N_i + \sum_{j \in \Psi(i)} G_\sigma(j) N_j. \quad (6)$$

The constant term C_0 and coefficients C_k are retrieved from the precomputed LUT based on the estimated Poisson parameter $\hat{\lambda}$ for each POI. Then the coefficients will be applied to Eq. (3) to get the processed line integral value y_{LBN}.

A.2 Step 2: Image reconstruction

Each energy bin is processed independently from Step.1. The line integral was then used to reconstruct bin-specific images using filtered backprojection.

A.3 Step 3: iMaD

Second-order polynomial fitting was implemented for iMaD to estimate the relative density for each basis material. The polynomial coefficients were estimated using bin-specific images from the noise-free, CS-affected calibration phantom.

B. CT-based assessment

We evaluated algorithm performance using CT images

simulated from the calibration phantom and the XCAT abdominal phantom [13]. Density maps of water and bone were generated, and forward projections were produced using the XCIST toolkit to simulate a PCD CT system with detector pixel size of (0.3 mm^2) and 4096 channels × 16 rows × 1024 views over 360° [14, 15]. The source-to-isocenter and source-to-detector distances were 600 mm and 1080 mm, respectively.

In this pilot study, we used four energies bins, and each of them are assigned to a monochromatic x-ray at [34, 57, 81, 103] keV. CS effects among four monochromatic energies were modeled using the Photon Counting Toolkit (PcTK) [16]. The energy of the split photons from each monochromatic energy beams are compared against the predefined bin thresholds [20, 45, 70, 95] keV to determine the probability of assigning additional counts to each lower-energy bin. This scheme allows us to exclude the beam-hardening effects and solely focus on the low-count problem.

iMaD for water and bone basis was performed to evaluate the performance of different data processing methods in spectral imaging tasks. Reconstruction and MaD performance were evaluated using 150 noise realizations (NRs) of the XCAT abdomen phantom [13].

III. Results

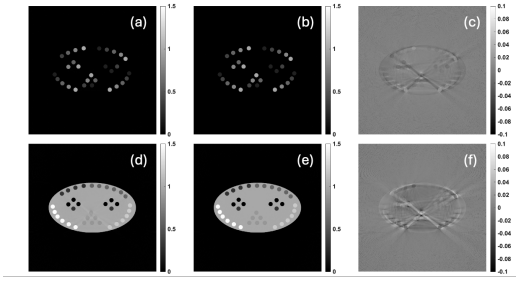


Fig. 1. Basis images for calibration phantom. Colorbar value represents the normalized basis density. (a) and (d) are the bone and water basis images acquired by applying the calibrated polynomial coefficients; (b) and (e) are the ground truth basis images; (c) and (f) are the difference image between the polynomial fitting results and the ground truth.

Fig. 1 shows the design of the calibration phantom and performance of iMaD calibration. Second order polynomials were implemented for iMaD. The coefficients were estimated based on the images reconstructed from noise-free, CS-affected count data. Fig. 1(a) and (d) show the generated basis image by applying the estimated coefficient to the calibration image.

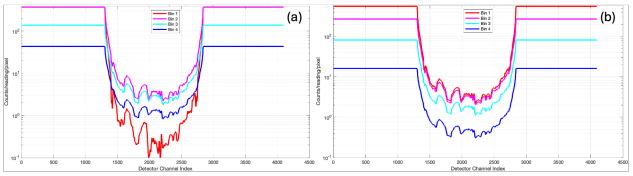


Fig. 2. Count distribution across energy bins for XCAT abdominal phantom. (a) Without CS effect; (b) with CS effect.

As shown in Fig. 2(b), the expected count in higher energy bin (bin 3 and bin 4, shown in light and dark lines) drops drastically compared to Fig. 2(a). The expected count in lower energy bin increases (bin 1 and bin 2, shown in red and magenta lines) due the splitting effect caused by CS. The lower energy bins behave less attenuated due to the artificially increased register count.

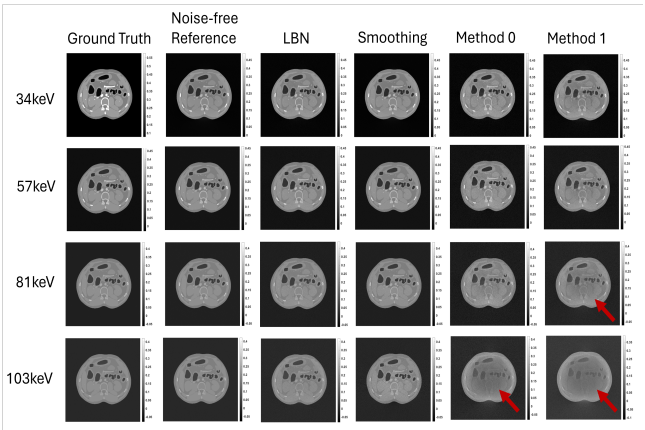


Fig. 3. Reconstructed images for each energy level. The ground truth is the reconstructed images directly from noise-free without CS effect. The other methods are the average results from 150 NRs. Images are leveled at the respective mean value for the adipose tissues with fixed width of 0.5 in unit of cm^{-1}. The red arrows indicate bias induced artifacts.

Fig. 3 shows the reconstructed images for each energy bin. Compared to the ground truth image where the CS effect is ignored, the lower energy bins all behave less attenuated as more counts were registered incorrectly to the lower energy bins. The difference among different processing methods increases as effective energy increases. Higher energy bins which suffered count loss show substantial performance variation among different processing methods due to low photon count. In bin 4, where majority of the counts fall into an extremely low count range ($\lambda < 1$), method 0 and 1 showed substantial shading artifacts and the soft tissue contrast in these two cases are almost gone in the center region (pointed by the red arrows).

Second order polynomial fitting based iMaD was used to estimate the relative basis densities for each image pixel. The same calibrated coefficients were applied to each NR for different processing methods.

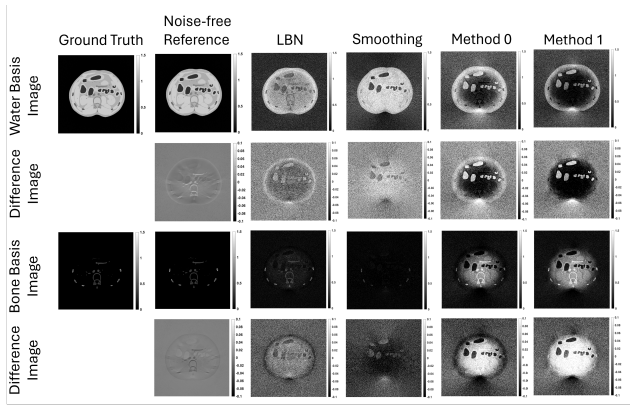


Fig. 4. iMaD results for the XCAT abdominal phantom. The ground truth is reconstructed directly from the basis projection. The noise-free refence image is noise-free, CS-affected result. Row 1 are the water basis images, and row 2 is the difference images between each method and the ground truth water basis image. Row 3 is bone basis images and row 4 is the difference image. Basis images are displayed at [0, 1.5] and difference images are displayed at [-0.1, 0.1] in units of relative basis density.

Fig. 4 shows the averaged basis images from 150 NR for different methods. The proposed LBN method and the smoothing method showed better performance whereas the other two methods showed bigger errors in iMaD results. The comparison between the Ground truth and noise-free reference image suggests there is bias in iMaD caused by CS effect.

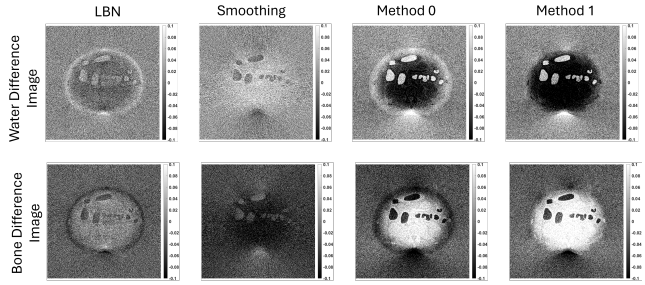


Fig. 5. Difference images between noise-free reference image and other processing methods. Images are displayed at [-0.1, 0.1] in units of relative basis density.

Fig. 5 shows the differences between the noise-free reference and the average iMaD results obtained using different processing methods. The noise-free, CS-affected results represent the theoretically best achievable performance under the current MaD framework. Because CS effect is present in all cases, this comparison isolates the effect of iMaD inaccuracies arising from low-count bias. Methods 0 and 1 perform poorly in basis estimation as the negative bias in bin 4 images dominate the results.

IV. Discussion and Conclusions

The low photon counts significantly degrade the image quality, specifically from increased bias. The proposed LBN method has demonstrated promising performance, improving both bias and variance simultaneously in low-dose settings. In the spectral imaging, the number of counts vary among energy bins. To achieve better quantification results in MaD, it is even more critical to handle the bias at low count settings.

In this study, we implemented the LBN method to individual energy bins. The LBN method showed overall best performance in reducing the bias and variance in line integral, reconstructed images, and in MaD results. Noticeably, LBN hugely reduced the bias and variance, and the second-best bias reduction was smoothing. LBN and smoothing has comparable line integral bias in bin 1 and bin 2, but LBN has bias -0.004 ± 0.047 and -0.004 ± 0.071 for bin 3 and bin 4 (extremely low count range data), whereas smoothing has -0.020 ± 0.031 and 0.123 ± 0.107. For MaD, the water and bone basis biases (mean ± SD) were 0.05 ± 0.64 and -0.09 ± 0.30 for LBN; 0.26 ± 0.75 and -0.27 ± 0.36 for smoothing; 0.05 ± 1.43 and -0.17 ± 0.72 for method 0; and -0.29 ± 1.22 and 0.05 ± 0.59 for method 1.

The bias can be broken into two different sources: 1) low photon count; 2) CS effect. For low-count bias, method 0 resulted in severe negative bias near the object center. This artifact substantially degraded object visibility and underscored the inadequacy of simple zero-replacement approaches. Methods 1 followed the expected trend that globally optimized coefficients for Laurent polynomials begin to fail as the expected count becomes extremely low: this method perform well under moderate count conditions, where only a small fraction of detector pixels receive counts below the valid range. In other words, for these methods to work, it is necessary to keep the dose level to a reasonable value otherwise the image quality deteriorates rapidly, which limits their applicability in low-dose imaging scenarios.

The effect of CS is also studied. The higher energy bins lose counts due to the splitting; lower energy bins gain false counts form downshifted events; and middle bins experience both loss and gain. There is intrinsic bias due to spectral contamination. For higher energy bin, where only count loss was seen, the bias is minimal or gone as there is no spectral contamination.

Overall, results from the XCAT simulations revealed that low expected counts can produce pronounced image artifacts, compromising visibility and diagnostic reliability. The proposed LBN method demonstrated strong capability in simultaneously mitigating bias and noise for different energy bins, even at the extremely low-count regime.

This study also revealed the combined bias arising from low photon count and CS effect. In high energy bins, image bias is dominated by low-photon count. In contrast, in low energy bins, the splitting effect introduced by CS mitigates logarithm conversion bias but introduce bias attribute to CS. The MaD results further demonstrate the importance of bin-specific bias handling. The caveat for the proposed method is that the bias caused by CS effect is not fully solved. In this

study the primary focus is on the low-count bias issue, but in the future, we are interested in further investigating the feasibility of applying a similar processing framework to address both sources of bias spontaneously.

REFERENCES

[1] K. Taguchi and J. S. Iwanczyk, "Vision 20/20: single photon counting x-ray detectors in medical imaging," *Med. Phys*., vol. 40, no. 10, p. 100901, 2013.

[2] M. Danielsson, M. Persson, and M. Sjölin, "Photon-counting x-ray detectors for CT," *Phys. Med. Biol.,* vol. 66, no. 3, p. 03TR01, 2021.

[3] S. Leng, M. Bruesewitz, S. Tao, K. Rajendran, A. F. Halaweish, N. G. Campeau, J. G. Fletcher, and C. H. McCollough, "Photon-counting detector CT: system design and clinical applications of an emerging technology," *Radiographics,* vol. 39, no. 3, pp. 729–743, 2019.

[4] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293–312, 2018.

[5] O. Lee, S. Kappler, C. Polster and K. Taguchi, "Estimation of Basis Line-Integrals in a Spectral Distortion-Modeled Photon Counting Detector Using Low-Rank Approximation-Based X-Ray Transmittance Modeling: K-Edge Imaging Application," *IEEE Trans. Med. Imaging*, vol. 36, no. 11, pp. 2389-2403, Nov. 2017, doi: 10.1109/TMI.2017.2746269.

[6] O. Lee, S. Kappler, C. Polster and K. Taguchi, "Estimation of Basis Line-Integrals in a Spectral Distortion-Modeled Photon Counting Detector Using Low-Order Polynomial Approximation of X-ray Transmittance," *IEEE Trans. Med. Imaging*, vol. 36, no. 2, pp. 560-573, Feb. 2017, doi: 10.1109/TMI.2016.2621821.

[7] K. Taguchi, C. Polster, W. P. Segars, N. Aygun, and K. Stierstorfer, "Model-based pulse pileup and charge sharing compensation for photon counting detectors: A simulation study," *Med. Phys*., vol. 49, no. 8, pp. 5038–5051, Aug. 2022, doi: 10.1002/mp.15779.

[8] A. Wunderlich and F. Noo, "Image covariance and lesion detectability in direct fan-beam x-ray computed tomography," *Phys. Med. Biol.*, vol. 53, no. 10, p. 2471, 2008.

[9] K. Li, J. R. Chen, and M. Feng, "Construction of a nearly unbiased statistical estimator of sinogram to address CT number bias issues in low-dose photon counting CT," *IEEE Trans. Med. Imaging*, vol. 42, no. 6, pp. 1846–1858, 2023.

[10] D. Bushe, R. Zhang, G.-H. Chen, and K. Li, "Unbiased zero-count correction method in low-dose high-resolution photon counting detector CT," *Phys. Med. Biol*., vol. 68, no. 11, p. 115002, 2023.

[11] J. Wei, D. Lee, K. Stierstorfer, G. S. K. Fung, C. Polster, S. Subramanian, and K. Taguchi, "Log conversion with fewer counts for photon counting CT: decreasing both bias and variance simultaneously," in *Proc. SPIE 13405, Medical Imaging 2025: Physics of Medical Imaging*, 134050H, Apr. 2025. [Online]. Available: https://doi.org/10.1117/12.3046256

[12] S. S. Hsieh and P. L. Rajbhandary, "Existence, uniqueness, and efficiency of numerically unbiased attenuation pathlength estimators for photon counting detectors at low count rates," *Med. Phys*., vol. 51, pp. 8742–8750, 2024. [Online]. Available: https://doi.org/10.1002/mp.17406

[13] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys*., vol. 37, no. 9, pp. 4902–4915, 2010. doi: 10.1118/1.3480985

[14] M. Wu, P. FitzGerald, J. Zhang, W. P. Segars, H. Yu, Y. Xu, and B. De Man, "XCIST—an open-access x-ray/CT simulation toolkit," *Phys. Med. Biol*., vol. 67, no. 19, 194002, 2022. doi: 10.1088/1361-6560/ac8f29

[15] XCIST GitHub repository. [Online]. Available: https://github.com/xcist. Accessed: Oct. 22, 2025.

[16] K. Taguchi, K. Stierstorfer, C. Polster, O. Lee, and S. Kappler, "Spatio-energetic cross-talk in photon counting detectors: Numerical detector model (PcTK) and workflow for CT image quality assessment," *Med. Phys.*, vol. 45, no. 5, pp. 1985–1998, May 2018. doi: 10.1002/mp.12863.

Neighbor2Inverse: Self-Supervised Denoising for Low-Dose Region-of-Interest Phase Contrast CT

Johannes B. Thalhammer[1,2,3], Lorenzo D'Amico[4], Lucy Costello[4], Sebastian Peterhansl[1], Tina Dorosti[1,2], Florian Schaff[1], Jannis Ahlers[4], Ronan Smith[5], Marcus Kitchen[4], Franz Pfeiffer[1,2,3], Martin Donnelley[5], Daniela Pfeiffer[2,3], Kaye S. Morgan[4]

***Abstract*—Propagation-based X-ray phase-contrast imaging (PBI) enables high-contrast visualization of lung structures and holds strong clinical potential. However, safe translation to the clinic will require a substantial radiation dose reduction, which inevitably increases image noise. Supervised convolutional-neural-network-based denoising can restore image quality but depends on paired low- and high-dose datasets, which are rarely available in practice. Self-supervised methods avoid this limitation, yet most are not well adapted to the inverse problem of PBI computed tomography (CT). We introduce *Neighbor2Inverse*, a self-supervised denoising framework for low-dose PBI CT. Building on the Neighbor2Neighbor principle, each noisy projection is subsampled into two variants that preserve structural information but contain independent noise realizations. These are reconstructed separately, and the resulting pairs are used to train a denoising network directly in the reconstruction domain. We benchmark the proposed method against established self-supervised denoising approaches. In region-of-interest PBI CT experiments, *Neighbor2Inverse* achieves superior noise suppression while preserving fine structural details, as demonstrated by improved contrast-to-noise ratio, spatial resolution, and composite image quality metrics. Competitive performance is also observed on clinical CT data under simulated low-dose conditions.**

Code and data are available at https://github.com/J-3TO/Neighbor2Inverse.

I. Introduction

PROPAGATION-BASED X-ray phase contrast imaging (PBI) is a technique that exploits the self-interference of a coherent wavefield during free-space propagation to retrieve phase information from an object [1], [2]. Unlike conventional attenuation-based imaging, PBI enhances soft-tissue contrast, making it particularly promising for applications such as cancer detection and characterization in lung imaging [3]–[6]. Currently, such characterization often relies on invasive biopsy procedures for histological validation [7]–[9]. If sufficient diagnostic detail can be obtained at low dose, PBI may reduce the need for biopsy or provide complementary information to improve treatment planning and understanding of tumor heterogeneity.

Minimizing radiation dose is essential for clinical translation of PBI. One effective strategy is region of interest (ROI) CT, also known as interior or local tomography, in which only a localized region is illuminated. Additional dose reduction can be achieved by decreasing the exposure time, at the cost of increased noise and reconstruction artifacts. Deep learning has achieved remarkable success in denoising and artifact reduction for low-dose CT. Most commonly, supervised learning with paired low-dose and high-dose images is used [10]–[12], but acquiring such datasets is often impractical.

Self-supervised alternatives address this limitation. The seminal Noise2Noise approach demonstrated that denoising can be learned directly from noisy image pairs without requiring clean targets [13], though it assumes access to multiple noisy realizations of the same object. To overcome this limitation, several techniques have emerged. *Neighbor2Neighbor* generates pseudo pairs from a single image by downsampling a single image to produce two versions with statistically independent noise but nearly identical underlying signal [14]. *Noise2Inverse* operates in the projection domain by splitting sinograms into sparse subsets for self-supervised training [15]. Building on these concepts, we propose Neighbor2Inverse, a hybrid self-supervised framework designed for low-dose PBI. By Neighbor subsampling in the projection domain, followed by reconstruction and image-domain training, our method combines the inverse problem formulation of *Noise2Inverse* with the statistical robustness of *Neighbor2Neighbor* without requiring paired high-dose data. We compare multiple denoising strategies against the proposed *Neighbor2Inverse* method to identify the most effective approach for low-dose ROI propagation-based phase contrast CT. We further assess robustness by applying the method to clinical CT data under simulated low-dose conditions.

1 Chair of Biomedical Physics, Department of Physics, School of Natural Sciences and the Munich Institute of Biomedical Engineering, both at the Technical University of Munich (TUM), Germany.

2 Institute for Diagnostic and Interventional Radiology, School of Medicine and Health, TUM Klinikum.

3 Institute for Advanced Study, TUM

4 X-ray Imaging Group within the School of Physics and Astronomy, Monash University, Australia

5 Adelaide Medical School and Robinson Research Institute, University of Adelaide, and the Respiratory and Sleep Medicine, Women's and Children's Hospital, both in Adelaide, Australia.

We acknowledge financial support through the European Research Council (ERC Synergy Grant SmartX, SyG 101167328), the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder, the TUM - IAS, as well as the Australian National Health and Medical Research Council (NMHRC) Synergy Grant APP2011204.

No animals were euthanised or subjected to procedures for this study. Calf tissue was collected post-mortem as scavenged material from animals slaughtered at a commercial knackery for non-research agricultural purposes. No live animal experimentation was performed, and in accordance with institutional and national guidelines, formal animal ethics approval was not required. The study complied with the ethical standards of the involved institutions.

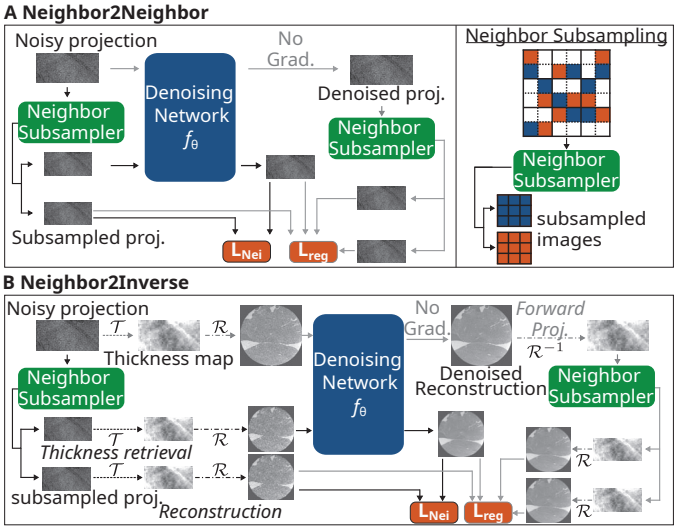


Fig. 1. A - Overview of the *Neighbor2Neighbor* approach applied to projections along with the Neighbor subsampling algorithm.
B - The *Neighbor2Inverse* pipeline.

II. Methods

A. Datasets

ROI propagation-based phase-contrast data were acquired at the Imaging and Medical Beamline (IMBL) of the Australian Synchrotron. An inflated calf lung was imaged at a propagation distance of 5 m using monochromatic 70 keV X-rays, following established protocols [4], [6]. Each ROI (2150x2560 pixels, 9 μm pixel size, scanned with of-center rotation axis) was scanned at high and low dose (200 ms and 15 ms exposure), with four ROIs used for training, one for validation, and one for testing. Minor sample motion caused a slight misalignment between the high- and low-dose acquisitions, rendering supervised training impossible. Projected thickness maps were obtained using single-material phase retrieval [2] and reconstructed via filtered backprojection.

For clinical evaluation, 200 exams from the RSNA Pulmonary Embolism Detection Challenge dataset [16] were used (120 for training, 40 for validation, 40 for testing). Low-dose data were simulated by forward projection into fan-beam sinograms with 2048 views, adding Poisson and Gaussian noise, and reconstructed again via FBP using torch-radon [17].

B. Noise Reduction

1) Neighbor2Neighbor: *Neighbor2Neighbor* (Figure 1A) enables self-supervised denoising by generating two sub-images from a single noisy projection [14]. Each projection $\mathbf{p}_\phi$ is divided into non-overlapping 2×2 patches, from which two neighboring pixels are randomly selected using a subsampling operator G, yielding $g_1(\mathbf{p}_\phi)$ and $g_2(\mathbf{p}_\phi)$. These sub-images share nearly identical signal content but uncorrelated noise under a pixel-wise independence assumption.

The denoising network f_θ is trained by minimizing $\mathbb{E}_{\mathbf{p}_\phi}\|f_\theta(g_1(\mathbf{p}_\phi)) - g_2(\mathbf{p}_\phi)\|^2$. To compensate for small signal discrepancies introduced by subsampling, a regularization term L_{reg} enforcing signal consistency is added, following [14].

2) Neighbor2Inverse: Figure 1B illustrates the proposed *Neighbor2Inverse* framework, which extends *Neighbor2Neighbor* to tomographic reconstruction. Since noise in reconstructed images is spatially correlated, subsampling is again performed in the projection domain. Each projection $\mathbf{p}_\phi$ is subsampled to $g_{1,2}(\mathbf{p}_\phi)$, followed by thickness retrieval $\mathcal{T}$ and FBP reconstruction $\mathcal{R}$, producing two images with nearly identical structure and independent noise.

One reconstruction is denoised using f_θ, while the other serves as the training target. To account for subsampling-induced signal mismatch, a regularization term analogous to *Neighbor2Neighbor* is applied after forward projection and resampling. The resulting objective is

$$\begin{aligned} L = &\left\| f_\theta\left(\mathcal{R}\left[\mathcal{T}g_1(\boldsymbol{p}_\phi)\right]\right) - \mathcal{R}\left[\mathcal{T}g_2(\boldsymbol{p}_\phi)\right]\right\|^2 \\ &+ \gamma\Big\| f_\theta\left(\mathcal{R}\left[\mathcal{T}g_1(\boldsymbol{p}_\phi)\right]\right) - \mathcal{R}[\mathcal{T}g_2(\boldsymbol{p}_\phi)] \\ &\quad - \mathcal{R}g_1\left(\mathcal{R}^{-1}f_\theta\left(\mathcal{R}[\mathcal{T}\boldsymbol{p}_\phi]\right)\right) \\ &\quad + \mathcal{R}g_2\left(\mathcal{R}^{-1}f_\theta\left(\mathcal{R}[\mathcal{T}\boldsymbol{p}_\phi]\right)\right)\Big\|^2 \\ = &\, L_{\text{Nei}} + \gamma L_{\text{reg}}. \end{aligned} \tag{1}$$

Here, L_{Nei} promotes noise suppression in the reconstruction domain, while L_{reg} compensates for signal discrepancies between subsampled projections and vanishes when the network approximates a perfect denoiser. The regularization parameter λ is linearly increased from 0 to 2 during the training.

3) Comparison Methods: For comparison, we trained a Wasserstein GAN with gradient penalty (WGAN-GP) [18] to generate synthetic low-dose projections from given high dose projections. Adding this learned low dose noise to high-dose measurements yields a paired dataset. After reconstruction, this dataset was used to train a denoising U-Net in a supervised manner, referred to as *RecoFakeNoiseNet*. We further compared against *Noise2Inverse* [15] and, for the clinical dataset, fully supervised denoising using available high-dose targets.

4) Training Details: All models were implemented in PyTorch and PyTorch Lightning and trained using Adam optimization. A U-Net backbone was used throughout [19]. Learning rates were selected using a learning-rate finder [20] and scheduled with `ReduceLROnPlateau`.

Neighbor2Inverse models were trained on full reconstructed images using a batch size of 1 with gradient accumulation. The regularization weight γ was linearly increased over 100 epochs, following [14]. All other models were trained with a batch size of 16. Final models were selected based on validation loss.

C. Quantitative Analysis

For the PBI dataset, image quality was assessed using contrast-to-noise ratio (CNR), spatial resolution (SR), and a composite metric $Q = \text{CNR}/\text{SR}$. Ten representative slices were analyzed. CNR was computed from paired air–tissue ROIs, while SR was estimated from tissue–air edges by fitting error functions and extracting the FWHM of the corresponding Gaussian derivative [21].

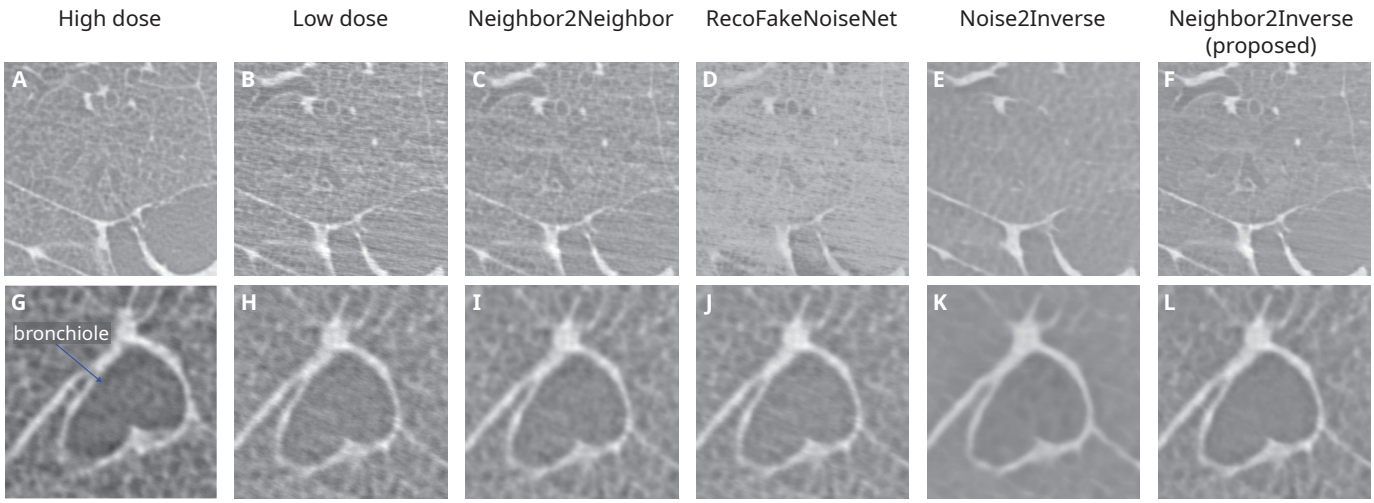


Fig. 2. Comparison of denoising methods for low-dose propagation-based CT reconstructions. Low-dose images (B, H) and corresponding high-dose references (A, G) are shown. Results from *Neighbor2Neighbor* [14] (C, I), *RecoFakeNoiseNet* (D, J), *Noise2Inverse* [15] (E, K), and the proposed *Neighbor2Inverse* (F, L) are compared. The top row shows a 300 × 300 pixel zoom, and the bottom row a 600 × 600 pixel region from the full 4675 × 4675 pixel image.

For the clinical dataset, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) were calculated with respect to the reference images.

III. Results

A. Qualitative Evaluation

Figure 2 compares low-dose reconstructions, the corresponding high-dose reference, and denoised images from different methods. All approaches reduce noise relative to the low-dose input, but with clear qualitative differences. Projection-domain *Neighbor2Neighbor* provides the weakest noise suppression. *RecoFakeNoiseNet* performs well in some regions (panel J) but introduces noticeable distortions elsewhere (panel D), likely due to a mismatch between real and synthetically generated noise. *Noise2Inverse* achieves strong noise reduction but oversmooths fine anatomical structures, including alveolar boundaries. In contrast, *Neighbor2Inverse* effectively suppresses noise while preserving small-scale features, although residual horizontal stripe artifacts remain.

Figure 3 shows results on simulated low-dose clinical CT data. Similar trends are observed: *Neighbor2Neighbor* leaves substantial residual noise, while supervised, Noise2Inverse, and *Neighbor2Inverse* yield comparable denoising performance. Among these, *Noise2Inverse* appears slightly blurrier than the supervised and *Neighbor2Inverse* results.

B. Quantitative Evaluation

Figure 4A summarizes quantitative results on the PBI test set. All methods improve contrast-to-noise ratio (CNR), while all except *Neighbor2Neighbor* also enhance or retain spatial resolution. *Neighbor2Inverse* achieves the highest composite image quality metric Q. Figure 4B reports PSNR and SSIM on the clinical dataset. Supervised and *Noise2Inverse* methods obtain the highest scores, followed by *Neighbor2Inverse* and *Neighbor2Neighbor*. The strong performance of *Noise2Inverse* is consistent with its tendency to produce smoother images, which favor these metrics.

IV. Conclusion

We introduced *Neighbor2Inverse*, a self-supervised denoising framework for low-dose propagation-based phase-contrast CT that operates in the reconstruction domain without requiring paired high-dose data. Across both experimental PBI and simulated low-dose clinical CT datasets, *Neighbor2Inverse* achieves effective noise suppression while preserving fine structural details. These results demonstrate the potential of *Neighbor2Inverse* as a practical and robust denoising strategy for low-dose phase-contrast and conventional CT imaging.

References

[1] P. Cloetens, R. Barrett, J. Baruchel, *et al.*, "Phase objects in synchrotron radiation hard x-ray imaging," *Journal of Physics D: Applied Physics*, vol. 29, no. 1, p. 133, 1996.

[2] D. Paganin, S. C. Mayo, T. E. Gureyev, *et al.*, "Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object," *Journal of Microscopy*, vol. 206, no. 1, 2002.

[3] N. Yagi, Y. Suzuki, K. Umetani, *et al.*, "Refraction-enhanced x-ray imaging of mouse lung using synchrotron radiation source," *Medical Physics*, vol. 26, no. 10, pp. 2190–2193, 1999.

[4] L. Costello, M. Donnelley, Y. Nesterets, *et al.*, "Evaluating the feasibility of region-of-interest x-ray phase contrast imaging for lung cancer diagnostics," *Scientific Reports*, vol. 15, Jun. 2025.

[5] J. N. Ahlers, L. D. Amico, H. Bast, *et al.*, "High-energy x-ray phase-contrast ct of an adult human chest phantom," *Scientific Reports*, vol. 15, no. 1, Aug. 2025.

[6] L. D'Amico, L. Costello, Y. Nesterets, *et al.*, "In situ propagation-based lung computed tomography for large animal models," *Journal of Synchrotron Radiation*, vol. 32, no. 6, pp. 1511–1522, Oct. 2025.

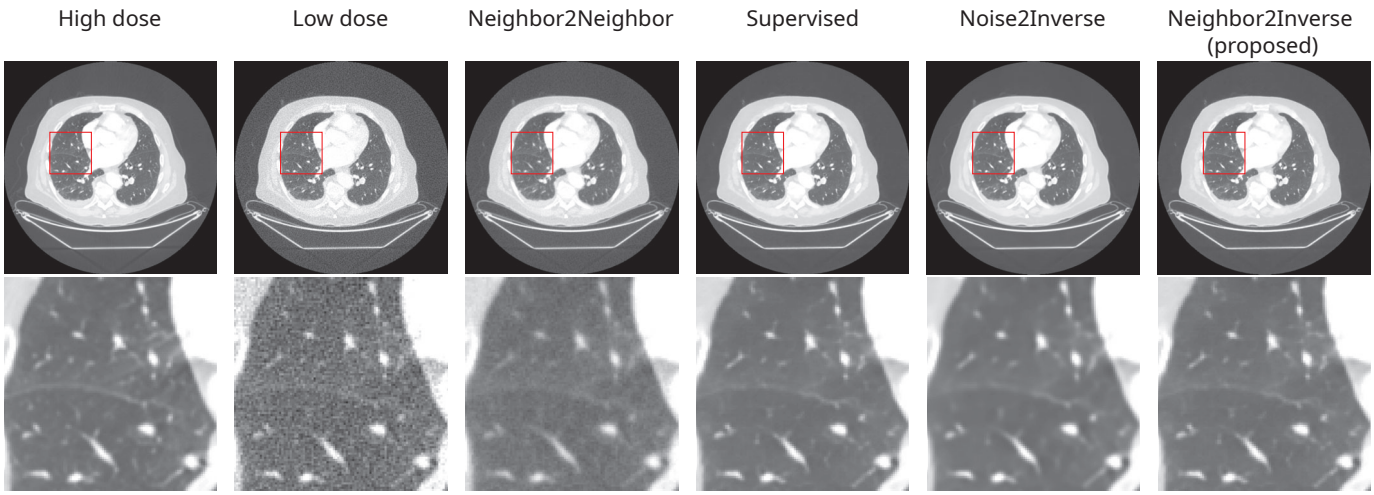


Fig. 3. Comparison of denoising methods applied to simulated low dose clinical CT with corresponding high dose reference. Denoising performed by the *Neighbor2Neighbor* approach [14] applied at the projection level, supervised denoising, the *Noise2Inverse* method [15] and the proposed *Neighbor2Inverse* approach. All images are displayed in lung window (-1350 to 150 HU). Top row: 512x512 pixel full image. Bottom row: 100x100 pixel zoomed-in region.

A Propagation-based X-ray phase-contrast Dataset (n=40)

Method	CNR↑	SR↓	Q↑
High dose	11.862 ± 2.119	8.398 ± 1.010	1.413 ± 0.304
Low dose	7.811 ± 1.515	8.699 ± 1.519	0.898 ± 0.234
Neighbor2Neighbor	9.902 ± 1.824	9.417 ± 1.484	1.051 ± 0.255
RecoFakeNoiseNet	17.203 ± 6.401	8.664 ± 1.318	1.985 ± 0.798
Noise2Inverse	8.470 ± 1.046	8.687 ± 1.105	0.975 ± 0.173
Neighbor2Inverse	16.178 ± 4.306	7.993 ± 1.471	2.024 ± 0.655

B Clinical Dataset (n=5,120)

Method	PSNR↑	SSIM↑
Low dose	34.52 ± 3.41	0.7541 ± 0.0963
Neighbor2Neighbor	41.14 ± 2.61	0.9479 ± 0.0196
Supervised	46.54 ± 2.41	0.984 ± 0.0068
Noise2Inverse	46.89 ± 2.41	0.9852 ± 0.0067
Neighbor2Inverse	44.39 ± 2.59	0.9795 ± 0.0079

best
worst

Fig. 4. (A) Denoising of measured low-dose PBI data, assessed using contrast-to-noise ratio (CNR), spatial resolution (SR), and the composite image quality metric Q=CNR/SR. (B) Denoising of simulated low-dose clinical CT data, evaluated using the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR). Depicted are mean values ± standard deviation.

[7] C. De Margerie-Mellon, C. De Bazelaire, and E. De Kerviler, *Image-guided biopsy in primary lung cancer: Why, when and how*, 2016.

[8] T. Melzer, C. M. Hackl, J. Walter, *et al.*, "Tumor seeding following CT- guided transthoracic needle biopsy in lung cancer. A case report," *BMC Pulmonary Medicine*, vol. 23, no. 1, 2023.

[9] Y. Zhang, L. Shi, M. J. Simoff, *et al.*, "Biopsy frequency and complications among lung cancer patients in the United States," *Lung Cancer Management*, vol. 9, no. 4, 2020.

[10] J. Zhang, W. Gong, L. Ye, *et al.*, *A Review of deep learning methods for denoising of medical low-dose CT images*, 2024.

[11] M. Han, H. Shim, and J. Baek, "Perceptual CT Loss: Implementing CT Image Specific Perceptual Loss for CNN-Based Low-Dose CT Denoiser," *IEEE Access*, vol. 10, 2022.

[12] Q. Yang, P. Yan, Y. Zhang, *et al.*, "Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, 2018.

[13] J. Lehtinen, J. Munkberg, J. Hasselgren, *et al.*, "Noise2Noise: Learning image restoration without clean data," in *35th International Conference on Machine Learning, ICML 2018*, vol. 7, 2018.

[14] T. Huang, S. Li, X. Jia, *et al.*, "Neighbor2Neighbor: Self-supervised denoising from single noisy images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2021.

[15] A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg, "Noise2Inverse: Self-Supervised Deep Convolutional Denoising for Tomography," *IEEE Transactions on Computational Imaging*, vol. 6, 2020.

[16] E. Colak, F. C. Kitamura, S. B. Hobbs, *et al.*, "The RSNA Pulmonary Embolism CT Dataset," *Radiology: Artificial Intelligence*, vol. 3, no. 2, 2021.

[17] M. Ronchetti, "Torchradon: Fast differentiable routines for computed tomography," *arXiv preprint arXiv:2009.14788*, 2020. eprint: arXiv:2009.14788.

[18] I. Gulrajani, F. Ahmed, M. Arjovsky, *et al.*, "Improved training of wasserstein GANs," in *Advances in Neural Information Processing Systems*, vol. 2017-December, 2017.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, 2015.

[20] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proceedings - 2017 IEEE Winter Conference on Applications of Computer Vision, WACV 2017*, 2017.

[21] J. Ahlers, *Fileswell*, 2025.

Bridging the Simulation-to-Reality Gap in Snapshot CAXRDT: A Confidence-Guiding Fine-Tuning Approach

Shengzi Zhao, Donghang Miao, and Yuxiang Xing, *Fellow, IEEE*

***Abstract*—Deep learning has demonstrated great performance in the field of medical imaging. However, in some scenarios such as snapshot coded aperture X-ray diffraction tomography (SCA-XRDT), acquiring large-scale real-world datasets for training is prohibitively difficult. Although training with simulated data offers a viable alternative, such networks often suffer from performance degradation when generalizing to real-world data due to the inevitable domain gap. In this work, we first investigate the discrepancies between simulated and real-world data through a comparative study using identical experimental setups. To address the problem that the cross-domain consistency exhibits significant pixel-wise variation, we propose a confidence-guiding (CG) module coupled with a two-step fine-tuning strategy (CGTune). By extracting features from the back-projection of confidence-weighted data, the module enables the network to discern the impact of data with varying degrees of simulation-to-reality discrepancy. This mechanism effectively mitigates the image quality degradation caused by such discrepancy while simultaneously enhancing data utilization. Simulation experiments demonstrate the effectiveness of the proposed method. It effectively eliminates structural artifacts when applied the CG tuned network to target data and improve the accuracy of reconstructed XRD patterns.**

***Index Terms*—deep learning, X-ray diffraction tomography, confidence-guiding module, two-stage finetuning.**

I. Introduction

DATA-DRIVEN deep learning (DL) methods have recently demonstrated good performance in medical imaging. By learning implicit representations from large-scale datasets, these methods effectively exploit underlying data patterns to improve the quality of reconstructed images. Nevertheless, in some scenarios, acquiring sufficiently large and representative real training datasets remains challenging, particularly for emerging imaging modalities or systems such as X-ray diffraction tomography (XRDT). This severely limits the applications of DL methods. A practical solution is to first generate a large volume of problem-specific simulated data using Monte Carlo (MC) simulations to train a network, and then develop generalization strategies to enhance its performance on real-world data. Fundamentally, this is a domain adaptation problem, where analyzing both the shared characteristics and the discrepancies between the source and target domains could be the key and incorporating the key information into method design could be critical.

There are various studies addressing the domain gap problem in the field of medical imaging. Darestani *et al*. proposed to pretrain a reconstruction network on a source domain P with supervision and subsequently perform unsupervised test-time fine-tuning on a target domain Q using only measurement consistency. This strategy substantially reduces the performance degradation caused by domain gaps in compressed sensing reconstruction [1]. Du *et al*. proposed a low-level feature-space unsupervised domain adaptation regularization that aligns source- and target-domain features extracted by a backbone network trained on simulated data. The method could transfer features learned from the simulated dataset to real-world data, thereby enhancing the network's generalization capability [2]. Barbano *et al*. proposed the Steerable Conditional Diffusion to synchronously update the model parameters during the generation process of a conditional diffusion model by enforcing constraints derived from the measurement data. The method significantly improves out-of-distribution reconstruction robustness across multiple medical imaging modalities [3]. Chen *et al*. proposed a significance score to quantify the importance of parameters in DL models, and enables a very efficient and robust group-to-individual adaptation for high quality patient-specific reconstruction for various scenarios in CT imaging [4]. These approaches focus on the differences between training and out-of-distribution test data in terms of image structures, such as artifact patterns and anatomical features. The fine-tune strategies enable neural networks to better leverage shared characteristics across domains while simultaneously adapting to the unique features of test data.

In this work, we study on a snapshot coded aperture XRDT (SCA-XRDT) system that is capable of retrieving the spatial distribution of XRD patterns on a cross-section of an object from single-view diffraction data. Based on Monte Carlo (MC) simulations [5], we developed a dual-domain deep learning framework for diffraction tomography reconstruction (DuDo-DTR) for the SCA-XRDT system, achieving good reconstruction results with a spatial resolution of ~2 mm [6].

This paper was submitted to CT meeting for review on Jan. 25th, 2026. This work was supported by National Natural Science Foundation of China under NNSFC 12275151. Corresponding author*: Yuxiang Xing (xingyx@mail.tsinghua.edu.cn). This work was also partially supported by Center of High Performance Computing, Tsinghua University.

However, it remains challenging to directly apply this method to real-world data. The primary difficulty is the absence of available real-world datasets for direct training as SCA-XRDT is a new imaging system. Moreover, large-scale phantom fabrication and scanning are too expensive and difficult to implement. To address this problem, this work investigates generalization strategies to adapt the model trained on simulated data to real-world scenarios, reducing the reliance of deep learning methods on expensive experimental data and promotes their application in the field of XRDT.

In the following, we first provide a brief overview of the forward model of a SCA-XRDT system and our deep learning framework. Then, we analyze the primary discrepancies between simulated and real-world data, based on which we propose a confidence-guiding (CG) module and a two-stage fine-tuning strategy (CGTune) to facilitate the network to generalize effectively from training data to target data. Finally, we validate the method using simulation experiments.

II. Method

A. The forward model of a SCA-XRDT system.

In an SCA-XRDT system shown in Figure 1, X-rays generated by the source are collimated into a fan beam by a slit collimator to irradiate the object. The incident photons undergo Rayleigh scattering, emerging at specific angles relative to the incident direction. Subsequently, the scattered photons are coded by a coded aperture and finally collected by the detector. The detected photons could be modeled as:

$$I(\vec{u},E)=\int R(E,E')\int_q\int_{\vec{x}} I_0(E')T(\vec{x},E',\vec{u})\mathrm{N_A}\frac{\mathrm{d}\sigma}{\mathrm{d}\theta_s}\times C(\vec{x},\vec{u})\Omega_s(\vec{x},\vec{u})f(\vec{x},q)\delta\left(q-\frac{E'}{\mathrm{hc}}\sin\frac{\theta_s}{2}\right)\mathrm{d}\vec{x}\mathrm{d}q\mathrm{d}E'. \quad (1)$$

Here, $I_0(E)$ is the spectrum of the X-ray source, $T(\vec{x},E,\vec{u})$ the attenuation by the object, $\mathrm{N_A}$ the Avogadro Constant, $\mathrm{d}\sigma/\mathrm{d}\theta_s$ the Thomson differential cross-section, θ_s the scattering angle, hc the product of Plank Constant and the speed of light, $\Omega_s(\vec{x},\vec{u})$ the scattering solid angle, and $C(\vec{x},\vec{u})$ the coded aperture. The function$\delta(q-E/\mathrm{hc}\sin(\theta_s/2))$ filters the scattered photons and $f(\vec{x},q)$ is the distribution of XRD patterns.

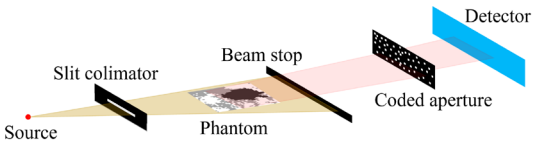


Figure 1. The diagram of the simulated SCA-XRDT system.

B. The overview of the DuDo-DTR

The SCA-XRDT system utilizes a pixelated energy-dispersive photon counting detector fabricated from semiconductor materials. Therefore, we use one network module (referred as Net-1) to compensate for non-ideal factors within the data, including charge sharing, noise, and multiple scattering. We use back-projection operator for transforming compensated data to a spectral image. The resulted image could be of low-quality with unsatisfying structure recovery. Thus, we use another network module (referred as Net-2) to further recover images details. The overall framework DuDo-DTR is illustrated in Figure 2.

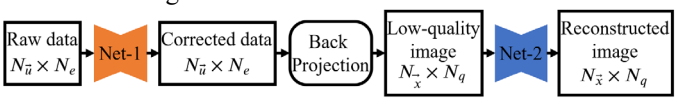


Figure 2. The diagram of the dual-domain deep learning framework for the diffraction tomography reconstruction.

C. Analysis of discrepancies between simulated and real-world data

Simulation tools have been devoted to model real-world data, but it is impossible to catch all the physical factors and variations in a real system. To visualize the discrepancies between simulated and real-world data, we conducted a comparative experiment using the simulation program described in [5] and an SCA-XRDT system in our laboratory. In the simulation program, we assigned identical parameters to all detector pixels and tuned the simulated detector response functions to closely match the real-world response matrix measured following the method in [7]. An identical phantom in both the simulated and real-world SCA-XRDT systems consisted of a water-filled PMMA cylinder with an outer diameter of 60 mm and an inner diameter of 58 mm. Additionally, identical geometric parameters were configured for both systems. After scanning, we performed a pixel-wise comparison of the acquired data and quantified the differences using the correlation coefficient defined as follows:

$$\rho_i(S_{\mathrm{simu}},S_{\mathrm{real}})=\frac{\sum_{i_E}S_{\mathrm{simu}}(i_E)S_{\mathrm{real}}(i_E)\Delta E}{\left[\left(\sum_{i_E}S^2_{\mathrm{simu}}(i_E)\,\Delta E\right)\left(\sum_{i_E}S^2_{\mathrm{real}}(i_E)\,\Delta E\right)\right]^{\frac{1}{2}}}. \quad (2)$$

Here, i and i_E index the detector pixels and energy bins. S_{simu} and S_{real} are the spectra acquired by the i^{th} pixel in simulation and real-world experiment, respectively.

Figure 3 compared results between simulated and real-world data. As shown in Figure 3(a), the simulated data exhibits good agreement with the real-world data in some region. The differences between the two data increases as the correlation coefficient decreases. Big deviation is primarily observed in the low-energy region, where the real-world data are of higher counts than the simulation data. In Figure (b), although the response functions of the simulated and real detectors are very close, the degree of agreement between simulated and real-world data varies significantly across different pixels. It is primarily attributed to the non-uniformity of the real detector pixels, which is difficult to accurately model in the simulation. In Figure 2(c), we apply a threshold of 0.99 to classify all pixels into two categories with white pixels meaning their $\rho > 0.99$. For our real SCA-XRDT system, data from more than half of the pixels have their $\rho > 0.99$, and the network can effectively handle the small difference. For the black pixels, the data contain useful information about the object but do not accord with simulation scenario well. Hence, the key challenge is how to leverage the information.

We take simulated data as the source domain for network training, aiming for robust generalization to the target domain of real-world data. The domain alignment exhibits significant pixel-wise variation. Inspired by WSL that assigns weights to

data to effectively utilize data of different noise level and quality, we introduce a data confidence mechanism into the deep-learning reconstruction framework, enabling the model to effectively utilize the data from all pixels.

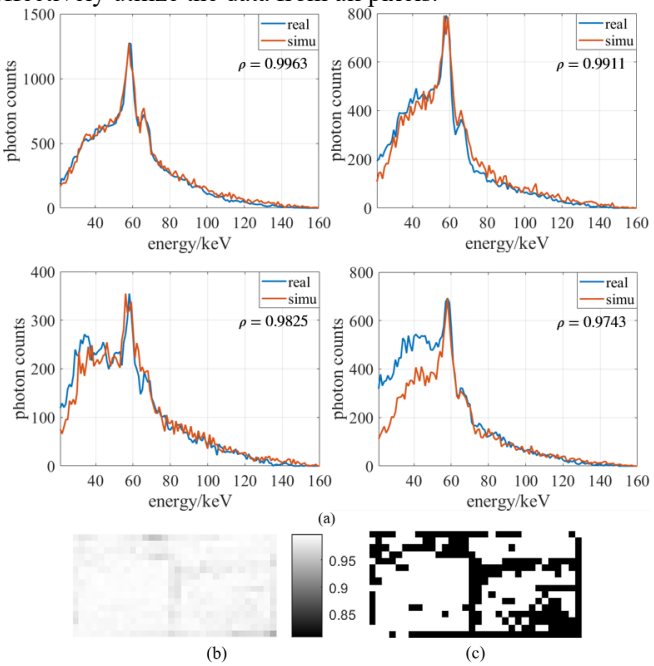


Figure 3. Comparison between simulated and real-world data. (a) Spectra measured at four selected pixels and their corresponding correlation coefficients. (b) Correlation coefficients for all pixels. (c) Spatial distribution of pixels with correlations coefficients larger than 0.99 (white pixels).

D. CGTune: CG module and finetune strategies

We incorporate a CG module into the image reconstruction network Net-2 in DuDo-DTR, as shown in Figure 4. The backbone of the Net-2 is a ResUNet receiving $\mathrm{A}^{\mathrm{T}}\mathbf{y}^{\mathrm{Net1}}$ as input, where A^{T} is the back-projection operator and $\mathbf{y}^{\mathrm{Net1}}$ is the corrected data from Net-1. The CG module in the red box utilizes $\mathrm{A}^{\mathrm{T}}(\mathbf{w}\odot\mathbf{y}^{\mathrm{Net1}})$ as input to extract joint features of confidence and data, and these features are then incorporated into the backbone. Here, $\mathbf{w}$ is the confidence matrix and $\odot$ the Hadamard product.

In the SCA-XRDT system, data collected by a single detector channel originates from multiple object pixels. Consequently, measurement errors are propagated across these corresponding pixels in the images. The module transforms the data confidence into the image domain and enables the network to localize structures influenced by low-confidence data by comparing the two inputs. It is helpful for the Net-2 to mitigate the impacts on reconstructed images due to differences between simulated and real-world data.

We train the CG module utilizing three datasets: Ω_{src}, Ω_{brdg} and Ω_{tgt}. Firstly, Ω_{src} is source domain dataset that are simulated and used to pre-train the Dudo-DTR without CG module. Secondly, we conjecture the higher photon counts in low-energy region of real-world data could be caused by difference in detector electric field, carrier mobility, electronic noise, etc. Hence, we produced another dataset Ω_{brdg} by adjust detector parameter to simulate this phenomenon. Finally, Ω_{tgt} is the target domain dataset that often consists of only a limited set of real-world cases.

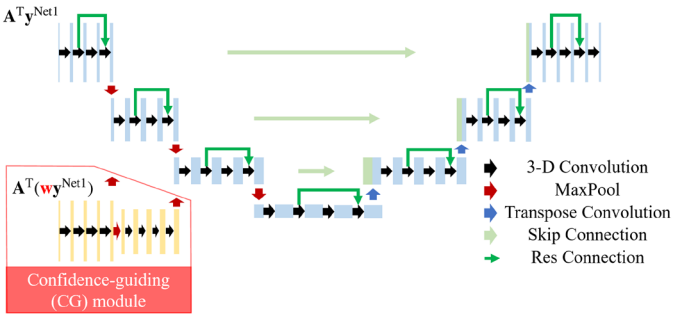


Figure 4. The diagram of the CG module and Net-2.

We conduct 'two-step finetuning' for the CG module. In the first step, we freeze the backbone parameters and train the CG module using a mixture dataset Ω_{mix} constructed from Ω_{src} and Ω_{brdg}. For a given phantom, the mixing process is obtained by:

$$\mathbf{I}_{i,\mathrm{mix}} = \boldsymbol{\alpha}_i \odot \mathbf{I}_{i,\mathrm{src}} + (\mathbf{1} - \boldsymbol{\alpha}_i) \odot \mathbf{I}_{i,\mathrm{brdg}}. \quad (3)$$

Here, $\boldsymbol{\alpha}_i \in [0,1]$, $I_{i,\mathrm{src}} \in \Omega_{\mathrm{src}}$, $I_{i,\mathrm{brdg}} \in \Omega_{\mathrm{brdg}}$ and $I_{i,\mathrm{mix}} \in \Omega_{\mathrm{mix}}$ with i indexing a phantom. For each training data, the confidence matrix $\mathbf{w}_i$ in the CG module is set to $\boldsymbol{\alpha}_i$. This is to bridge the gap between source and target domains by enabling the network to learn the impact of data with varying confidence levels on reconstructed images, and equipping it with the capability to correct location-specific errors between outputs and labels. In the second step, we freeze the parameters of the CG module and employ Low-Rank Adaptation (LoRA) [8] to finetune the confidence matrix $\mathbf{w}$ using Ω_{tgt}.

E. Configurations of the experiment

We construct our phantoms based on the CBIS-DDSM dataset [9], with each composed of three components: adipose, glandular and cancerous tissues. The datasets Ω_{src} and Ω_{brdg} were generated via MC simulations [5]. Because we do not have real-world available and using simulated target domain data will be convenient for method performance evaluation, we construct a separate dataset $\Omega_{\mathrm{tgt}}^{\mathrm{simu}}$ of 10 phantoms different from those used in the first step with detector parameters of some pixels different from Ω_{src} and Ω_{brdg}.

III. Experimental Results

A. The results of the first finetuning step

We show the results obtained after the first fine-tuning in Figure 5. For the three test phantoms, the trained backbone network yields good reconstruction results. However, when the network is tested on the mixed dataset Ω_{mix}, significant structural errors emerge in the reconstructed images. Specifically, in the boxed region of Phantom 1 and the area indicated by the arrow in Phantom 2, glandular tissues are erroneously reconstructed as cancerous tissues. Conversely, within the box in Phantom 3, cancerous tissues are incorrectly reconstructed as glandular tissues. Upon incorporating the CG module and fine-tuning the network, the aforementioned reconstruction artifacts are significantly suppressed.

Furthermore, when evaluating the fine-tuned network on dataset Ω_{src}, the reconstruction quality remains highly consistent with the ground truth. This demonstrates that the network effectively processes diverse data based on confidence levels, thereby mitigating the impact of data discrepancies on the final results. The XRD patterns at the blue points for phantom 1 and 3 presented in Figure 5(b) further demonstrate the effectiveness of our method.

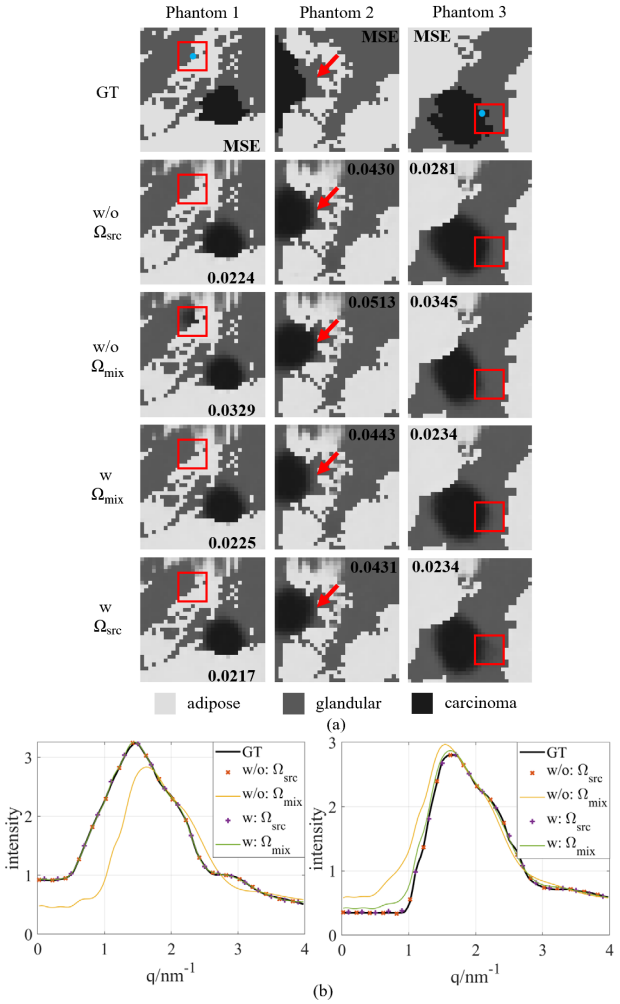


Figure 5. The results of the first finetuning. (a) Tomography at $q = 1.1$ nm^{-1}. Ω_{src} and Ω_{mix} represents the test datasets. "w" and "w/o" mean the results from the network with and without CG module, respectively. (b) The XRD patterns at the blue points for phantom 1 and 3.

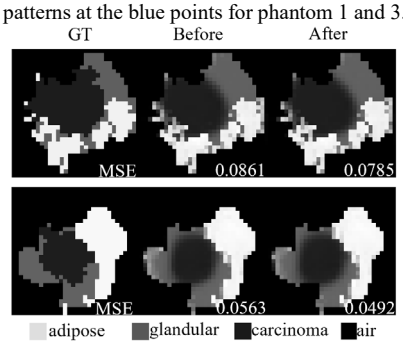


Figure 6. Results of the second fine-tuning.

B. The results of the second finetuning step

Figure 6 shows the results of the second fine-tuning stage. Following the fine-tuning of the confidence matrix using dataset Ω_{tgt}^{simu}, the MSE of the reconstruction is reduced. However, in terms of image structures, the fine-tuned results exhibit no significant improvement. This may suggest that MSE may be insufficient for accurately evaluating the quality of XRDT images. Another potential reason is that we currently employ only six phantoms to fine-tune **w**, resulting in severely insufficient data diversity.

IV. Discussion & Conclusion

For SCA-XRDT systems, acquiring large-scale real-world datasets for network training is extremely challenging. Consequently, training networks on simulated data and generalizing them to real data is a feasible approach. However, due to the challenge of accurately modeling each detector pixel, the agreement between simulated and real-world data differs and even poor for some pixels. This leads to performance degradation when generalizing the network trained on simulated data to real-world data. We propose a CG module coupled with a two-step fine-tuning strategy (CGTune), to help the network fully exploit the information contained within the data of every detector pixel and mitigate the degradation in reconstruction quality. The results demonstrate that this module significantly mitigates the impact of cross-domain discrepancies on reconstruction quality. In future work, we are to increase the size of the training set for the second fine-tuning and explore additional methods to improve the effectiveness of the fine-tuning process.

In this work, we propose a CG module to assist the network in handling the discrepancies between the training dataset and the target dataset intended for generalization. Simulation experiments demonstrate the effectiveness of the proposed method. In the future, we will continue to refine this approach and evaluate the results using real-world data in the second fine-tuning.

References

[1] Darestani MZ, et al. Test-time training can close the natural distribution shift performance gap in deep learning based compressed sensing. International conference on machine learning, 2022: PMLR, 4754-4776.
[2] Du MG, et al. Deep-learning-based metal artefact reduction with unsupervised domain adaptation regularization for practical CT images. IEEE Transactions on Medical Imaging. 2023;42(8): 2133-2145. Doi: 10.1109/TMI.2023.3244252.
[3] Barbano R, et al. Steerable Conditional Diffusion for Out-of-Distribution Adaptation in Medical Image Reconstruction. IEEE Transactions on Medical Imaging. 2025;44(5): 2093-2104. Doi: 10.1109/TMI.2024.3524797.
[4] Chen CY, et al. PWLS-SOM: alternative PWLS reconstruction for limited-view CT by strategic optimization of a deep learning model. Physics in Medicine & Biology. 2025;70(18): 185011. Doi: 10.1088/1361-6560/adffe0.
[5] Shen L, et al. Efficient GPU-accelerated Monte Carlo simulation for x-ray diffraction imaging with spectroscopic detector modeling. Medical Imaging 2025: Physics of Medical Imaging. Vol. 13405. SPIE, 2025.
[6] Zhao SZ, et al. A deep learning reconstruction framework for snapshot coded aperture X-ray diffraction tomography. 18th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, 2025.
[7] Liang KC, et al. Pixelated Energy-dispersive Photon counting Detector Calibration with X-ray Diffraction. IEEE NSS/MIC/RSTD 2022, Milano, Italy.
[8] Hu EJ, et al., Lora: Low-rank adaptation of large language models, ICLR, 2022.
[9] Sawyer-Le R., et al. Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM). Cancer Imaging Arch, 2016. DOI: 10.7937/K9/TCIA.2016.7O02S9CY.

Geometric Calibration for X-ray Fluorescence Tomography using Multi-Support Consistency

Meo van Duijnen Montijn, Jérôme Lesaint, Jean Michel Létang, and Simon Rit

***Abstract*—We propose a novel geometric calibration approach for computed tomography (CT) based on data consistency conditions that rely solely on sinogram supports, i.e., the regions where the sinogram values are non-zero. This approach enables the estimation of shifts, from which the center of rotation may also be directly inferred. The approach is agnostic to attenuation and applies to sets of supports arising from different spatial, spectral, and other measurement domains. While the formulation is general, X-ray fluorescence CT serves as the primary application considered in this work. Numerical simulations demonstrate the accuracy of the proposed approach.**

I. Introduction

GEOMETRIC calibration is a prerequisite for accurate reconstructions in all computed tomography (CT) modalities. Errors in the acquisition geometry—such as misestimated centers of rotation or uncorrected sample motion—cause reconstruction artifacts and loss of resolution. In transmission-based tomography, analytical calibration approaches grounded in data consistency conditions (DCCs), most notably those derived from the Helgason-Ludwig conditions [1], provide data-driven means to address these issues. Such methods enable center-of-rotation determination and shift correction directly from projection data [2], eliminating the need for external markers or heuristic alignment procedures.

However, Helgason-Ludwig DCCs rely on assumptions that may be violated in modalities governed by multiple interaction mechanisms, most notably X-ray fluorescence CT (XFCT). In X-ray fluorescence, atoms are excited through X-ray irradiation and subsequently emit secondary fluorescence photons isotropically. In XFCT, fluorescence photons emitted following X-ray excitation are recorded as functions of translation and rotation and, after spectral fitting, yield elemental sinograms from a single measurement, as illustrated in Fig. 1. A major difficulty, however, is that both the incident beam and the emitted fluorescence are attenuated along their respective paths, thereby eliminating the general validity of existing analytical calibration approaches.

Crucially, although attenuation alters sinogram intensities, it may not affect the sinogram support, so calibration-relevant information may be preserved even in the presence of unknown attenuation. Early work showed that single-sinogram support boundaries could be used for coarse geometric calibration [3]. More recently, accurate calibration has been achieved by heuristically matching large sets of supports via iterative reprojection [4], but at high computational cost, as repeated image reconstructions are required, and without an analytical foundation. Here, we extend upon the DCCs introduced in [3] to obtain an analytical, attenuation-agnostic calibration approach using sets of supports. The proposed approach enables direct estimation of geometric shifts and the center of rotation. Although the formulation is general, XFCT is used as the primary validation modality.

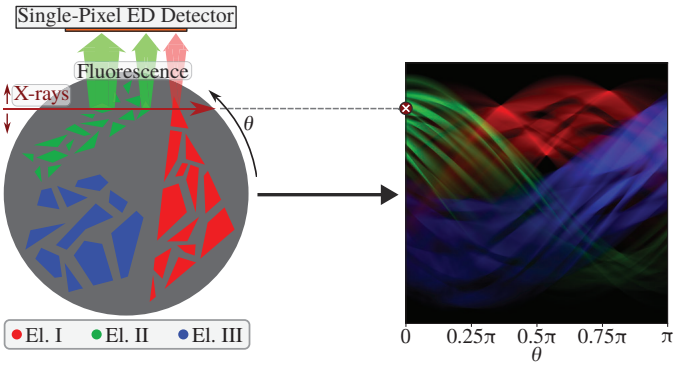


Fig. 1. XFCT: (a) Experimental setup in which an X-ray pencil beam traverses the sample at each projection angle and excites atoms in a phantom comprising distinct elements (here, three), generating fluorescence photons that are recorded by a single-pixel energy-dispersive (ED) detector positioned normal to the beam direction. (b) Combined X-ray fluorescence sinograms exhibiting attenuation effects. The white cross indicates the pencil beam position in (a).

II. Mathematics

In XFCT, one collects sinograms for different elements and possibly at different out-of-plane offsets. From each sinogram, a sinogram support $Y \subset \mathbb{R}^2$, defined by its nonzero entries, can be extracted, together forming the set $\mathbb{Y}$. Each sinogram support Y in this set is determined by the projection of a convex envelope $X \subset \mathbb{R}^2$ in object space, with the corresponding set of envelopes denoted by $\mathbb{X}$, and is related through the mapping $\mathcal{R}_{\text{sup}} : X \mapsto Y$ as illustrated in Fig 2. For each element of $\mathbb{Y}$, the outer sinogram boundaries $h_Y^+(\theta)$ and $h_Y^-(\theta)$ may be extracted, denoting the lower and upper borders, respectively, in the direction of translation. Although

M.Q. van Duijnen Montijn is with ESRF, The European Synchrotron, 71 Avenue des Martyrs, CS 40220, 38043 Grenoble Cedex 9, France, and with INSA Lyon, Université Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621 Villeurbanne, France (e-mail: meo.van-duijnen-montijn@esrf.fr).

J. Lesaint is with ESRF, The European Synchrotron, 71 Avenue des Martyrs, CS 40220, 38043 Grenoble Cedex 9, France (e-mail: jerome.lesaint@esrf.fr).

J.M. Létang and S. Rit are with INSA Lyon, Université Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69621 Villeurbanne, France (e-mail: jean.letang@insa-lyon.fr; simon.rit@creatis.insa-lyon.fr).

Manuscript received January 28, 2026.

additional boundaries may in principle be identified, they are not considered in this work.

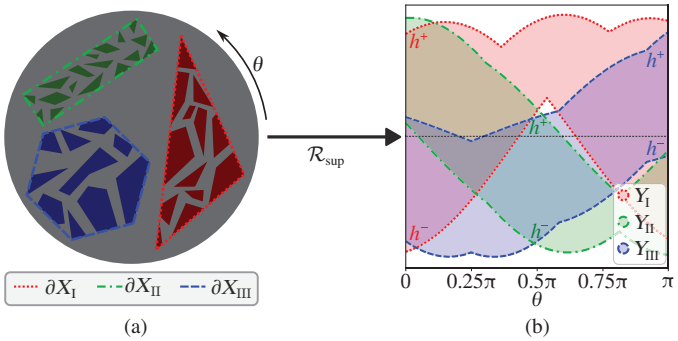


Fig. 2. Convex envelopes (X), support sinograms (Y) and their mapping: (a) Three spatial domains containing different fluorescing elements, with the borders of their convex envelopes (∂X) highlighted. (b) The corresponding fluorescence support sinograms and their upper and lower boundaries are denoted h^+ and h^-, respectively.

The main sinogram boundaries are traced, for each projection angle, by the extreme points of the object's convex envelope in the translational dimension. The boundaries of the convex envelope in object space are therefore directly related to the boundaries of its support sinogram. Since the object-space boundaries correspond to those of a convex shape, their radii of curvature do not change sign between different boundary points on the same side of the sinogram. It was shown in [3] that these radii of curvature are related to the boundaries of the support sinograms via

$$\rho_X^{\pm}(\theta) = \mathcal{L} h_Y^{\pm}(\theta), \quad \pm\rho_X^{\pm}(\theta) \geq 0, \tag{1}$$

where the linear operator is defined as

$$\mathcal{L} := \mathrm{Id} + \frac{\partial^2}{\partial\theta^2}, \tag{2}$$

and Id is the identity operator. The functions $\rho_X^+(\theta)$ and $\rho_X^-(\theta)$ correspond to the $+$ and $-$ branches of Eq. 1 and denote the radii of curvature at the upper and lower extreme points, respectively, of the convex envelope X at angle θ. Boundaries that are not twice differentiable are interpreted in the sense of distributions. Two examples illustrating the validity of Eq. 1 are shown in Fig 3.

A. Perturbations

In-plane perturbations manifest themselves as angle-dependent shifts $\Delta(\theta)$ in the translational dimension, yielding perturbed sinogram boundaries

$$\tilde{h}_Y^{\pm}(\theta) := h_Y^{\pm}(\theta) + \Delta(\theta). \tag{3}$$

Such perturbations may violate the DCCs of Eq. 1. However, enforcing these conditions by applying shifts to a perturbed sinogram does not, in general, lead to a unique solution. For each sinogram Y, there may exist an entire set of admissible shifts $\mathbb{A}_Y$ that restore the DCCs when applied. Since the true shifts are identical for all Y and are contained in each of these sets, it follows that

$$\Delta(\theta) \in \mathbb{A} := \bigcap_{Y\in\mathbb{Y}} \mathbb{A}_Y, \tag{4}$$

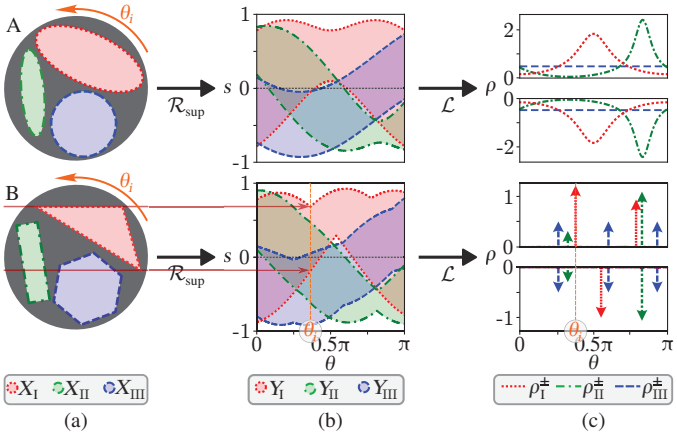


Fig. 3. Radii of curvature for two phantoms with a highlighted case. (a) Phantoms A and B, exhibiting jagged and smooth convex envelopes X of fluorescent supports, respectively, within a gray capillary. The two pencil beams correspond to irradiated extreme points of X_{I} for phantom B at angle θ_i with infinity (top) and zero (bottom) radii of curvature. (b) Support sinograms. The vertical dashed line highlights Y_{I} of phantom B at angle θ_i, featuring a cusp and a sinusoidal segment at the boundaries. (c) Corresponding radii of curvature (ρ^+ and ρ^- in top and bottom plots, respectively), with arrows indicating Dirac delta contributions.

where $\mathbb{A}$ denotes the ambiguity in $\Delta(\theta)$ when considering the full set $\mathbb{Y}$ of sinogram supports Y.

B. Ambiguity analysis

Combining Eqs. 1 and 3 yields the relation

$$\pm\rho_X^{\pm}(\theta) = \pm\mathcal{L}\left[\tilde{h}_Y^{\pm}(\theta) - \Delta(\theta)\right] \geq 0. \tag{5}$$

Taking both signs of this equation into account then leads to

$$\mathcal{L}\tilde{h}_Y^{-}(\theta) \leq \mathcal{L}\Delta(\theta) \leq \mathcal{L}\tilde{h}_Y^{+}(\theta) \quad \forall Y \in \mathbb{Y}. \tag{6}$$

The sets of admissible shifts $\mathbb{A}_Y$ thus comprises all functions $\Delta(\theta)$ that satisfy this condition. The properties of this set have been investigated: any two entries differ only in their odd-frequency Fourier components, whose amplitudes are constrained and mutually linked. Evaluating this set, for example, for the case in which X is a circle, yields the ambiguity in the object domain corresponding to the family of Reuleaux polygons [5], which themselves, consistently, exist only for odd orders. These and related properties are exploitable, but are left for future work. Since $\mathbb{A}_Y$ may differ for each Y, an increasing number of sinograms in the set $\mathbb{Y}$ may progressively constrain the ambiguity in $\Delta(\theta)$ via Eq. 4, although not every additional element necessarily leads to a further reduction. Neglecting a fundamental harmonic component in the shifts, which merely induces a global displacement of the reconstruction, a criterion can be derived under which the intersection of admissible shifts reduces to a singleton. For this to occur, the set $\mathbb{X}$ must contain objects X with extreme points having zero-radius of curvature for a sufficient amount of projection angles. E.g., this is satisfied by phantom B alone (Fig 3). If $\mathbb{X}$ fails to satisfy this criterion for uniqueness—which will likely be difficult to assess in practice—an ambiguity in the shifts remains. Nevertheless, this ambiguity may still be sufficiently constrained for practical purposes and may furthermore be determined using only the perturbed projection data.

III. Methods

XFCT acquisitions were simulated on two phantoms, as illustrated in Fig 3a, using 256 translational steps and 360 rotational steps over a half-turn. Each phantom contained three convex envelopes emitting different fluorescence, within an attenuating capillary. The exact separation, whether in the spatial or energy domain, is irrelevant for the approach, provided that the supports are distinguishable, however. The simulations were carried out using the `PyCorrectedEmissionCT` package (https://pypi.org/project/corrct/) [6]. Fluorescence sinograms were generated by assigning three arbitrarily chosen fluorescence-emitting elements (I, II, III) to the supports, and by attaching attenuation maps to both the supports and the capillary. Shifts were modelled as semi-random combinations of trigonometric functions superimposed on noise. The sinograms were subsequently perturbed by applying these shifts through linear interpolation of pixel intensities, before their pixel-accurate boundaries were extracted. The applied shifts and the resulting perturbed support sinograms are illustrated in Fig. 4.

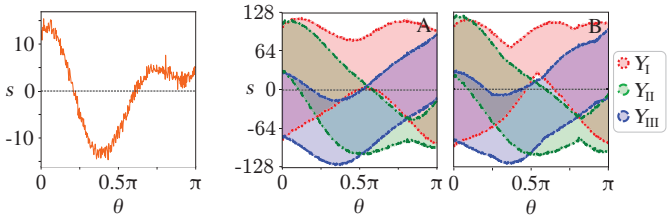


Fig. 4. (a) Simulated perturbation and (b) corresponding perturbed support sinograms for phantoms A and B.

Determining a solution for the shifts under the constraint of Eq. 6 requires evaluating second-order derivatives of the perturbed boundary positions. Direct computation of these derivatives using finite differences, however, is numerically unstable. In the sinogram domain, each point in object space traces a fundamental (co)sine of period 2π. For every boundary point, there must therefore exist a cosine that passes through it and remains entirely within the sinogram support. This condition corresponds to the edge case of Eq. 1, for which $\mathcal{L}h_Y^{\pm}(\theta) = 0$. Each such cosine has only a single degree of freedom, as it is constrained to pass through a given boundary point. Accordingly, for the perturbed boundary points of a sinogram Y at projection index i, after applying a candidate shift ($\hat{\Delta}_i$), this function can be expressed as a cosine with relative phase ϕ via

$$m_{Y,i}^{\pm}(\hat{\Delta}_i, \phi_{Y,i}^{\pm}, \theta) := \frac{\tilde{h}_{Y,i}^{\pm} - \hat{\Delta}_i}{\cos \phi_{Y,i}^{\pm}} \times \cos(\theta - \theta_i + \phi_{Y,i}^{\pm}), \quad (7)$$

where the amplitude (in front of the cosine) is chosen such that the function passes through the boundary point, as illustrated in Fig. 5a. The objective is to minimize violations where these functions leave the sinogram support. Denoting the discrete arrays of the perturbed boundaries, shifts, and projection angles by $\tilde{\mathbf{h}}_Y^{\pm}$, $\hat{\boldsymbol{\Delta}}$ and $\boldsymbol{\theta}$, respectively, these violations are quantified by the violation vectors

$$\begin{aligned}\mathbf{v}_{Y,i}^{\pm}(\hat{\boldsymbol{\Delta}}, \phi_{Y,i}^{\pm}) := \\ \max\left(\mathbf{m}_{Y,i}^{\pm}(\hat{\Delta}_i, \phi_{Y,i}^{\pm}, \boldsymbol{\theta}) - (\tilde{\mathbf{h}}_Y^{+} - \hat{\boldsymbol{\Delta}}), \mathbf{0}\right) \\ + \min\left(\mathbf{m}_{Y,i}^{\pm}(\hat{\Delta}_i, \phi_{Y,i}^{\pm}, \boldsymbol{\theta}) - (\tilde{\mathbf{h}}_Y^{-} - \hat{\boldsymbol{\Delta}}), \mathbf{0}\right),\end{aligned} \quad (8)$$

where the first and second terms track the function passing out of the upper and lower support boundaries, respectively. This then gives the entries and magnitudes of such violations,

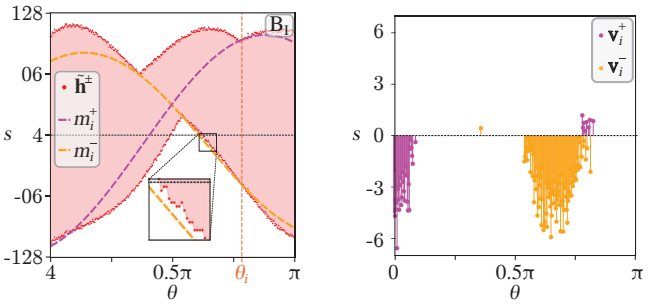


Fig. 5. Cosine fitting at angle θ_i for element I of phantom B: (a) Perturbed support sinogram with the best-fitting cosines ($m_i^{\pm}$). (b) Violations of the fitted cosines ($\mathbf{v}_i^{\pm}$), i.e. cosine exceedances as defined in Eq. 8, with zero-valued entries omitted.

as illustrated in Fig. 5b. These vectors must be minimized simultaneously over all projections and for all entries of $\mathbb{Y}$. To achieve this, we define a cost function depending on the candidate shift $\hat{\boldsymbol{\Delta}}$ and on the complete set of relative phases $\Phi := \{\boldsymbol{\phi}_Y^{+}, \boldsymbol{\phi}_Y^{-}\}_{Y \in \mathbb{Y}}$ as

$$C\left(\hat{\boldsymbol{\Delta}}, \Phi\right) := \sum_{Y,i} \left\|\mathbf{v}_{Y,i}^{+}(\hat{\boldsymbol{\Delta}}, \phi_{Y,i}^{+})\right\|_2^2 + \left\|\mathbf{v}_{Y,i}^{-}(\hat{\boldsymbol{\Delta}}, \phi_{Y,i}^{-})\right\|_2^2. \quad (9)$$

The best-fitting cosines may not verify Eq. 1 and this must therefore be incorporated as a constraint in the fitting procedure. Denoting $\hat{\mathbf{h}}_Y^{\pm} := \tilde{\mathbf{h}}_Y^{\pm} - \hat{\boldsymbol{\Delta}}$, the slopes of the cosines provide the first-order derivatives at their corresponding boundary points via

$$\frac{\partial}{\partial \theta} m_{Y,i}^{\pm}(\cdot, \theta_i) = -\hat{h}_{Y,i}^{\pm} \tan \phi_{Y,i}, \quad (10)$$

while the second-order derivatives can be obtained by finite differences of these first-order derivatives. To achieve this, the discrete second-order derivative operator $\mathbf{D}$ is defined as

$$\left(\mathbf{D}\hat{\mathbf{h}}_Y^{\pm}\right)_i := \frac{\frac{\partial}{\partial \theta} m_{Y,i+1}^{\pm}(\cdot, \theta_{i+1}) - \frac{\partial}{\partial \theta} m_{Y,i-1}^{\pm}(\cdot, \theta_{i-1})}{\theta_{i+1} - \theta_{i-1}}, \quad (11)$$

with boundary conditions for $k \in \{-1, N\}$, where N denotes the number of projections, given by

$$\begin{aligned}(\hat{h}_{Y,k}^{\pm}, \phi_{Y,k}^{\pm}) = -\ (\hat{h}_{Y,k \bmod N}^{\mp}, \phi_{Y,k \bmod N}^{\mp}), \\ \theta_k = \theta_{k \bmod N} + \pi \operatorname{sgn}(k).\end{aligned}$$

This allows the definition of the estimated discrete radii of curvature and express its constraint via

$$\hat{\boldsymbol{\rho}}_X^{\pm} := (1 + \mathbf{D})\hat{\mathbf{h}}_Y^{\pm} \quad (12a)$$

$$\text{s.t.} \quad \pm \hat{\boldsymbol{\rho}}_X^{\pm} \geq \mathbf{0}. \quad (12b)$$

The cosines are fitted around a set zero in the translational domain, thereby fixing the center of rotation at this position and determining shifts relative to it. As an initial guess, all relative phases are set to zero (thus satisfying Eq. 12b), and the shifts are initialized as the mean boundary positions, making the performance independent of the shifts. The minimization of the cost function in Eq. 9 involves $N \times (1+2|\mathbb{Y}|)$ variables and therefore scales rapidly with the size of $\mathbb{Y}$. However, by approximating the min and max functions in Eq. 8 with smooth functions, highly sparse associated Jacobians and Hessians can be derived.

Although the precise method for boundary extraction is beyond the scope of this paper, we can investigate the impact of projection data noise on calibration performance by introducing random errors into the extracted boundaries. This is achieved by offsetting the boundary values using pixel-binned Gaussian perturbations, with varying standard deviations.

IV. Results

The shift estimation results for phantoms A and B are shown in Fig. 6. The algorithm shows similar performance on both phantoms. Most deviations from the GT are within one pixel; only at 3 and 1 of the 360 projections is this exceeded for phantoms A and B, respectively. Reconstructions from the

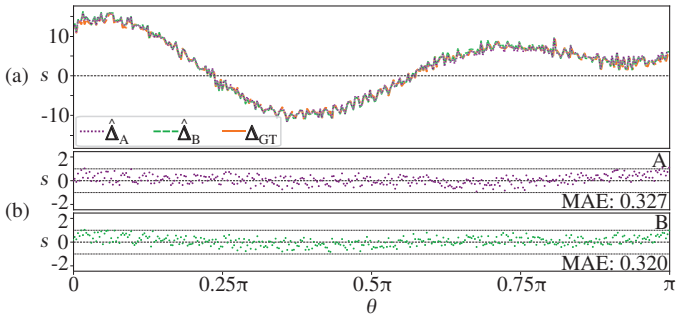


Fig. 6. Simulation results for shift estimations for phantoms A and B: (a) comparison between found and GT shifts, (b) residuals between found and GT shifts and their mean absolute error (MAE).

corrected sinograms, as shown in Fig. 7, show a substantial improvement over the uncorrected case and are close to the GT reconstructions. Attenuation artifacts are present in the reconstructions, as this effect is not corrected for.

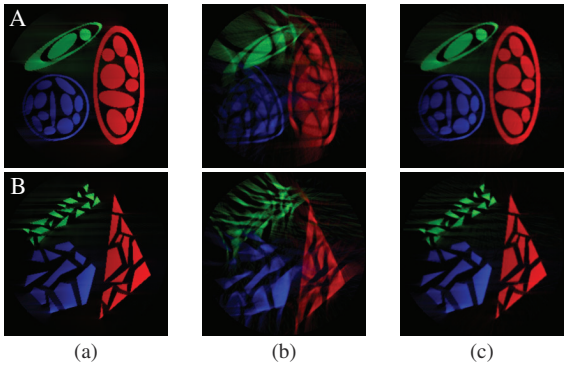


Fig. 7. Combined filtered backprojections of both phantoms: (a) using the GT geometry, (b) from the perturbed sinograms, and (c) from the sinograms corrected with the proposed method.

The calibration performance as a function of the boundary extraction accuracy for both phantoms is shown in Fig. 8. While for a given extraction accuracy a range of calibration outcomes may be obtained, the mean performance deteriorates as accuracy decreases, with a slightly more pronounced decline for phantom B. The mean MAE with both phantoms increases by 50% and 100% for standard deviations of roughly 0.9 and 1.5 pixels, respectively, at which 53% and 28% of the boundaries are determined with pixel-accuracy.

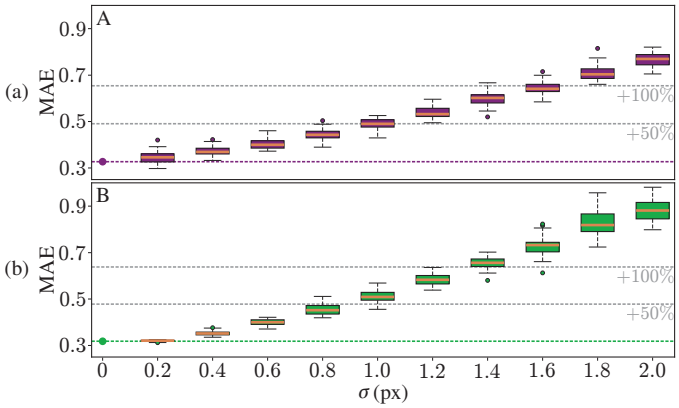


Fig. 8. MAE of shift estimations as a function of the standard deviation (σ) in boundary extraction, evaluated across 40 random offsets per (non-zero) σ-value: (a) phantom A, (b) phantom B.

V. Discussion

For both phantoms, shifts are estimated with near pixel-accuracy (Fig. 6), and correcting the sinograms accordingly yields substantially improved reconstruction qualities (Fig. 7), nearing that of the GT.

In the ideal theoretical case, phantom B was expected to yield perfect performance. In practice, finite translational and angular sampling limits the accuracy of the estimated radii of curvature, and the resulting discretization errors propagate into deviations from the ideal case. Since these radii provide the primary constraint on shift ambiguity, their imperfect estimation degrades calibration accuracy. The empirically observed non-convexity of the optimization, with different initializations converging to distinct minima, also plausibly explains the remaining deviations and indicates the need for a more robust optimization approach.

Conversely, due to its less constraining radii of curvature, phantom A was expected to achieve much lower accuracy. However, its performance is comparable. This may suggest that its set of admissible shifts is structured rather than uniform. The symmetry and smoothness of its supports may bias the optimization toward this solution, causing similar, though not identical, phantoms to converge to the same result. While this could explain the strong performance, it also suggests that smooth samples may yield less reliable calibration.

In practice, boundary extraction is affected by noise and partial volume effects, leading to imperfect estimates. Approximating these errors with a pixel-binned Gaussian of varying standard deviation (Fig. 8) indicates that small extraction

inaccuracies do not immediately lead to substantial performance degradation. However, performance may deteriorate when boundaries cannot be extracted with sufficient accuracy. For such cases, DCCs could be formulated to support the extraction processes, however. Finally, experimental datasets will typically contain substantially larger sets of sinograms than those considered here; this increased diversity may impose stronger cumulative constraints on the admissible shifts, thereby reducing ambiguity and potentially improving calibration accuracy.

VI. Conclusion

A novel tomographic calibration approach based on a pre-existing class of data consistency conditions is proposed and validated through numerical simulations. The method exploits sinogram support boundaries to recover shifts with respect to a chosen center of rotation. Provided these boundaries can be reliably extracted, the approach is invariant to attenuation effects, remains applicable to data acquired over half a rotation or less and does not require any reconstruction. Numerical simulations on shift-affected XFCT data demonstrate the effectiveness of the method. While applicable to a broader class of tomographic modalities, this work constitutes the first analytical calibration approach that is agnostic to attenuation effects in XFCT.

Acknowledgement

This work was supported by the European Synchrotron Radiation Facility (ESRF) and was performed within the framework of the SIRIC LYriCAN+ (INCa-DGOS-INSERM-ITMO cancer 18003) and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program 'Investissements d'Avenir' (ANR- 11-IDEX-0007) operated by the ANR. Simon Rit is a Royal Society Wolfson Visiting Fellow at the University College of London funded by the Wolfson Foundation and Royal Society.

References

[1] S. Helgason, *The Radon Transform*, ser. Progress in Mathematics. Boston, MA, USA: Birkhäuser, 1999.

[2] M. Guizar-Sicairos, A. Diaz, M. Holler, M. S. Lucas, A. Menzel, R. A. Wepf, and O. Bunk, "Phase tomography from x-ray coherent diffractive imaging projections," *Opt. Express*, vol. 19, no. 22, pp. 21 345–21 357, 2011.

[3] J. L. Prince and A. S. Willsky, "Estimation algorithms for reconstructing a convex set given noisy measurements of its support lines," Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, Cambridge, MA, Tech. Rep. LIDS-P-1638, 1987.

[4] Z. Zhang, Z. Dong, H. Yan, A. Pattammattel, X. Bi, Y. Dong, G. Liu, X. Sun, and Y. Zhang, "A general image misalignment correction method for tomography experiments," *iScience*, vol. 26, no. 10, p. 107932, 2023.

[5] H. Ryan, "On extreme constant width bodies in $\mathbb{R}^3$," *Discrete & Computational Geometry*, 2025.

[6] N. R. Viganò and V. A. Solé, "Physically corrected forward operators for induced emission tomography: a simulation study," *Measurement Science and Technology*, vol. 29, no. 3, p. 034005, 2018.

X-ray Dark-field Imaging of Iron Oxide Nanorods for Instantaneous DSA (iDSA)

Xiangyang Tang*, Yi Yang, Duhee Joen, Huiqiao Xie and Hui Mao

***Abstract*—X-ray dark-field (D-F) signal, which is a complex quantity, is making the grating interferometry-based X-ray imaging more capable of potential clinical applications. We have investigated the potential of X-ray D-F imaging with ellipsoidal microbubbles and rod-shaped iron oxide particles at both micro and nanoscales being the contrast agents for conspicuity enhancement in preclinical and/or clinical applications. In this work, we continue to study the feasibility of X-ray D-F imaging with iron oxide nanorods (IONRs) as the contrast agent, over IONRs' orientation, concentration and particle dimension. The preliminary data quantitatively show that, at nanoscales, the X-ray D-F signal generated by IONRs is much stronger than its counterpart in X-ray attenuation imaging. Moreover, it varies dramatically over IONRs' orientation, which, in principle, may serve as the foundation for advanced clinical applications, such as instantaneous digital subtraction angiography (iDSA).**

***Index Terms*—X-ray dark-field signal, X-ray dark-field imaging, grating interferometry-based X-ray imaging, iron oxide nanorods, nanoparticulated X-ray contrast agent.**

I. Introduction

The research and development in grating interferometry-based X-ray dark-field (D-F) imaging is gaining the momentum to establish its clinical utility for applications in pulmonary imaging [1], [2]. We investigated the signal corresponding to the D-F contrast generated by structures at microscale and found that the microstructures may not only generate strong D-F signal but also contribute to the phase contrast [3], revealing the root causes that account for the artifacts in X-ray differential phase contrast computed tomography while asymmetric microstructures are imaged [4]–[6]. We introduced the concept of complex X-ray D-F signal to characterize microstructures' effects on X-ray's D-F and phase signals [3]. Focusing on the real part of complex X-ray D-F signal, we have studied the potentials of X-ray D-F imaging with ellipsoidal microbubbles [7] and iron oxide microrods [8] as the agents for contrast enhancement in potential preclinical and clinical applications. Recently, we studied the X-ray D-F signal generated by iron oxide nanorods (INORs) and its variation over INORs' orientation [9]. Via three-dimensional (3D) phantom studies in this work, we continue to investigate the variation of X-ray D-F signal over INORs' concentration and dimension and explore the potential of X-ray D-F imaging with IORNs as the contrast agent for instantaneous digital subtraction angiography (iDSA).

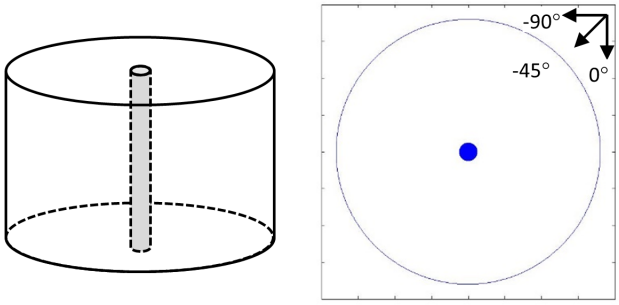


Fig. 1. Perspective (left) and transverse (right) views of the 3-D phantom (51.61 mm diameter, mimicking the neck) with a central target (3.71 mm diameter, mimicking the carotid) filled up with IONRs. The arrows indicate the angles at which the projections are generated and recorded.

II. Materials and Methods

The study's design of experiment is presented below.

A. Modeling and simulation of grating interferometry-based X-ray D-F imaging

As has been evaluated and verified in our previous work [3] and other publications in the literature [4]–[5], [10], the modeling and simulation of grating interferometry-based X-ray imaging is based on Fresnel-Kirchhoff analysis. In the simulation study, the period of gratings G_1 and G_2 are 4 and 2 μm, respectively. The 2-D detector consists of 12 rows and 228 columns, with the dimension of each element being set at 256×256 μm^2, and the X-ray source is assumed mono-chromatic at 35 and 80 keV, and the distance between the gratings G_1 and G_2 is 48.4 or 129.1 mm, corresponding to the respective first fractional Talbot distance. While X-ray's quantum property is considered, the X-ray flux is set at 10^8 photon/cm^2 and observes the Poisson distribution. To ensure high fidelity in the simulation study, the dimension of finite X-ray along the x-direction is 10 nm, which corresponds to 200 samplings within one period of grating G_2, while one of the gratings shifts 100 steps to acquire the data for each sampling.

Manuscript submitted Jan. 28, 2026, revised May 3, 2026, and published June 1, 2026.

X. Tang, D. Jeon and H. Mao are with the Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA 30322 USA (e-mail: xiangyang.tang@emory.edu).

Y. Yang was with the Department of Radiology and Imaging Sciences, Emory University School of Medicine, Atlanta, GA 30322 USA (e-mail: yangyi86@hotmail.com).

H. Xie is with the Department of Radiation Oncology, Emory University School of Medicine, GA 30322, USA (huiqiao.xie@emory.edu).

B. Phantom study via computational simulation

With a cylindrical target, which consists of IONRs being embedded at the center of a cylindrical neck/carotid phantom (Fig. 1), the computational simulation is designed to study the attenuation and D-F signals of the iron oxide nanoparticles and their orientational variation in X-ray D-F imaging. The neck/carotid phantom's diameter is 51.61 mm, while that of the central target is 3.71 mm. The refractive index of water is $1 - 1.8813\times10^{-7} + i8.8402\times10^{-11}$ at 35 keV and $1 - 3.598\times10^{-8} + i2.266\times10^{-11}$ at 80 keV, while its counterpart of iron oxide nanoparticles (Fe_3O_4) is $1 - 3.1264\times10^{-7} + i9.481\times10^{-10}$ and $1 - 5.9842\times10^{-8} + i6.7899\times10^{-11}$, respec-tively. At view angles -90°, -45° and 0°, the projections of X-ray attenuation and D-F signals are generated and plotted for quantitative analysis.

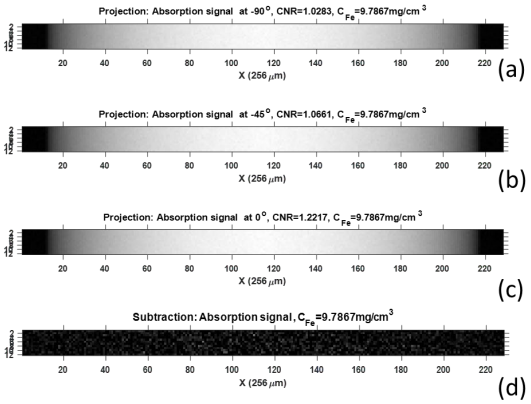


Fig. 2. X-ray's attenuation projection of the 3-D neck/carotid phantom at view angles -90° (a), -45° (b) and 0° (c) and the subtraction of projection at -90° from that at 0° (d).

With IORNs in the central target, three experiments are carried out in the 3-D phantom study: (I) X-ray D-F signal generated by INORs and its variation over nanorods' orientation; (II) X-ray D-F signal generated by INORs and its variation over nanorods' concentration; (III) X-ray D-F signal generated by INORs and its variation over nanos' dimension.

III. Results

The 3-D phantom study is carried out at 35 keV and 80 keV, respectively, but only the preliminary data corresponding to the former are presented here due to space limitation.

A. Experiment I: INORs' X-ray D-F signal and its variation over orientation

Presented in Fig. 2 are the 3-D neck/carotid phantom's projection corresponding to X-ray's attenuation signal over view angles -90° (a), -45° (b) and 0° (c), while their counter-parts in D-F signal are in Fig. 3. Visually, the central target (vessel) that is filled up with IONRs are not differentiable from the backgound (water) in the phantom's attenuation projection over all three view angles. However, in the phantom's D-F projection, the central IONR target manifests itself at view angle 0° (Fig. 3(c)), but submerged in noise at other view angles (-90°/-45°: Fig. 3(a)/(b)). Quantitatively, the contrast-to-noise ratio (CNR) between the central IONR target and background corresponding to Fig. 2 and Fig. 3 are plotted in Fig. 4 (top), in which, it is observed that the CNR varies little over IONRs' orientation in the conventional attenuation-based X-ray imaging, but varies substantially in X-ray D-F imaging.

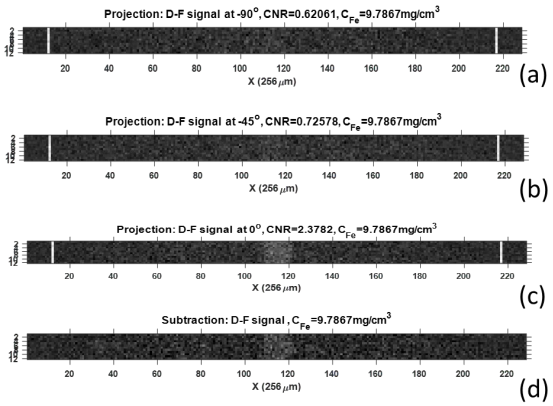


Fig. 3. X-ray's D-F projection of the 3-D neck/carotid phantom at view angles -90° (a), -45° (b) and 0° (c) and the subtraction of projection at -90° from that at 0° (d).

B. Experiment II: IONRs' X-ray D-F signal its variation over particle concentration

At 0° view angle, the CNR in the phantom's attenuation and D-F projections over IORNs' concentration are plotted in Fig 4 (bottom). It is observed that the CNR of the target against its background in both attenuation and D-F projections behave similarly, i.e., increases proportionally with concentration.

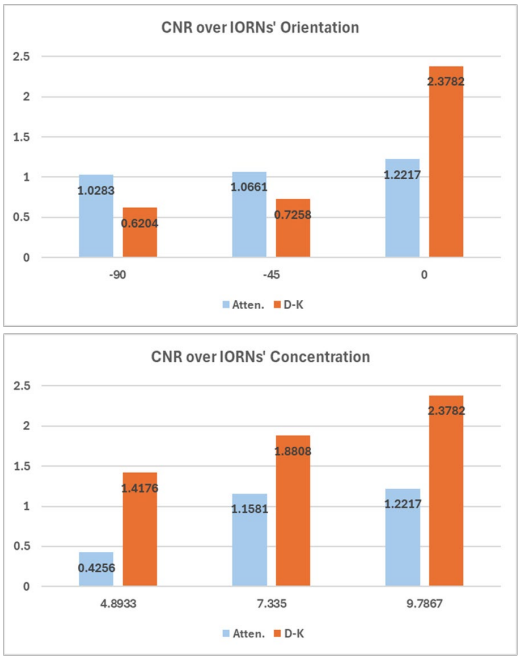


Fig. 4. The CNR measured in the neck/carotid phantom's attenuation and D-F projections over the view angles -90°, -45° and 0° (top) and the IONRs' concentration (bottom).

C. Experiment III: IONRs' X-ray D-F signal and its variation over particles' dimension

Given identical IORN concentration in the central target,

the CNR in phantom's attenuation projection at 0° view angle as a function over IORNs' dimension are plotted in Fig 5 (top), while that corresponding to the D-F projection are in Fig. 5 (bottom). Notably, at identical IORN concentration, the CNR of the central target decreases slightly with IONRs' dimension in the attenuation projection, whereas it decreases moderately with IONRs' dimension in the D-F projection.

D. Experiment IV: IONRs' X-ray D-F signal and its variation over particles' orientational distribution

Given identical IORN concentration in the central target, the CNR in phantom's attenuation and D-F projection at 0° view angle against the IORNs' orientational distribution are shown in Fig 6. Notably, at identical IORN concentration, the CNR in phantom's D-F projection decreases moderately with increasing SD in the particle's orientational distribution, while its counterpart in attenuation projection only fluctuates mildly.

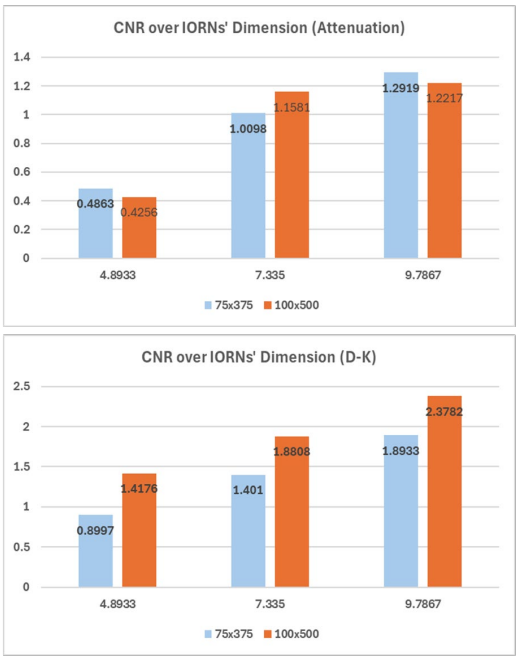


Fig. 5. Variation of the CNR between the 3-D phantom's central target and its background over IONRs' dimension: 75×375 um^2 and 100×500 um^2 (projections are generated at 0° view angle).

IV. Discussion

Though exogenous imaging contrast agents, e.g., the blood-compartment iodinated contrast agent used in conventional X-ray based imagings, is commonly employed in the clinic for conspicuity enhancement of pathophysiological abnormalilties and improve their detection and diagnosis, the R&D in contrast agents specific for X-ray D-F imaging will still be of interest. As shwon in Fig. 3(c), the IONRs in the central target generate strong contrast while they are lining up in orientation. Notably, the adoption of IONRs as the agent for contrast enhancement in X-ray D-F imaging is of significance and relevance, in terms of toxicity for potential preclinical and clinical applications, especially in light of the fact that the iron oxide nanoparticles have been approved by FDA as the contrast agents in clinical MRI applications.

It is not hard to imagine that, at a view angle (e.g., 0° in Fig. 1), two X-ray D-F images that are similar to the "contrast" and "mask" images in state-of-the-art digital subtraction angiography (DSA) can be acquired milliseconds, rather than seconds, apart; the "contrast" image is acquired while the anisotropic IRORs are lined up by applying an external magnetic field at 0° orientation; the "mask" image is taken while the external magnetic field is being tuned to another orientation, e.g., -45° or -90° (see Fig.1); then, the central target should appear at large contrast in the image formed by subtracting the "mask" one from the "contrast" one, whereas the surrounding tissues or organs should appear no or little contrast since they undergo no change in their orientations between the data acquisitions. This can serve as the foundation for implementing instantaneous DSA (iDSA) for diagnosis, therapeutical intervention and efficacy assessment of cardio-vascular, cerebrovascular and peripheral vascular diseases (see Fig. 3(d) and its comparison to Fig. 2(d)).

The iron oxide nanorods have been developed for biomedical applications such as MRI contrast agent [12]–[16], and hyperthermia therapies, due to their desirable properties in (i) chemical stabilities and reliable dispersion in liquid/fluid media, (ii) high saturation magnetization, (iii) large surface areas for binding or loading bioactive or functional moieties, and (iv) biocompatibility. The team led by one of the co-authors (H. Mao) has been engaged in synthesis and study of IONRs with shape and size-dependent physicochemical properties to tailor the nanoparticles for high performance in biomedical applications [17]–[20]. When varying its shape, the atomic arrangements in each facet of the nanocrystal can be altered, thus conferring new properties and applications [20]. Specifically, the anisotropic-shaped iron oxide nano-particles, e.g., nanorods with high aspect ratio, have been shown to be strong MRI contrast agents, given their increased transverse relaxitivity induced by the linear clustering effect. For *in vivo* applications, the anisotropic nanorods in a certain range of aspect ratios exhibit favorable pharmacokinetics and longer half-life in the circulation system compared to nanospheres [20]. We have worked on synthesizing iron oxide nanorods at various aspect ratios and studied the IONRs' magnetic properties at different aspect ratios (see the magnetic moment vs. applied magnetic field (M−H) curves presented in Fig. 7 [20] as examples).

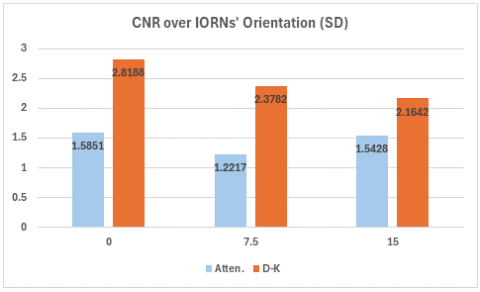


Fig. 6. Variation of the CNR between the 3-D phantom's central target and its background over IONRs' oriental distribution (projections are generated at 0° view angle).

Anisotropic iron oxide nanorods at high aspect ratio exhibit strong magnetic moments along their long axis and yield high efficiency in electromagnetic controlling of orientation and distribution, even under relatively weak magnetic field [15], [20]. If two images are acquired milli-seconds apart at the orientations corresponding to the nanorods' max/min D-F signal, the nanorods may appear in large contrast in the subtraction image against their surroundings that undergo no changes in orientation, as shown in Fig. 3(d) and compared to Fig. 2(d). As such, iDSA at substantially higher contrast sensitivity without motion artifacts becomes possible and thus overcomes the major challenges in the conventional DSA.

Finally, we want to indicate that only 12 rows were assumed in the X-ray detectors in the phantom studies so far, mainly due to computational limitations. Further studies that assume more rows in X-ray detectors are under the way and the data will be reported by us soon.

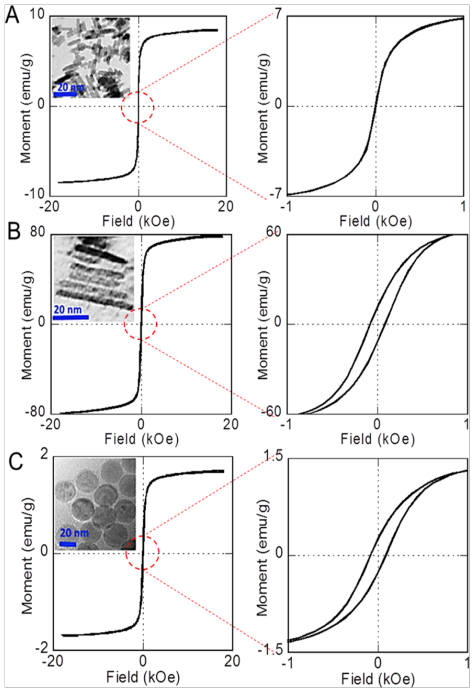


Fig. 7. Transmission Electron Microscopy (TEM) of 25 nm (a), 50 nm (b) IONRs, and nanospheres (c), and their magnetic-hysteresis (M-H) curves.

V. Conclusions

The preliminary data in this work show that the rod-shaped IONRs are anisotropic in morphology and may be utilized as the contrast agent for implementing instantaneous digital subtraction angiography (iDSA) that can overcome the challenges existing in the conventional DSA for diagnosis, therapeutic intervention, and therapeutic efficacy assessment in cerebrovascular and cardiovascular diseases.

VI. Acknowledgements

This work was supported in part by the National Institute of Health grant NS140900 (PI: X. Tang).

References

[1] M Viermetz, N Gustschin, C Schimid, J Haeusele, M. von Teuffenbach, P Meyer, F Bergner, T Lasser, T Koehler and F Pfeiffer, "Dark-field computed tomography reaches the human scale," *PNAS*, 119(8): e2118799119, Feb. 2022, 10.1073/pnas. 2118799119.

[2] E Willer, AA Fingerle, W Noichl, F De Marco, M Frank and T Urban et al, "X-ray dark-field chest imaging for detection and quantification of emphysema in patients with chronic obstructive pulmonary disease: a diagnostic accuracy study," Lancet Digital Health, 3:e733-44 2021.

[3] Y Yang and X Tang, "Complex dark-field contrast and its retrieval in X-ray phase contrast imaging implemented with Talbot interferometry," *Med. Phys*., 41(10):101914, 19 pages, Oct. 2014.

[4] YI Nesterets, "On the origins of decoherence and extinction contrast in phase-contrast imaging," *Opt. Commun.*, 281(4):533–542, Feb. 2008.

[5] SK Lynch, V Pai, J Auxier, AF Stein, EE Bennett, CK Kemble, X Xiao, W Lee, NY Morgan and HH Wen, "Interpretation of dark-field contrast and particle-size selectivity in grating interferometers," *Appl. Opt.*, 50(22):4310–4319, Aug. 2011.

[6] P Modregger, F Scattarella, BR Pinzer, C David, R Bellotti and M Stampanoni, "Imaging the ultrasmall-angle X-ray scattering distribution with grating interferometry," *Phys. Rev. Lett*., 108(4): 048101, 4 pages, Jan. 2012.

[7] X Tang and Y Yang, "Dark-field X-ray Imaging of Anisotropic Subdel Microstructures: Rod-shaped Air-bubbles," *Proce of 8th Internat. Conf. on Image Formation in X-ray CT*, pp. 443–446, 2024.

[8] X Tang, Y Yang and H Mao, "X-ray Dark-field Imaging of Anisotropic Rod-shaped Magnetic Iron Oxide Particles," *IEEE NSS MIC RTSD*, Tampa FL, Oct. 26–Nov. 2, 2024.

[9] Y Yang, H Xie, H Mao and X Tang, "The Signal Generated by Anisotropic Iron Oxide Nanorods in X-ray Dark-Field Imaging and Its Orientational Variation", *Proceeding of the 18th International Meeting on Fully Three-dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pp.590–593, 2025.

[10] M Bech, O Bunk, T Donath, R Feidenhans'I, C David and F Pfeiffer, "Quantitative X-ray dark-field computed tomography," *Phys. Med. Biol*., 55(18):5529–5539, Aug. 2010.

[11] W Yahiro, S Harasse, K Kawabata, H Kuwabara, T Yamazaki and A Momose 2011, "Distribution of unresolvable anisotropic microstructure in visibility-contrast images using X-ray Talbot interferometry," *Phys. Rev. B*, 84(9):094106, 9 pages.

[12] SJ Son, J Reichel, B He, M Schuchman, and SB Lee 2005, "Magnetic Nanotubes for Magnetic-Field-Assisted Bioseparation, Biointeraction, and Drug Delivery," *J. Am. Chem. Soc.,* 127:7316-7317.

[13] Wahajuddin and S Arora 2012, "Superparamagnetic iron oxide nanoparticles: magnetic nanoplatforms as drug carriers," *Int. J. Nanomed.* 2012:7 3445–3471.

[14] A Hervaultab and NTK Thanh 2014, "Magnetic nanoparticle-based therapeutic agents for thermo-chemotherapy treatment of cancer," *Nanoscale*, 2014, 6, 11553, DOI: 10.1039/c4nr03482a.

[15] C Pucci, A Degl'Innocenti, MB Gümüş and G Ciofani 2022, "Superparamagnetic iron oxide nanoparticles for magnetic hyperthermia: recent advancements, molecular effects, and future directions in the omics era," *Biomater. Sci.*, 2022, 10, 2103.

[16] K Li, H Nejadnik and HE Daldrup-Link 2017, "Next-generation superparamagnetic iron oxide nanoparticles for cancer theranostics," *Drug Discovery Today*, 22(9):1421–1429.

[17] J Huang, L Wang, R Lin, AY Wang, L Yang, M Kuang, W Qian and H Mao 2013, "Casein-coated Iron Oxide Nanoparticles for High MRI Contrast Enhancement and Efficient Cell Targeting," *ACS Appl Mater Interfaces*, 5(11): 4632–4639.

[18] L Wang, J Huang, H Chen, H Wu, Y Xu, Y Li, H Yi, YA Wang, L Yang and H Mao 2017, "Exerting Enhanced Permeability and Retention Effect Driven Delivery by Ultrafine Iron Oxide Nanoparticles with T1–T2 Switchable Magnetic Resonance Imaging Contrast," *ACS Nano.*, 11(5):4582–4592.

[19] R Lin, Y Li, T MacDonald, H Wu, J Provenzale, X Peng, J Huang, L Wang, AY Wang, J Yang and H Mao 2017, "Improving sensitivity and specificity of capturing and detecting targeted cancer cells with anti-biofouling polymer coated magnetic iron oxide nanoparticles," *Colloids Surf B Biointerfaces*, 150:261-270.

[20] A. Orza, H Wu, Y Xu, Q Lu and H Mao 2017, "One-Step Facile Synthesis of Highly Magnetic and Surface Functionalized Iron Oxide Nanorods for Biomarker-Targeted Applications," *ACS Appl. Mater. Interfaces*, 9:20719−20727.

Structured Loss Amplification for U-Net-based Human-scale Dark-field CT Streak Reduction

D. Frey*[1,2], T. Dorosti*[1,2,3], J. B. Thalhammer[1,2,3,4], J. F. Hilmer[1,2], P. Bleuel[1,2], T. Hiu[1,2], S. Peterhansl[1,2], J. McGinnis[5,6], T. Koehler[4,7], D. Pfeiffer[3,4], F. Pfeiffer[1,2,3,4], D. Rueckert[5,6,8], F. Schaff[1,2]

***Abstract*—Streak artifacts remain a substantial challenge for the clinical translation of human-scale dark-field computed tomography. Supervised learning approaches using convolutional neural networks have shown promise in artifact suppression, but typically omit domain-specific prior knowledge. In this work, we present a structured loss formulation that integrates spatial and frequency-informed bias into U-Net-based streak reduction by selectively amplifying pixel-wise loss, guiding the model towards high-fidelity predictions during training. Performance was analyzed across independently trained models and quantified via contrast-to-noise ratio (CNR) and full-width-at-half-maximum (FWHM) of the line spread function at the air-tissue interface. As a substitute for clinical patient data, test metrics were computed from a scan of ventilated ex vivo porcine lungs. Using center- and high-frequency amplification, CNR was enhanced by 24 % for L_1 and 33 % for L_2 loss while preserving adequate image resolution. These results motivate the use of structured loss amplification as a flexible design choice for neural network architectures and demonstrate its effectiveness for advanced streak reduction in dark-field imaging.**

***Index Terms*—Computed tomography, convolutional neural networks, dark-field imaging, deep learning, streak artifact reduction**

I. Introduction

X-RAY dark-field computed tomography (DFCT) measures small-angle scattering as complementary information to conventional attenuation-based imaging [1]. Recently, DFCT imaging has been translated to the human scale, but reconstruction quality remains limited by prominent streak artifacts intensified by the mechanically unstable environment of a rotating grating-based interferometer [2]. While these vibrations enable phase retrieval without an explicit stepping mechanism, fluctuations in the relative grating motion induce artifacts during sliding-window processing [3]. Previous approaches to streak reduction have focused on advancing conventional processing algorithms, for example by suppressing vibrational modes during sample reconstruction [4].

Convolutional neural networks (CNNs) have demonstrated their worth for a wide range of medical imaging tasks. Although predominantly applied to image segmentation, the structure-preserving nature of U-Net architectures makes them well-suited for regression-based image restoration [5]. In the context of attenuation-based CT reconstruction, U-Nets have been successfully applied to the mitigation of undersampling artifacts in low-dose sparse-view conditions [6], [7]. Metal artifact reduction faces a similar problem, for which some solutions extend the traditional supervised formula via self-supervised CNNs or prior-guided algorithms trained on the sinogram domain [8]–[10].

While previous work in these closely related fields provides an inspiring foundation for the development of advanced streak reduction algorithms, the complexity of DFCT artifacts impedes direct translation and dictates careful consideration of design choices, such as the loss function. Approaches such as feature-aware perceptual loss or spectral methods like EAGLE provide alternatives to the traditional loss computation between terminal network prediction and ground truth (GT) [11], [12]. Recently, a normalized loss formulation adopted as part of a dual-frame 3D U-Net has proven feasible for enhanced DFCT reconstructions by accounting for the lower intensity of lung tissue compared to bone [13].

Building on this idea, our work focuses on modulating loss functions tailored to DFCT streak reduction. Contrary to previous strategies, we modify loss on a pixel-wise level by selectively amplifying errors guided by prior knowledge in spatial and frequency domain. In doing so, a structured bias is imposed to purposefully constrain the solution space.

II. Methods

A. Dataset

The training data consisted of 89 reconstructions of CT scans acquired with a human-scale dark-field CT prototype scanner at the Munich Institute of Biomedical Imaging. Samples included a mix of different phantom materials as well as

*Equal contribution.

No live experiments were performed. Ethical approval was not required.

The authors declare no conflicts of interest.

[1]Chair of Biomedical Physics, Technical University of Munich (TUM), Germany
[2]Munich Institute of Biomedical Engineering, TUM, Germany
[3]Institute for Diagnostic and Interventional Radiology, TUM University Hospital, TUM, Germany
[4]TUM Institute for Advanced Study, TUM, Germany
[5]Chair for AI in Healthcare and Medicine, TUM, Germany
[6]Munich Center for Machine Learning (MCML), TUM, Germany
[7]Philips GmbH Innovative Technologies, Hamburg, Germany
[8]Department of Computing, Imperial College London, U.K.

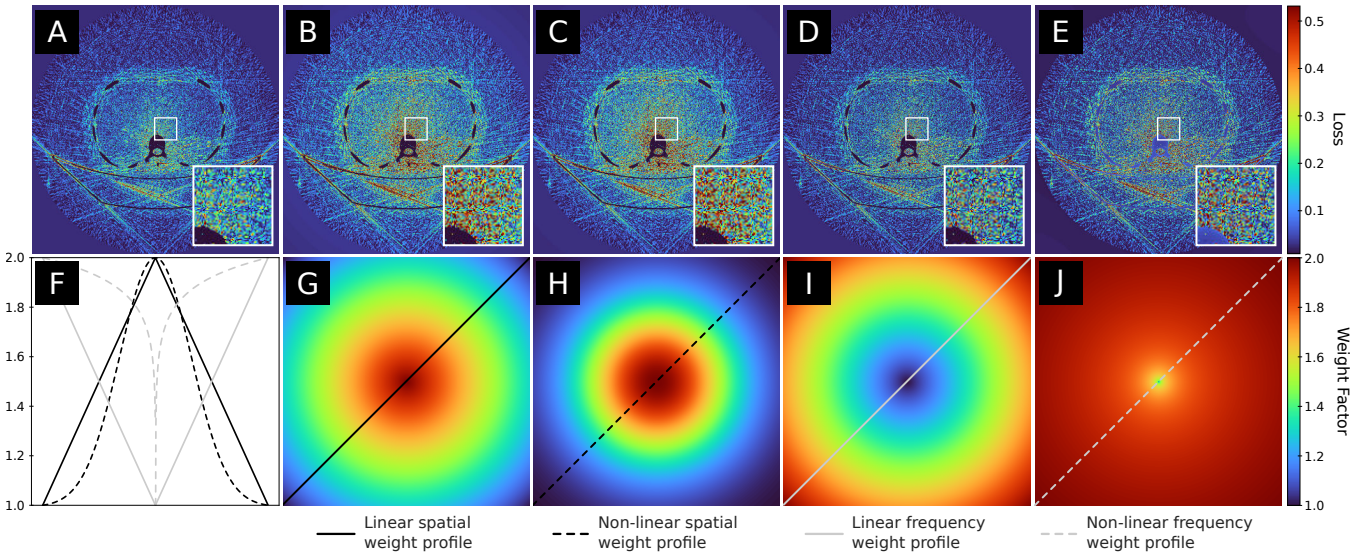


Fig. 1. Loss amplification strategies in spatial and frequency domain. For illustration, the pixel-wise loss map (A) is computed using an L_1 loss between a normalized single-rotation input slice and its multi-rotation GT. Spatially weighted loss maps (B, C) increase contributions towards the image center with linear and non-linear weights (G, H). Frequency-weighted loss maps (D, E) enhance high-frequency components by multiplication with linear and non-linear weights (I, J) in Fourier space. Designed weighting functions are highlighted from line profiles (F). Images are windowed consistently for each row.

ex vivo porcine lungs ventilated inside an anthropomorphic "Lungman" thorax phantom (Kyoto Kagaku, Tokyo, Japan). Experimental conditions were maintained at $80\,\mathrm{kVp}$, $550\,\mathrm{mA}$ and $1.5\,\mathrm{s}$ per full revolution.

Each scan comprised multiple (up to 20) rotations with 2400 shots each to create a stack of full-view projections. Input data at clinically feasible dose was gathered per sliding-window phase retrieval and filtered back-projection (FBP). Higher-quality GT data with significantly diminished artifacts and noise were obtained by virtual phase-stepping across all available rotations and FBP. From each volume, only the central axial slice (800×800 pixels) was used to limit training stochasticity.

B. Model and Training

Throughout all experiments, a fixed model architecture and consistent hyperparameters (learning rate $1\cdot10^{-3}$, weight decay $1\cdot10^{-1}$, Adam optimizer, linear learning rate scheduler with step size 3 and $\gamma = 0.9$, batch size 32) were used to isolate the effect of loss design. Hyperparameters were selected via grid-search using an L_1 loss. The validation split was set to $20\,\%$.

We employed a standard four-block 2D U-Net architecture with bilinear upsampling, batch normalization and ReLU as activation function in a residual-learning formulation. Each layer doubled the number of channels from 64 to 1024 at the deepest level. A global residual connection added the network prediction to the input image. During training, intensities of single- and multi-rotation data were clipped consistently to emphasize streaks and normalized to the range $[0, 1]$. For each input, a random patch (128×128 pixels) was picked to improve generalization; weights were cropped to the same patch during amplification.

Inference on the validation split, as well as on the hold-out test set containing scans of porcine lungs from a different specimen than those used for training, was performed on full-sized slices. Data augmentation comprised horizontal and vertical flips and in-plane rotations with a maximum angle of $\pm15°$ at $50\,\%$ chance. Training was terminated using early stopping with a patience of 20 epochs and a maximum of 100 epochs. Experiments were implemented in Python (v3.12.10) with PyTorch (v2.5.1, CUDA 12.4) and executed on an NVIDIA A100 GPU with $80\,\mathrm{GB}$ VRAM. While training and inference can be performed with substantially reduced VRAM, we chose to maximize batch sizes to reduce run-to-run variance given the limited dataset size.

C. Structured Loss Design

We based our design choices on previous knowledge collected from human-scale DFCT data. Pixel-wise loss maps $\mathcal{L} \in \mathbb{R}^{H\times W}$ for slices of height H and width W were multiplied element-wise with amplification weights $\mathcal{A}$. Weights were constrained to a range of $[1, 2]$ to mitigate the risk of vanishing or exploding gradients. Fig. 1 illustrates how this strategy visually enhanced structurally relevant components.

Focus towards the image center was enforced by increasing spatial weights $\mathcal{A}_S$ towards the iso-center to penalize errors more heavily in sample regions of interest (ROIs):

$$\mathcal{L}_S = \mathcal{A}_S \cdot \mathcal{L}. \tag{1}$$

This approach also emphasized regions where streaks typically appear stronger. The non-linear radial profile provided a smoother decay towards the image edges and a more distinct transition between ROI and background.

Frequency-space weights $\mathcal{A}_F$ were designed to amplify high-frequency contributions and multiplied with the Fourier-transformed loss map. The modulated loss follows as the real part of the inverse transformation after weighting:

$$\mathcal{L}_F = \mathrm{Re}\left\{\mathcal{F}_{2\mathrm{D}}^{-1}\left\{\mathcal{A}_F \cdot \mathcal{F}_{2\mathrm{D}}\left\{\mathcal{L}\right\}\right\}\right\}, \tag{2}$$

where $\mathcal{F}_{2\mathrm{D}}$ and $\mathcal{F}_{2\mathrm{D}}^{-1}$ denote the 2D Fourier transform and its inverse, and Re the real part. This approach was fine-tuned to the typical signature of DFCT streaks by further increasing the weight of mid- and high-frequency bands in the non-linear version. Concatenated weighting was implemented as consecutive application of (1) and (2) or their reversed order.

D. Performance Evaluation

For a systematic assessment of our loss amplification approaches, we trained all possible experimental configurations for ten repetitions at individually fixed random seeds. Configurations combined the following aspects:

- L_1 vs. L_2 as the baseline loss,
- linear vs. non-linear weighting functions, and
- spatial (S) vs. frequency (F) amplification, as well as their combined application.

Underperforming models were included to assess the robustness of each configuration.

After training, model performance was evaluated on test data and quantified for a single example slice via contrast-to-noise ratio (CNR) incorporating both signal (lung tissue) and background (air) noise:

$$\mathrm{CNR} = \frac{\mu_{\mathrm{lung}} - \mu_{\mathrm{air}}}{\sqrt{\sigma_{\mathrm{lung}}^2 + \sigma_{\mathrm{air}}^2}}, \tag{3}$$

where μ denotes the mean and σ the standard deviation of each ROI. Additionally, we analyzed the trade-off with image resolution by computing the full-width-at-half-maximum (FWHM) of the edge spread function derivative. ROIs (30×30 pixels) were picked for lung tissue as signal region and air as background reference close to the image center. The line profile for FWHM computation was drawn between the centers of both ROIs.

III. Results and Discussion

A. Quantitative Comparison of Model Performances

A comprehensive comparison of test performance metrics across differently amplified models is presented in Fig. 2. Median values with their standard deviations are listed in Tab. I for L_1 and Tab. II for L_2, maximum CNR values per weighting scheme are highlighted in bold font. Using L_2, loss amplification leads to modest improvements in CNR while

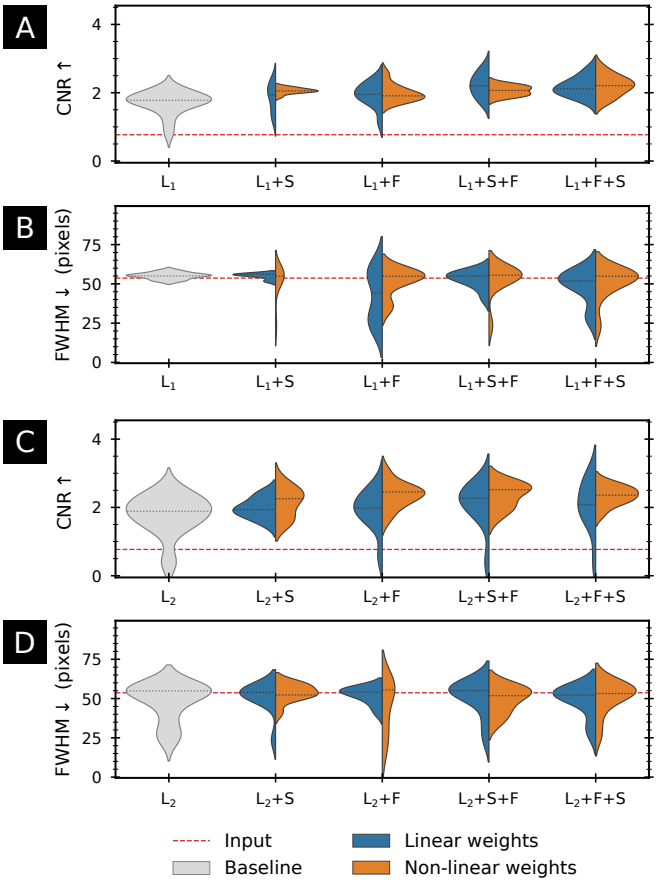


Fig. 2. Test performance for ten independently trained models in each experimental configuration. Violin plots illustrate the distributions of contrast-to-noise ratio (CNR) and full-width-at-half-maximum (FWHM) of the air-tissue edge for the original input image and predictions from models trained on amplified L_1 (A, B) and L_2 (C, D). Dashed lines indicate median values. Loss amplifications were realized with linear (blue) and non-linear weightings (orange). Abbreviations denote spatial weights (S), frequency weights (F), and their concatenations (S+F and F+S).

largely preserving FWHM, with the strongest gains observed for concatenated weighting schemes. Relative improvement in CNR reaches up to 24 % ($\mathrm{CNR} = 2.21 \pm 0.30$) compared to the baseline model ($\mathrm{CNR} = 1.78 \pm 0.34$). Non-linear weighting appears more robust against suboptimal outlier models, reflected by a slightly reduced variance across repetitions.

L_2-based models exhibit more pronounced performance gains from amplification. All configurations yield higher CNR values, achieving at maximum a relative increase of 33 % ($\mathrm{CNR} = 2.52 \pm 0.37$) from baseline ($\mathrm{CNR} = 1.89 \pm 0.58$), again observed for the concatenated configuration. Non-linear L_2 amplification consistently outperforms linearly weighted models. For the highest CNR gains, variance is reduced as well.

Loss amplification preserves spatial resolution at comparable levels to the baseline in exchange for substantial CNR

TABLE I
TEST METRICS FOR L_1 ($n = 10$, MEDIAN $\pm$ STANDARD DEVIATION)

	CNR ↑		**FWHM ↓ (pixels)**	
Experiment	Linear	Non-linear	Linear	Non-linear
L_1	1.78 ± 0.34		55 ± 3	
L_1+S	1.93 ± 0.41	2.05 ± 0.08	56 ± 2	55 ± 11
L_1+F	1.95 ± 0.36	1.91 ± 0.25	44 ± 16	55 ± 9
L_1+S+F	$\mathbf{2.20 \pm 0.35}$	2.07 ± 0.14	55 ± 6	56 ± 11
L_1+F+S	2.12 ± 0.28	$\mathbf{2.21 \pm 0.30}$	52 ± 11	55 ± 11

TABLE II
TEST METRICS FOR L_2 ($n = 10$, MEDIAN $\pm$ STANDARD DEVIATION)

	CNR ↑		**FWHM ↓ (pixels)**	
Experiment	Linear	Non-linear	Linear	Non-linear
L_2	1.89 ± 0.58		55 ± 12	
L_2+S	1.93 ± 0.30	2.26 ± 0.43	54 ± 10	52 ± 6
L_2+F	1.98 ± 0.58	2.45 ± 0.37	55 ± 6	56 ± 17
L_2+S+F	$\mathbf{2.27 \pm 0.66}$	$\mathbf{2.52 \pm 0.37}$	56 ± 12	52 ± 9
L_2+F+S	2.08 ± 0.74	2.36 ± 0.27	53 ± 10	53 ± 11

gains. The observed differences between configurations suggest that spatial and frequency-based structuring contribute complementary information during training.

B. *Qualitative Comparison of Model Performances*

Fig. 3 illustrates input, GT, and prediction slices from the test dataset for the assessment of streak reduction efficacy with non-linearly amplified L_2 loss. The single-rotation reconstruction exhibits pronounced streaks of varying intensity superimposed on the lung signal. These artifacts are largely suppressed in the multi-rotation reconstruction, as phase information is sampled more densely from the surplus of information. Although dose-wise clinically unfeasible, the quality of this reconstruction emphasizes the inherent potential of human-scale DFCT for pulmonary imaging.

Inference with the U-Net trained using baseline L_2 loss achieves only partial streak reduction. Residual artifacts appear as circular bands of alternating intensity. In contrast, all loss-amplified predictions are capable of mitigating these residual streaks globally.

Spatial loss amplification (L_2+S) enhances CNR while simultaneously improving resolution, as evidenced by sharper structural detail within the lung parenchyma. This observation suggests that emphasizing central regions during training is critical for effective streak mitigation. Frequency loss amplification (L_2+F) further increases CNR at the cost of FWHM, enhancing contrast at boundaries near large air cavities. Notably, this configuration treats intensity variations in homogeneous regions more conservatively than spatial amplification, which may help prevent the misinterpretation of residual artifacts as anatomical features.

Concatenated weighting (shown for L_2+S+F) prioritizes contrast enhancement over resolution, yielding a visually balanced trade-off between the two approaches. Lung tissue appears smoother at reasonable resolution. Although none of the predictions fully match the sharpness of the GT, CNR is clearly superior, as even multi-rotation reconstructions retain fine streak artifacts. In general, the observed gains in CNR can be attributed to reduced noise in both tissue and air regions at similar signal levels.

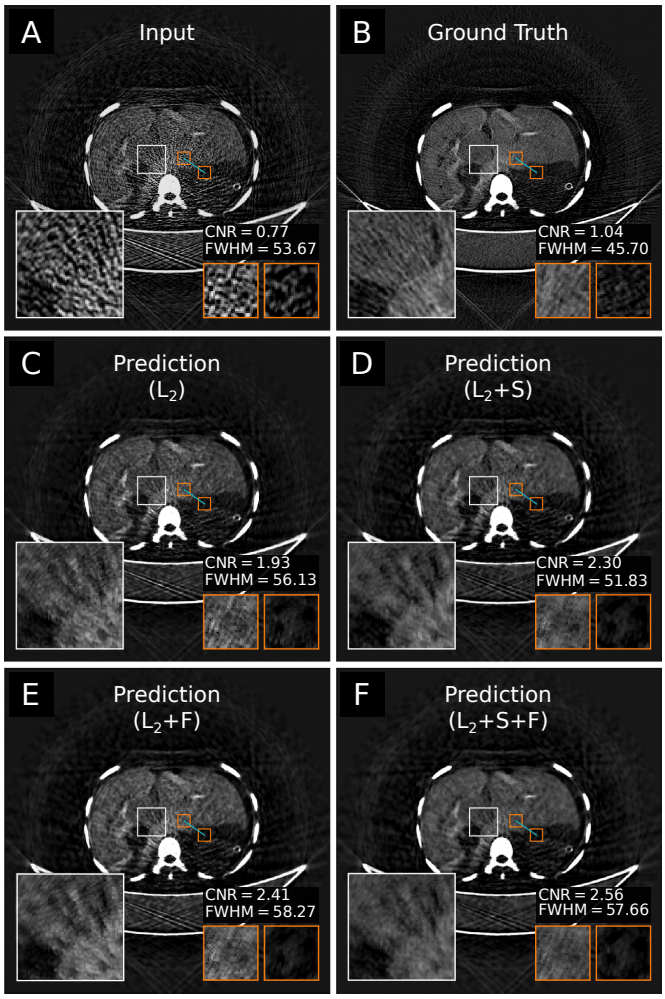


Fig. 3. Improved streak reduction of an axial DFCT slice from the test dataset with non-linear L_2 amplification. Inference tested on single-rotation input data (A) with a multi-rotation GT reconstruction for reference (B). Predictions retrieved using the baseline (C), spatially amplified (D), frequency-amplified (E), and concatenated spatially and frequency-amplified model (F). Out of all ten options per configuration, the prediction with CNR closest to each respective median is shown as representative slice. ROIs (orange) and line profile (cyan) for contrast-to-noise ratio (CNR) and full-width-at-half-maximum (FWHM) computation are indicated.

IV. CONCLUSION

This work demonstrates that incorporating domain knowledge directly into pixel-wise loss computation provides a powerful and interpretable mechanism for guiding deep learning based artifact suppression. Structured loss design offers a pathway toward narrowing the gap between domain knowledge and data-driven learning, ultimately supporting

the development of physics-informed models for clinical imaging. The direct comparison between linear and non-linear weighting strategies further indicates that fine-tuning loss maps can unlock additional performance gains, even when applied to comparatively simple U-Net architectures.

Notably, these improvements were achieved despite very limited training data, suggesting that further gains can be expected with continued data acquisition. Beyond the 2D setting, extending loss amplification to 3D U-Net architectures might further improve training based on global contextual CT information. While we originally aimed at removing DFCT streaks, our results indicate that CNNs also suppress other types of noise, an effect that should be examined more broadly in future work.

As DFCT streaks appear non-local and directional, more sophisticated weighting schemes incorporating masks adapted to streak orientations may be more effective than isotropic weighting functions, and should therefore be examined. Considering that even the multi-rotation GT exhibits fine streaks, a simulation study could provide further insights into how U-Net training is affected, while also offering a more concise basis for comparative evaluation. Additionally, unbiased reader studies will be essential for fine-tuning informed loss design choices.

Acknowledgment

Financial support through the European Research Council (ERC Smart Detectors for Darkfield X-ray Imaging, SyG 101167328), and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the States, as well as by the Technical University of Munich – Institute for Advanced Study.

References

[1] F. Pfeiffer, M. Bech, O. Bunk, P. Kraft, E. F. Eikenberry, C. Brönnimann, C. Grünzweig, and C. David, "Hard-X-ray dark-field imaging using a grating interferometer," *Nature Materials*, vol. 7, no. 2, pp. 134–137, 2008.

[2] M. Viermetz, N. Gustschin, C. Schmid, J. Haeusele, M. v. Teuffenbach, P. Meyer, F. Bergner, T. Lasser, R. Proksa, T. Koehler, and F. Pfeiffer, "Dark-field computed tomography reaches the human scale," *Proceedings of the National Academy of Sciences*, vol. 119, no. 8, 2022.

[3] C. Schmid, M. Viermetz, N. Gustschin, W. Noichl, J. Haeusele, T. Lasser, T. Koehler, and F. Pfeiffer, "Modeling vibrations of a tiled Talbot-Lau interferometer on a clinical CT," *IEEE Transactions on Medical Imaging*, vol. 42, no. 3, pp. 774–784, 2023.

[4] J. Haeusele, C. Schmid, M. Viermetz, N. Gustschin, T. Lasser, T. Koehler, and F. Pfeiffer, "Robust sample information retrieval in dark-field computed tomography with a vibrating Talbot-Lau interferometer," *IEEE Transactions on Medical Imaging*, 2024.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Springer International Publishing, 2015, pp. 234–241.

[6] J. Thalhammer, M. Schultheiß, T. Dorosti, T. Lasser, F. Pfeiffer, D. Pfeiffer, and F. Schaff, "Improving automated hemorrhage detection at sparse-view CT via U-net–based artifact reduction," *Radiology: Artificial Intelligence*, vol. 6, no. 4, p. e230275, 2024.

[7] A. Ries, T. Dorosti, J. Thalhammer, D. Sasse, A. Sauter, F. Meurer, A. Benne, T. Lasser, F. Pfeiffer, F. Schaff, and D. Pfeiffer, "Improving image quality of sparse-view lung tumor CT images with U-Net," *European Radiology Experimental*, vol. 8, no. 1, p. 54, 2024.

[8] L. Gjesteby, Q. Yang, Y. Xi, B. Claus, Y. Jin, B. De Man, and G. Wang, "Reducing metal streak artifacts in CT images via deep learning: Pilot results," in *The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 14, no. 6, 2017, pp. 611–614.

[9] B. Kim, H. Shim, and J. Baek, "A streak artifact reduction algorithm in sparse-view CT using a self-supervised neural representation," *Medical Physics*, vol. 49, no. 12, pp. 7497–7515, 2022.

[10] L. Yu, Z. Zhang, X. Li, and L. Xing, "Deep sinogram completion with image prior for metal artifact reduction in CT images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 228–238, 2021.

[11] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Springer International Publishing, 2016, pp. 694–711.

[12] Y. Sun, Y. Huang, Z. Yang, L.-S. Schneider, M. Thies, M. Gu, S. Mei, S. Bayer, F. G. Zöllner, and A. Maier, "EAGLE: An edge-aware gradient localization enhanced loss for CT image reconstruction," *Journal of Medical Imaging*, vol. 12, no. 1, pp. 014 001–014 001, 2025.

[13] T. Kumschier, J. Thalhammer, C. Schmid, J. Haeusele, T. Koehler, F. Pfeiffer, T. Lasser, and F. Schaff, "Streak artefact removal in X-ray dark-field computed tomography using a convolutional neural network," *Medical Physics*, vol. 51, no. 10, pp. 7404–7414, 2024.

Beam Hardening Correction for Dark-Field Imaging in a Human-Scale CT Prototype

Lennard Kayser, Daniel Frey, Josepha F. Hilmer, Sofia Demianova, Tina Dorosti, Maximilian E. Lochschmidt, Thomas Koehler, Daniela Pfeiffer, and Franz Pfeiffer

***Abstract*—Clinical dark-field CT provides a complementary dark-field image alongside conventional attenuation contrast. The dark-field signal is generated due to small-angle X-ray scattering (SAXS) at the microstructure of the sample, thereby providing insights into the functionality of, for example, lung tissue. However, hardening of the polychromatic X-ray energy spectrum produces not only the well-known beam hardening (BH) artifacts in the attenuation image but also an artificial dark-field signal that is not generated by SAXS in the sample. A look-up table-based correction was performed on the BH-induced dark-field signal, reducing the artificial contributions from bones and soft tissue to isolate the dark-field contrast of SAXS-generating structures of diagnostic interest. The calibration was performed using aluminum as the bone model and water as the soft tissue equivalent. Subsequently, a weighted combined look-up table was generated using regularized linear regression. The impact of varying material weightings on the correction was studied. The correction method was validated with Lungman phantom measurements. While demonstrated on scout scans, the proposed correction scheme is directly transferable to projection data and thus represents a necessary prerequisite for quantitative 3D dark-field CT.**

***Index Terms*—Beam hardening correction, computed tomography, dark-field imaging.**

I. Introduction

X-RAY dark-field computed tomography (DFCT) provides a complementary imaging signal to the conventional attenuation-based CT. The dark-field (DF) signal generates contrast due to small-angle X-ray scattering (SAXS), which creates spatial fluctuations of the wavefront due to microstructural material interfaces in the sample [1]. Therefore, DF imaging is particularly sensitive to areas with many material interfaces, such as the alveolar structure in the lungs [2]. This property of DF imaging has been shown to provide information on alveolar impairments [3]. As these are scattering effects below the resolution of conventional detectors, the DF signal is made accessible using a three-grating Talbot-Lau interferometer [4]. DF imaging has already demonstrated its diagnostic potential in a clinical radiography setup, outperforming conventional attenuation-based radiography [5]. Subsequent advancements have enabled the translation of DF imaging to a human-scale CT prototype [6]. While stationary interferometer setups employ a stepping movement of the gratings to retrieve the DF signal, this approach is not feasible in a rotating gantry setup. Therefore, the CT prototype uses the gantry vibrations to perform a sliding window phase retrieval [7]. While the attenuation signal is proportional to the mean intensity reduction when a sample is introduced, the DF signal is retrieved by analyzing the amplitude over the mean intensity of the interferometer pattern, called visibility. Because visibility also depends on the mean intensity, the DF signal strength is energy-dependent. Therefore, the reconstructed DF image is significantly impacted by artifacts, such as beam hardening (BH), which introduces an artificial DF signal that does not originate from the microstructure of the sample. The aforementioned effect makes it hard to distinguish true DF generating areas from areas with strong attenuation, generating an artificial signal.

This work applies a method to correct for this artifact and isolate the true DF signal originating from the microstructure of the sample. In contrast to previous beam-filtration [8] and simulation-based approaches [9] using single materials, this study employs a fast and simple weighted look-up table (LUT) based method to correct for material combinations of bone and soft tissue in the human thorax. As a preliminary experiment, the correction method was applied to CT scout scans. In the future, the method will be adapted to 3D CT scans. The methodology and the initial application on scout scans are outlined in this work.

II. Materials and Methods

A. Dark-Field CT Prototype

The CT system used is a modified Philips Brilliance iCT SP setup, as shown in Fig. 1. It is equipped with three additional gratings, G0, just after the source, dividing the incoherent X-ray source into many smaller, individually coherent line sources to make interferometry possible. The second grating upstream of the sample is called G1. It acts as a phase mask by imprinting a periodic phase modulation on the incoming wave. The phase modulation is converted to a high-frequency intensity modulation during propagation and de-modulated by

No live experiments were performed. Ethical approval was not required. The authors declare no conflicts of interest.

Financial support through the European Research Council (ERC Smart Detectors for Dark-Field X-ray Imaging, SyG 101167328), and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the States, as well as by the Technical University of Munich.

L. Kayser, D. Frey, J. Hilmer, S. Demianova, T. Dorosti, M. Lochschmidt, and F. Pfeiffer are with the Chair of Biomedical Physics, Department of Physics, School of Natural Sciences, Technical University of Munich, Garching, Germany and the Munich Institute for Biomedical Engineering, Technical University of Munich, Garching, Germany; (e-mail: lennard.kayser@tum.de).

L. Kayser, T. Dorosti, M. Lochschmidt, D. Pfeiffer, and F. Pfeiffer are with the Department of Diagnostic and Interventional Radiology, School of Medicine and Health, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany.

T. Koehler is with Philips Innovative Technologies, Hamburg, Germany.

T. Koehler and F. Pfeiffer are with the Institute for Advanced Study, TUM, Garching, Germany.

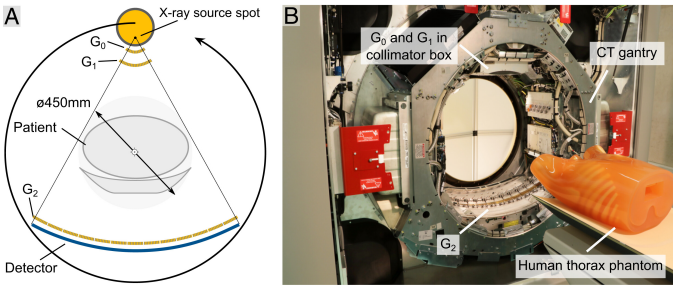


Fig. 1. A: Sketch of the CT setup. G0, G1, and G2 are the gratings of the Talbot-Lau interferometer to retrieve the dark-field signal. B: Picture of the Lungman human thorax phantom measurement in the CT. Figure adapted from [6].

the third grating, G2, just upstream of the detector [10]. This intensity pattern can be read out by the detector and used for the signal reconstruction. The reconstruction was done using sliding window processing [7]. For this work, only 2D scout scans were performed in which the source is fixed at the top, and the gantry does not rotate. The table is moved throughout the scan to cover a distance of $45\,\mathrm{cm}$ along the table. The measurements were taken with a tube voltage of $80\,\mathrm{kV}$ and a current of $200\,\mathrm{mA}$. An air scan was performed with the same measurement parameters but without the table and sample.

B. Image Formation

Dark-field contrast is formed via the combined effect of the three gratings which results in a periodic fringe pattern at the detector plane. The vibration of the setup itself slightly varies the relative positions of the three gratings resulting in an oscillation of the intensity signal in each pixel. The fundamental idea behind dark-field imaging is analyzing the local changes of this oscillation pattern induced by a sample [4]. The normalized amplitude of this intensity oscillation is called visibility

$$V(\bar{E}) = \frac{I_{max}(\bar{E}) - I_{min}(\bar{E})}{I_{max}(\bar{E}) + I_{min}(\bar{E})}. \tag{1}$$

Once the parameters, mean intensity $I_0(\bar{E})$ for the reference scan and $I(\bar{E})$ for the sample scan, as well as the visibility for both the sample scan V and the reference scan V_0 are retrieved, the attenuation and DF signals can be calculated. The attenuation signal is given by

$$A(\bar{E}) = -\ln\left(\frac{I(\bar{E})}{I_0(\bar{E})}\right) \tag{2}$$

and the DF signal is calculated using

$$D(\bar{E}) = -\ln\left(\frac{V(\bar{E})}{V_0(\bar{E})}\right) \tag{3}$$

and (1) [1]. The parameters visibility, attenuation, and DF signal are all dependent on the mean energy of the X-ray beam $\bar{E}$.

C. Correction Concept for Beam Hardening Induced Dark-Field Signal

Based on the attenuation in the human thorax, tissues can be grouped into the main classes bone and soft tissue, both of which produce negligible SAXS. However, both materials harden the polychromatic source spectrum, $I_0(\bar{E})$, thereby altering the measured visibility (1) and introducing an artificial DF contribution (3) superimposed on the true SAXS signal. Since this artificial contribution does not deliver microstructural information, it is considered as an artifact that should be removed from the measured DF image. In contrast, lung tissue produces a true SAXS signal originating from the alveolar microstructure, which represents the component of interest. Using (3), the measured DF signal D can be decomposed into a SAXS term D' and a BH-induced term D_{BH} [11]:

$$\begin{aligned} D &= -\ln\left(\frac{V}{V_0^{\mathrm{BH}}} \cdot \frac{V_0^{\mathrm{BH}}}{V_0}\right) \\ &= -\ln\left(\frac{V}{V_0^{\mathrm{BH}}}\right) - \ln\left(\frac{V_0^{\mathrm{BH}}}{V_0}\right) \\ &= D' + D^{\mathrm{BH}}. \end{aligned} \tag{4}$$

V_0^{BH} denotes the reference visibility measured with the same beam-hardened spectrum as in the sample measurement, i.e., the visibility measured in a calibration measurement matching the sample attenuation.
To estimate D_{BH}, calibration measurements were performed and combined into a single LUT. As beam hardening increases monotonically with attenuation, the measured attenuation, A, served as a surrogate for beam hardness and as the LUT input. For homogeneous attenuators without microstructure, (4) reduces to

$$D = D^{\mathrm{BH}}, \tag{5}$$

which enables direct estimation of the BH-induced DF signal as a function of attenuation [11]. Calibration measurements with materials lacking microstructure in multiple thicknesses were performed with the samples placed on the patient table. Following standard CT practice, water and aluminum were used as surrogates for soft tissue and bone, respectively. For each configuration, attenuation images, A, and DF images, D, were acquired using the standard protocol [12], and the mean signal within a region of interest (ROI) was used to generate the LUTs. Two LUTs were created: One for pure water, $D_{\mathrm{H_2O}}(A)$, and one for pure aluminum, $D_{\mathrm{Al}}(A)$, with spline fits. The spline fit was performed using a regularized linear regression [13]:

$$\sum_{i=1}^{n} |D_i - f(A_i)|^2 + \lambda \int_{A_1}^{A_n} \left(f^{(2)}(u)\right)^2 \mathrm{d}u, \tag{6}$$

with λ determined by generalized cross validation [14]. A_i and D_i are the attenuation and DF values of the n calibration measurements, whereas $f(A_i)$ denotes the fitted spline function at the corresponding attenuation input value. These two LUTs were then combined to $D_{w_{\mathrm{Al}}}$ by a weighted mean in

$$D_{w_{\mathrm{Al}}}(A) = (1 - w_{\mathrm{Al}}) \cdot D_{\mathrm{H_2O}}(A) + w_{\mathrm{Al}} \cdot D_{\mathrm{Al}}(A), \tag{7}$$

TABLE I
ALUMINUM AND WATER THICKNESSES [MM] USED FOR THE CALIBRATION MEASUREMENTS.

Water	0	50	100	125	150	175	200	230	
Alu	2	4	6	8	10	12	14	16	18

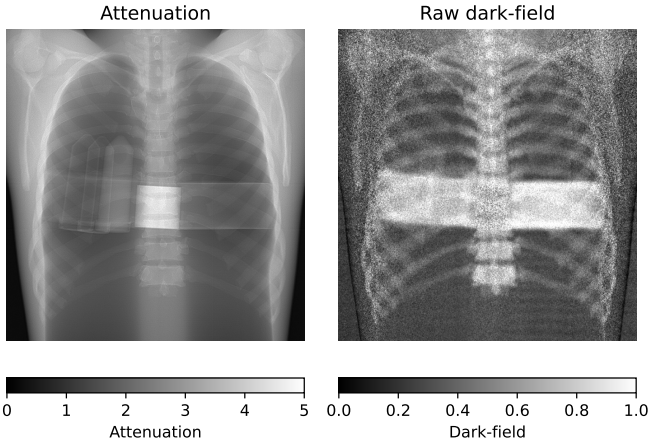


Fig. 2. Attenuation (left) and uncorrected raw dark-field (right) images of the Lungman phantom measurement. The images were retrieved using (2) and (3), respectively.

with w_{Al} being the weight for the aluminum contribution in the correction. The attenuation image, A, of a sample scan was then used as the input for $D_{w_{\mathrm{Al}}}(A)$ to estimate the amount of BH-induced DF signal, which was subtracted from the raw DF image.

D. Phantoms for Calibration and Validation

Multiple thicknesses of aluminum and water were used for the calibration indicated in Table I. The value 0 represents a measurement taken only from the table, without a sample. The air scan, required for image reconstruction, was performed without the patient table, as the material composition of the table varies along its full length. We used the Lungman phantom (Multipurpose Chest Phantom N1 "Lungman", Kyoto Kagaku) as a validation for the implemented correction on human-like samples. An image of the phantom is visible in Fig. 1B. The phantom models the thorax of an average Japanese man (65 kg, 168 cm) with soft tissue replicas made of urethane-based resin and bone replicas made of epoxy resin [15]. In addition, one neoprene foam inlet was placed in the Lungman thorax to mimic lung tissue, together with water and wool falcon tubes and a Polyoxymethylene (POM) plate in the foam inlet center to test the correction performance with multiple attenuators in the beam.

III. Results and Discussion

A. Attenuation and Raw Dark-Field Images

Fig. 2 shows the two images obtained from a scout scan of the Lungman phantom. The strongly attenuating materials, i.e., the bone replicas and the POM plate, are clearly visible in the horizontal center band of the attenuation image of the phantom. The neoprene foam is hardly visible in the attenuation image, as expected. However, the foam inlet is clearly visible in the raw DF image, as it scatters due to its foamy structure with many material interfaces. The bones are also clearly visible. Differentiating between materials with SAXS-induced "real" DF signal (e.g., neoprene/lung tissue) and strongly beam-hardening materials (e.g., bones/aluminum) that create a BH-induced DF signal is very challenging in this case. This is especially relevant in regions where the DF generated by the neoprene foam and the BH-induced signal by the spine and POM plate overlap such as in the center region of figure 2.

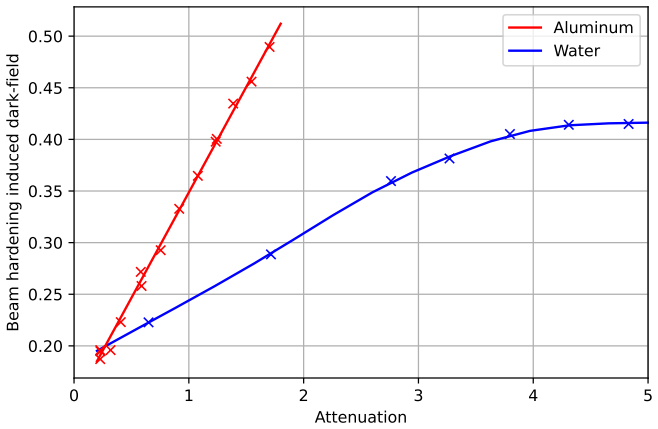


Fig. 3. Beam hardening induced dark-field signal of the calibration measurements for aluminum (red) and water (blue). The aluminum/water data were fit using regularized linear regression, as indicated by the solid red/blue lines. These two materials produce negligible SAXS. Therefore the measured dark-field signal can be attributed to beam hardening.

B. Retrieval of the Look-Up Tables

The attenuation and DF values for each thickness of aluminum and water from the calibration measurements are shown in Fig. 3. These homogeneous materials were chosen as bone (aluminum) and soft tissue (water) equivalents, as they ideally only produce attenuation and no SAXS-induced DF signal. Therefore, the measured DF signal of these materials can be attributed to the BH-induced DF signal. The data points were fit using smoothing spline functions for each material to create two LUTs: $D_{\mathrm{H_2O}}(A)$ (blue curve) and $D_{\mathrm{Al}}(A)$ (red curve). The two curves start with an offset in attenuation and DF because the table itself creates some attenuation and BH-induced DF signal that is also present in the regular sample images.

C. Comparison of Material Weightings

The two LUTs for aluminum and water were combined into a single weighted LUT with the weight w_{Al} giving the aluminum contribution used for the correction. To determine the optimal weight, different aluminum contributions were evaluated, balancing bone correction strength with minimal over-correction. Fig. 4 shows the corrected DF images of the Lungman phantom for three relative amounts of aluminum:

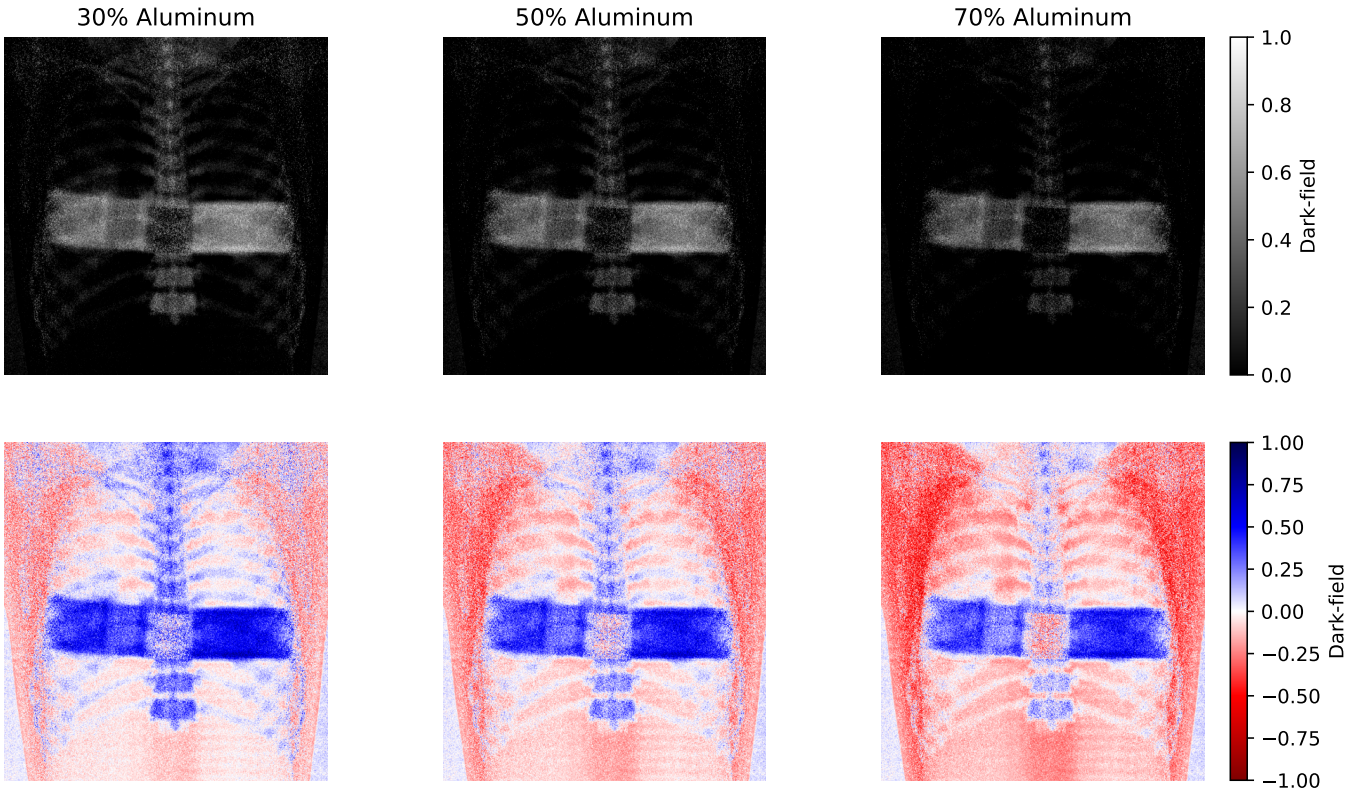


Fig. 4. The three columns show corrected dark-field images with three different contributions of aluminum (30 %, 50 %, and 70 %) in the weighted combined LUTs. For each material contribution the upper row shows the dark-field images windowed between 0 and 1 as you would normally observe them, whereas the lower row shows the images scaled from -1 to 1 highlighting regions of over-correction. Red corresponds to regions with over-correction, and blue indicates regions with positive dark-field signal.

30 %, 50 %, and 70 % as columns. The top row indicates the corrected DF images, whereas the bottom row subplots show regions of over-correction in red and regions with positive DF signal in blue. In an ideal scenario, there will be no over-corrections (red areas) visible, and only true positive DF signal from the foam (blue areas) would be measured, with no DF signal contributions from the bones and soft tissue. The 30 % column still has a noticeable signal from the bones, while the 70 % column has a strong over-correction signal in the soft tissue region. Additionally, the outer parts of the foam inlet are also notched by the correction. The 50 % column balances over-correction of soft tissue and signal from the bones. Hence, we use a 50 % aluminum contribution to start with.

Overall, the bone signal was reduced, allowing the SAXS DF signal from the foam inlet to be isolated and recognized as the area of interest. Some bone signal is remaining but overall it is very much reduced in the 50 % aluminum contribution in comparison to the raw DF image in Fig. 2. The ideal contribution of aluminum should be between 30 % and 70 % and will be determined in detail in the future using a suitable figure of merit.

IV. Conclusion

Beam hardening induced DF signal creates an artificial signal that is not due to SAXS. We developed a regularized linear regression method using aluminum and water calibrations to correct beam-hardening artifacts in dark-field imaging. Lungman phantom tests confirmed the method reduces unwanted bone and tissue contributions while preserving the SAXS signal. This proof-of-concept on CT scout scans paves the way for 3D CT applications.

References

[1] W. Yashiro, Y. Terui, K. Kawabata, and A. Momose, "On the origin of visibility contrast in X-ray Talbot interferometry," *Optics Express*, vol. 18, no. 16, p. 16890, 2010.

[2] K. Willer, A. A. Fingerle, W. Noichl, F. De Marco, M. Frank, T. Urban, R. Schick, A. Gustschin, B. Gleich, J. Herzen *et al.*, "X-ray dark-field chest imaging for detection and quantification of emphysema in patients with chronic obstructive pulmonary disease: a diagnostic accuracy study," *The Lancet Digital Health*, vol. 3, no. 11, pp. e733–e744, 2021.

[3] A. Yaroshenko, K. Hellbach, A. Ö. Yildirim, T. M. Conlon, I. E. Fernandez, M. Bech, A. Velroyen, F. G. Meinel, S. Auweter, M. Reiser *et al.*, "Improved in vivo assessment of pulmonary fibrosis in mice using X-ray dark-field radiography," *Scientific Reports*, vol. 5, no. 1, pp. 1–8, 2015.

[4] F. Pfeiffer, M. Bech, O. Bunk, and et al., "Hard-x-ray dark-field imaging using a grating interferometer," *Nature Materials*, vol. 7, pp. 134–137, 2008.

[5] T. Urban, A. P. Sauter, M. Frank, K. Willer, W. Noichl, H. Bast, R. Schick, J. Herzen, T. Koehler, F. T. Gassert, J. H. Bodden, A. A. Fingerle, B. Gleich, B. Renger, M. R. Makowski, F. Pfeiffer, and D. Pfeiffer, "Dark-field chest radiography outperforms conventional chest radiography for the diagnosis and staging of pulmonary emphysema," *Investigative Radiology*, vol. 58, no. 11, pp. 775–781, Nov. 2023.

[6] M. Viermetz, N. Gustschin, C. Schmid, J. Haeusele, M. von Teuffenbach, P. Meyer, F. Bergner, T. Lasser, R. Proksa, T. Koehler, and F. Pfeiffer, "Dark-field computed tomography reaches the human scale," *Proceedings of the National Academy of Sciences*, vol. 119, no. 8, p. e2118799119, 2022.

[7] J. Haeusele, C. Schmid, M. Viermetz, N. Gustschin, T. Lasser, T. Koehler, and F. Pfeiffer, "Advanced phase-retrieval for stepping-free x-ray dark-field computed tomography," *IEEE Transactions on Medical Imaging*, vol. 42, no. 10, pp. 2876–2885, Oct. 2023.

[8] N. Bevins, K. Li, J. Zambelli, and G.-H. Chen, "Type II beam hardening artifacts in phase contrast imaging," in *Medical Imaging 2013: Physics of Medical Imaging*, vol. 8668. SPIE, 2013, pp. 308–313.

[9] G. Pelzer, G. Anton, F. Horn, J. Rieger, A. Ritter, J. Wandner, T. Weber, and T. Michel, "A beam hardening and dispersion correction for X-ray dark-field radiography," *Medical Physics*, vol. 43, no. 6Part1, pp. 2774–2779, 2016.

[10] F. Pfeiffer, T. Weitkamp, O. Bunk, and C. David, "Phase retrieval and differential phase-contrast imaging with low-brilliance X-ray sources," *Nature Physics*, vol. 2, no. 4, pp. 258–261, 2006.

[11] M. E. Lochschmidt, T. Urban, L. Kaster, R. Schick, T. Koehler, D. Pfeiffer, and F. Pfeiffer, "Look-up table-correction for beam hardening-induced signal of clinical dark-field chest radiographs," 2025. [Online]. Available: https://arxiv.org/abs/2508.05422

[12] J. Haeusele, C. Schmid, M. Viermetz, N. Gustschin, T. Lasser, T. Koehler, and F. Pfeiffer, "Advancing x-ray dark-field imaging for human-scale applications: helical CT and surview imaging," in *Medical Imaging 2024: Physics of Medical Imaging*, R. Fahrig, J. M. Sabol, and K. Li, Eds., vol. 12925, International Society for Optics and Photonics. SPIE, 2024, p. 129250G. [Online]. Available: https://doi.org/10.1117/12.3006374

[13] H. J. Woltring, "A fortran package for generalized, cross-validatory spline smoothing and differentiation," *Advances in Engineering Software*, vol. 8, no. 2, pp. 104–113, 1986.

[14] G. Wahba, "Estimating the smoothing parameter," in *Spline Models for Observational Data*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics, 1990, pp. 45–65.

[15] Kyoto Kagaku Co., Ltd., "Ph-1 multipurpose chest phantom n1 ii "lungman" – product data."

Simulation Study of Scatter Rejection in Time-of-Flight CT

Altea Lorenzon, Xiao Jiang, J. Webster Stayman

***Abstract*—Scattered radiation poses significant challenges in X-ray computed tomography (CT). Unmitigated scatter introduces a number of artifacts, degrades image quality, and results in quantitative inaccuracy in the reconstruction. Recent advances in high temporal resolution photon counting detectors and ultra-fast X-ray sources have enabled a novel modality to differentiate between primary and scattered radiation based on photon time of flight (TOF). In essence, while primary (unscattered) photons travel a straight path, photons subject to Compton and Rayleigh interactions change direction and undergo longer overall path lengths, hence longer TOF before arriving at the detector. Thus, if TOF can be estimated with sufficient accuracy, individual detection events can be characterized as scatter and rejected. In this work, we develop a TOF CT simulator that includes a GPU-based Monte Carlo simulator. We implement a timing threshold based scatter rejection method, and evaluate the performance of TOF CT under different scenarios of X-ray pulse duration and detector timing resolution.**

I. Introduction

X-RAY computed tomography (CT) is a cornerstone of modern medical imaging, with a wide range of diagnostic and interventional applications. However, like all imaging modalities involving X-rays, one must contend with scattered radiation. Scattered photons measured by the detector carry little to no information about the object and generally present as an unknown additive contribution to the projection data. If not rejected or corrected in CT, scatter can lead to significant image quality degradations including cupping and streak artifacts, loss of contrast, decreased contrast-to-noise ratio and inaccuracy of the reconstructed attenuation values [1]. In cone-beam CT (CBCT), which uses large cone angles and tends to capture more scattered photons, scatter fractions can be particularly high, making effective scatter control even more important.

Several solutions have been proposed to reduce the impact of scatter in X-ray imaging [2], [3]. This includes *scatter rejection*, where one reduces the number of scattered photons that are detected; and *scatter correction*, where the scatter contribution is estimated and subtracted from the projection data. Scatter rejection methods include hardware devices placed on the source side, such as beam collimators [4] and bowtie filters [5]; and on the detector side, such as anti-scatter grids [6]. Increasing the air gap can also help to reduce scattered photons directed to the detector, but it is not always a practical solution. Among scatter correction methods, specialized hardware can be utilized to characterize the scatter component in the measured data, such as beam blockers [7] and thin collimation on a wide detector. Many software-based methods have been developed, including Monte Carlo simulations [8], scatter kernel superposition methods [9], and more recently, machine learning-based approaches [10]. Despite their effectiveness in mitigating scatter artifacts, many of these methods are limited by the requirement of additional hardware, computational complexity, or modeling assumptions. Moreover, while scatter correction methods can generally account for the mean scatter contribution, the quantum noise associated with those photons will remain present in the projection data.

Altea Lorenzon, Xiao Jiang, and J. Webster Stayman are in Biomedical Engineering at Johns Hopkins University. Contact author e-mail: web.stayman@jhu.edu

Recent advances using ultra fast X-ray sources and high temporal resolution detectors have enabled a novel, physically motivated technique to discriminate between primary and scattered radiation based on photon time-of-flight (TOF) [11]. Specifically, while primary photons travel a straight path from the source to the detector, photons undergoing Compton and Rayleigh scattering interactions are subject to direction changes that increase their path length. Therefore, assuming that all photons are emitted at the same time, the arrival time of scattered photons is delayed with respect to the primary photons. Whether it is possible to leverage this time delay to reject scattered photons depends on the characteristics of the X-ray source (e.g. pulse width) and the capabilities of the system to accurately measure the TOF.

In this work, we developed a Monte Carlo simulator for fast, GPU-based computation of photon interactions in a user-specified object. The simulator records detection events including TOF information and can model varied X-ray pulse shapes/widths and detector temporal response. We also implemented a scatter rejection method based on simple thresholding to discard late-arriving photons and investigated performance in a simulated CT system. Performance was evaluated in three scenarios corresponding to different source pulse durations and detector timing resolutions.

II. Methods

A. Forward Model

In this work, we adopt the following polyenergetic forward model for CT measurements, where the expected value of the i^{th} measurement is given by:

$$\bar{y}_i = \sum_{k=1}^{N_E} B_i(E_k)\Delta E \exp\left[-\sum_{m=1}^{N_m}\mu^m(E_k)\sum_{j=1}^{N}A_{ij}\rho_j^m\right] + s_i. \tag{1}$$

In this model, we assume that the object is made of N_m material bases, each characterized by a density ρ^m and an energy-dependent mass attenuation coefficient $\mu^m(E_k)$. The function $B_i(E_k)$ accounts for both the poly-energetic source spectrum and the detector energy response - which is discretized into N_E energy bins of width ΔE. The system matrix that performs linear projection has elements denoted by A_{ij} describing the contribution of the j^{th} voxel to the i^{th} measurement. The additive scatter contribution is represented by the term s_i, and, in this work, will be modeled using our Monte Carlo simulator.

B. Monte Carlo Simulation of Scatter

To estimate the amount of scatter produced by a specific CT system and object, we implemented a Monte Carlo simulation that models the transport of X-ray photons in the system in the energy range $10-150$ keV. The simulation setup includes a point polychromatic source and a photon-counting detector that can either be flat (e.g. flat-panel CBCT) or curved (e.g. traditional multi-row CT). The pulse shape $PS(t)$ of the X-ray source can

be described by arbitrary functions - e.g., a Dirac delta (instantaneous emission), a uniform distribution with variable width, or a Gaussian distribution with variable FWHM. Similarly, the detector timing resolution or jitter $TR(t)$ can be also be specified by selection of the distribution - e.g., Dirac delta (infinite resolution), a uniform distribution or a Gaussian distribution.

Object being scanned is defined as a voxelized volume, where each voxel is assigned a material, elemental or compound, which is defined by its chemical composition and mass density. For each material, the photon interactions are governed by the mass attenuation coefficients $\mu_{tot}(E)$ for the dominant effects contributing to the X-ray attenuation in this energy range, namely the photoelectric effect (*Pe*), Compton (*Co*) and Rayleigh (*Ra*) scattering. The total mass attenuation coefficient is given by:

$$\mu_{tot}(E) = \mu_{Pe}(E) + \mu_{Co}(E) + \mu_{Ra}(E). \quad (2)$$

For compound materials, the mass attenuation coefficient is calculated as

$$\mu_{compound}(E) = \sum_i w_i, \mu_i(E), \text{ with } w_i = \frac{n_i M_i}{\sum_j n_j M_j}, \quad (3)$$

where n_i and M_i denote the number of atoms and the atomic mass of the i^{th} element in the compound, respectively.

The simulation process, which is repeated for every projection angle, consists of three main steps:

1) **Generation** Photons are initialized with direction vectors pointing to individual detector pixels and projected to the volume surface. Each photon is assigned an energy, such that, for each energy bin, the number of generated photons is proportional to the product of the pixel fluence and the spectral intensity.
2) **Interaction** Within the volume, photons are transported according to the delta tracking method [12]. To determine the type of interaction, a number ζ is drawn from a uniform distribution between 0 and 1: if $\zeta < \mu_{Pe}/\mu_{tot}$ a photoelectric interaction occurs and the photon is not propagated further; if $\mu_{Pe}/\mu_{tot} \leq \zeta < (\mu_{Pe} + \mu_{Co})/\mu_{tot}$, a Compton interaction takes place, resulting in an updated photon energy and direction according to [13]. Otherwise, a Rayleigh scattering occurs, and the photon direction is updated following [14]. This process is repeated until the photon energy falls below the threshold $E_{min} = 10\,\text{keV}$ or the photon exits the volume.
3) **Detection** Once photons exit the volume, they are projected onto the detector's surface based on their final direction. If they hit the detector's surface, the scatter map - discretized in (N_T, N_v, N_u) bins - is updated according to the photon time delay T and the detector pixel (u, v) where the photon is recorded. The photon time delay is defined as

$$T = T_{arrival} - T_{min}, \text{ with } T_{arrival} = t_{tot} + \tau + \xi, \quad (4)$$

where t_{tot} is the total TOF of the photon, which is approximated as the total path length traveled by the photon divided by c, while τ and ξ are random variables sampled from the distributions $PS(t)$ and $TR(t)$, respectively. T_{min} is the time to travel the distance between the source and the detection position at speed c (i.e., the ideal time of a ballistic photon).

The simulation code is implemented using CUDA to leverage GPU acceleration.

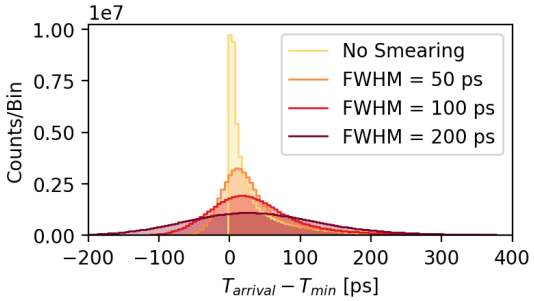


Fig. 1. Time distributions T of scattered photons in one projection obtained from simulations with different total time smearing, namely 50 ps, 100 ps and 200 ps. They are compared with the time distribution of scattered photons obtained from a simulation assuming instantaneous source emission and no timing jitter.

C. Scatter Rejection

If the TOF of scattered photons is higher than that of primary photons, a scatter rejection method can be implemented leveraging the time information. Based on (4), the time delay T of primary photons is simply $T = \tau + \xi$. Therefore, provided that the X-ray source emits sufficiently fast pulses and the detector exhibits low timing jitter, it should be possible to separate the scatter tails from the primary signal. In this work, we implement a scatter rejection method based on thresholding the combined primary and scatter time distribution T and evaluate its effectiveness in different scenarios, namely varying the source pulse shape and detector timing jitter. The threshold is chosen so that late-arriving photons are rejected while preserving as much as possible of the primary signal.

To investigate relative trade-offs for different thresholds and scenarios, we define the *signal efficiency* ε as the fraction of the total primary events P that fall below the time threshold T':

$$\varepsilon = \frac{P(T < T')}{P_{tot}}, \quad (5)$$

as well as the *scatter rejection efficiency* η as the fraction of scatter events S correctly excluded by the threshold:

$$\eta = 1 - \frac{S(T < T')}{S_{tot}} \quad (6)$$

For each scenario, the threshold T' is determined from the primary distribution so that a certain signal efficiency is achieved. To evaluate the performance of the proposed method, we will compare the results of filtered-backprojection (FBP) reconstructions obtained before and after setting thresholds T' corresponding to $\varepsilon \approx 56\%$, 65%, 73%, 80%, 86%, 90%, 96% and 99%. Specifically, we will compute the RMSE of these reconstructions using the reconstruction obtained from data without scatter as ground truth.

D. Simulated CT System Setup

The proposed TOF scatter rejection method was evaluated using the following simulated CT system. The source spectrum was generated using Spektr [15], with a 90 kVp (tungsten anode) spectrum and 6.0 mm Al filtration. The bare beam intensity was set to $I_0 = 10^5$ photons/pixel and was uniform across the detector. We modeled a flat photon counting detector with 512×128 pixels of dimensions $0.6 \times 0.6\,\text{mm}^2$. Detection events were discretized into 501 time bins of 5 ps width, spanning the interval from -500 ps to 2000 ps, with one bin reserved for underflow and overflow events. In this study, we presumed that both the source pulse $PS(t)$ and the detector timing resolution $TR(t)$

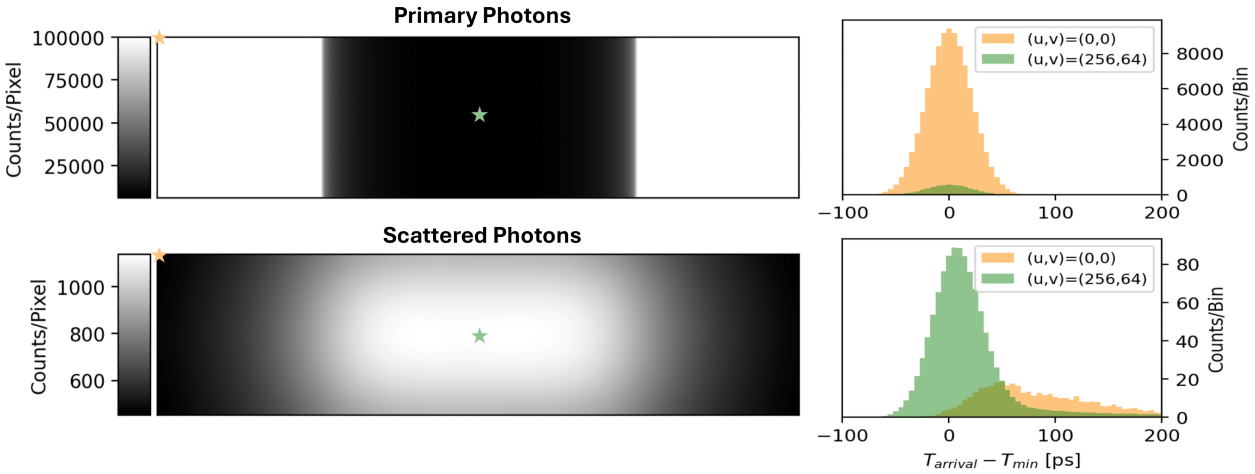


Fig. 2. Left: two-dimensional histogram of a single projection photon counts as a function of the detector coordinate. Right: time delay distribution in two detector pixels. The top row figures are related to primary photons, the bottom row to scatter.

have a Gaussian shape, so that the combined effect on the time measurements is that of a Gaussian with FWHM = 50 ps, 100 ps and 200 ps, respectively.The simulated digital phantom was a water cylinder, with diameter of 100 mm and height of 128 mm. The reconstruction volume had dimensions of $512 \times 512 \times 128$ voxels, with dimensions $0.5 \times 0.5 \times 1.0\,\text{mm}^3$. The source-to-axis distance was set to 600 mm and the source-to-detector distance to 900 mm. A total of 180 projections were generated, with 2° sampling steps. The Monte Carlo estimate of the scatter map was smoothed with a Gaussian kernel ($\sigma = 10$ pixels) before being added to the projection data. Finally, Poisson noise was applied.

III. Results

A. Simulation Results

Figure 1 shows the results of four scatter Monte Carlo simulations. Specifically, the time delay distributions T obtained from simulation assuming a total Gaussian time smearing with $\text{FWHM} = \sqrt{\text{FWHM}^2_{PS(t)} + \text{FWHM}^2_{TR(t)}} = 50\,\text{ps},\ 100\,\text{ps}$ and 200 ps are compared with the time distribution obtained under the assumption of instantaneous source emission and no detector timing jitter. In the absence of time smearing, the quasi-exponential tail of the scatter TOF is clearly visible. Introducing a temporal smearing results in a convolution with the combined $PS(t)$ and $TR(t)$ functions, which increasingly blur this tail.

Figure 2 presents the photon distribution on the detector for primary (polyenergetic Beer's law) and scattered photons (Monte Carlo simulation). In this case, the total temporal FWHM was set to 50 ps. The time delay distribution of T is also shown for a pixel in the center and in the corner of the detector: for primary photons, the distribution is obtained by scaling a Gaussian probability density function with the chosen FWHM by the total number of counts in each pixel. For scatter, the shape of the time delay histogram can vary depending on the areas of the detector considered. In particular, scattered photons detected near the corner of the detector generally have a higher time delay, as they travel longer path lengths (they have to pass through the object and to the detector corner). This observation suggests that setting spatially-varying thresholds T' could be beneficial and improve scatter rejection. In this initial study, a single global threshold will be adopted for simplicity.

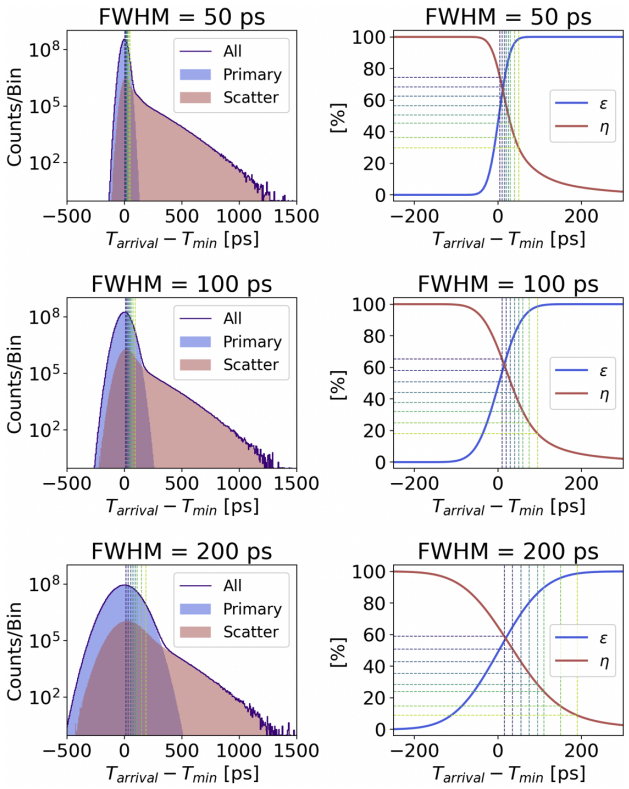


Fig. 3. Time delay distribution of primary and scattered photons in a single projection (left), signal efficiency ε and scatter rejection efficiency η (right) for each time smearing scenario (FWHM $= 50, 100, 200$ ps). The dashed lines represent the thresholds selected for scatter rejection and the corresponding scatter rejection efficiency values. The same color is used for thresholds corresponding to the same signal efficiency.

B. Scatter Rejection Results

Given the time delay distributions of primary and scattered photons, signal efficiency ε and scatter rejection efficiency η are calculated according to (5) and (6). Figure 3 shows the results for one projection. For the three FWHM cases considered, the scatter rejection thresholds were selected to achieve $\varepsilon \approx 56\%,\ 65\%,\ 73\%,\ 80\%,\ 86\%,\ 90\%,\ 96\%$ and 99% (with small variations due to the binning of the time delay histograms). Specifically, for FWHM $= 50$ ps, the corresponding thresholds are $T' = 5,\ 10,\ 15,\ 20,\ 25,\ 30,\ 40$ and 50 ps; for FWHM $= 100$ ps, $T' = 10,\ 20,\ 30,\ 40,\ 50,\ 60,\ 75$ and 95 ps;

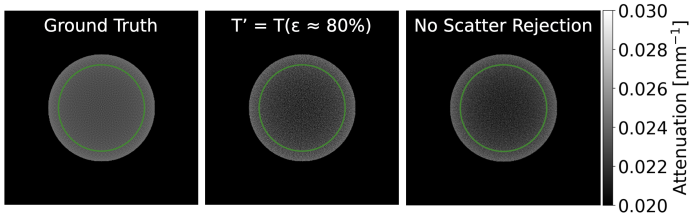


Fig. 4. From left to right: central slice of FBP reconstruction of projection data assuming FWHM $= 50\,\text{ps}$ with no added scatter (ground truth); with scatter and threshold of $T' = 20\,\text{ps}$ applied; and with scatter no threshold applied. The green circle delimits the ROI used for RMSE calculation.

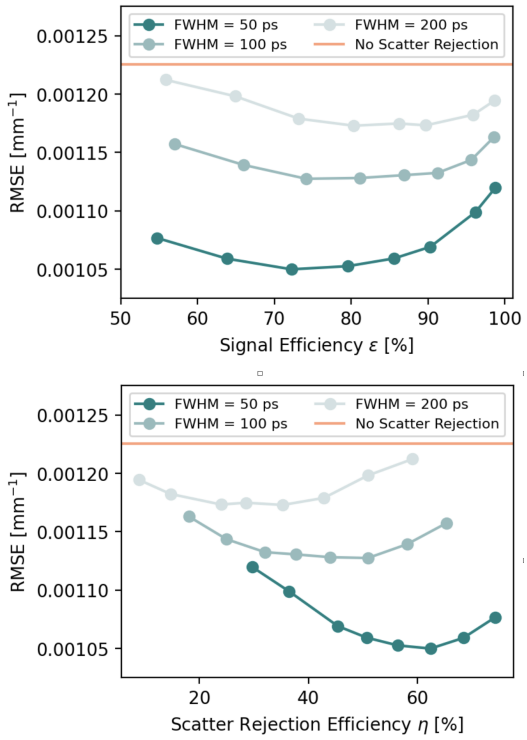


Fig. 5. RMSE as a function of the signal efficiency (top) and the scatter rejection efficiency (bottom) of the selected thresholds for scatter rejection. The RMSE of the uncorrected reconstruction is also shown for comparison.

and for FWHM $= 200\,\text{ps}$, $T' = 15, 35, 55, 75, 95, 110, 150$ and $190\,\text{ps}$. For each of the three FWHM cases considered, the scatter rejection efficiency η at the selected thresholds decreases with increasing signal efficiency. In addition, the same signal efficiency corresponds to a lower rejection efficiency as the time smearing increases. Therefore, we expect that the reconstruction quality will degrade with higher thresholds and larger time smearing. However, reducing the signal efficiency implies that a smaller fraction of the primary photons are used in the reconstruction, leading to noisier and inaccurate images. Figure 3 also illustrates the time delay distributions T of primary photons and scatter for one projection together with the selected threshold locations. Due to the symmetry of the setup, the same thresholds are applied to all projection angles.

Figure 4 shows FBP reconstructions corresponding to the scenarios where FWHM $= 50\,\text{ps}$. The cupping artifact is clearly visible in the central slice of the reconstruction obtained from projection data where no scatter rejection is applied, and it is reduced when a threshold $T' = 20\,\text{ps}$ is applied.

Figure 5 summarizes performance in terms of RMSE for all reconstructed data (TOF resolution and thresholds). RMSE is computed on the central slice of the reconstructed volume, in a circular area of diameter $80\,\text{mm}$ centered on the water cylinder, where cupping artifacts due to residual scatter are more pronounced. As expected, there is a trade-off between signal efficiency and scatter rejection efficiency, with higher scatter rejection efficiency resulting in lower RMSE, but only as long as a large fraction ($\varepsilon \geq 70 - 80\,\%$) of primary photons is retained. For lower values of ε, the image quality degrades again due to insufficient primary and quantum noise. Finally, the results demonstrate that the threshold-based scatter rejection method is more effective when the combined pulse shape and detector time resolution FWHM are smaller.

IV. Conclusion

In this work, we implemented a Monte Carlo simulator to model a TOF X-ray CT system. There are a few limitations on the presented simulation process that will be improved in future work: for instance, we are not considering the object's refractive index for the computation of the TOF; we are neglecting the time smear inside pixels and additional detection-related issues, such as deadtime, pile-up, and clock stability. We also plan to incorporate more realistic pulse shape and time resolution functions.

The presented scatter rejection method based on time delay thresholding has been demonstrated to improve image quality. However, it exhibits a trade-off between signal efficiency and scatter rejection efficiency. Using image quality metrics such as RMSE helps identify the optimal operating range, achieving a balance between primary photon retention and scatter rejection.

In ongoing work, we aim at developing more advanced TOF scatter control techniques that will combine scatter rejection (e.g. via thresholding) with scatter correction (e.g. by estimating the residual scatter component using the tail of the scatter distribution). These efforts will help inform the development of future TOF CT systems.

Acknowledgments

This research has been funded by the Industrial Technology Innovation Program [P0030253, Development of the world 1st TOF detector DXA for fast and low-radiation scan] of the Ministry of Trade, Industry and Energy of the Republic of Korea.

References

[1] Siewerdsen, J. H and Jaffray, D. A, *Medical Physics*, vol. 28, no. 2, pp. 220–231, 2 2001. [Online]. Available: http://doi.wiley.com/10.1118/1.1339879
[2] Rührnschopf, E.-P and Klingenbeck, K, *Medical physics*, vol. 38, no. 7, pp. 4296–4311, 2011.
[3] ——, *Medical physics*, vol. 38, no. 9, pp. 5186–5199, 2011.
[4] Barnes, G. T, *RadioGraphics*, vol. 11, no. 2, pp. 307–323, 3 1991.
[5] Mail, N *et al.*, *Medical Physics*, vol. 36, no. 1, pp. 22–32, 1 2009.
[6] Neitzel, U, *Medical Physics*, vol. 19, no. 2, pp. 475–481, 3 1992.
[7] Zhu, L *et al.*, *Medical Physics*, vol. 36, no. 6Part1, pp. 2258–2268, 6 2009.
[8] Zbijewski, W and Beekman, F, *IEEE Transactions on Medical Imaging*, vol. 25, no. 7, pp. 817–827, 7 2006.
[9] Sun, M and Star-Lack, J. M, *Physics in Medicine and Biology*, vol. 55, no. 22, pp. 6695–6720, 11 2010. [Online]. Available: https://iopscience.iop.org/article/10.1088/0031-9155/55/22/007
[10] Lorenzon, A *et al.*, in *Medical Imaging 2025: Physics of Medical Imaging*, vol. 13405. SPIE, 2025, pp. 190–195.
[11] Rossignol, J *et al.*, *Physics in Medicine & Biology*, vol. 65, p. 085013, 4 2020.
[12] Woodcock, E *et al.*, in *Proc. Conf. Applications of Computing Methods to Reactor Problems*, vol. 557, no. 2. Argonne National Laboratory, 1965.
[13] Kawrakow, I, *NRCC Report Pirs-701*, 2001.
[14] Salvat, F *et al.*, *Workshop Proceedings, Data Bank NEA/NSC/DOC(2011)5, Barcelona*, 07 2006.
[15] Punnoose, J *et al.*, *Medical Physics*, vol. 43, no. 8Part1, pp. 4711–4717, 2016.

Incorporating Tube Current Latency in Organ Risk-Based CT Tube Current Modulation

Catherine Knobloch, Philipp Oelze, Andreas Prokein, Michael Grasruck, Matthias Baer-Beck, and Marc Kachelrieß

***Abstract*—Tube current modulation (TCM) in computed tomography (CT) has become an established technique for dose reduction. However, newly proposed effective dose-minimizing TCM neglects tube current latency, which may lead to physically infeasible tube current curves, which could in practice be distorted significantly and lead to higher effective dose levels. This work proposes three methods to incorporate the tube current latency into effective dose minimizing TCM: a retrospective smoothing, direct optimization of the latent tube current, and indirect optimization by optimizing the filament current in the x-ray tube, and compares these to mAs- and dose-minimizing TCM without latency consideration. Additionally, the benefit of turning off the tube voltage in angular ranges in which effective dose-minimizing TCM is zero to force the tube current to zero (zero push) is investigated. Our results show that riskTCM with latency still reduces the effective dose when compared to mAsTCM and reaches similar dose levels as our effective dose-minimizing TCM when the zero push is incorporated into the optimizers.**

***Index Terms*—automatic exposure control, tube current modulation, dose reduction, CT, tube current latency.**

I. Introduction

Computed tomography (CT) exposes patients to ionizing radiation. Consequently, minimizing patient dose, while maintaining image quality, is of critical importance. Automatic exposure techniques seek to achieve this, one form being tube current modulation (TCM) [1] [2], which changes the tube current depending on the position of the x-ray tube during a scan. An established, conventional method is mAsTCM, which aims to minimize the current-time product for a certain noise level in the patient. This does not take into account the sensitivity of the individual organs to radiation. As such, riskTCM, which minimized the effective dose to the patient, was developed as an organ-based TCM method [3]. However, due to the anatomical nature of patients and organ positions, the effective dose that is administered can vary strongly when changing the tube current angle. This results in quick changes in the TCM curve. Since the tube current cannot be adjusted at an arbitrary rate, this poses an issue. The tube current is regulated by the filament current in the x-ray tube. A higher filament current generates an increase in the number of electrons that are emitted and accelerated towards the anode, generating a higher tube current. The opposite is the case when lowering the tube current. Fewer electrons leave the filament, in turn generating a lower tube current. Although increasing the tube current can be done relatively quickly, decreasing the tube current is limited by the slow cooling rate of the filament. Assuming that both Newton's law of cooling can be applied to the filament temperature/current and that the tube current is directly proportional to the filament current, the fastest lowering of the tube current occurs when the filament current is set to $0\,\mathrm{mA}$ and would result in the exponential decay of the tube current. With a rotation time of $t_{\mathrm{rot}} = 500$ ms and a cooling constant $\tau = 90\,\mathrm{ms}$, typical values for abdominal CT scans, a reduction to 1% would take $414\,\mathrm{ms}$ (83% of the rotation). Since riskTCM also has entire angular ranges in which the tube current should be set to zero, this is a rather significant limitation that must be addressed. This work focuses on incorporating the tube current latency into the riskTCM optimization framework, minimizing the effective dose, while respecting the rate of tube current adjustment. Specifically, this work will focus on the limitation when lowering the tube current, since this has a higher latency. Of course, the presented methods can also be extended to the dynamics of increasing the tube current.

Catherine Knobloch, Philipp Oelze and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls University, Heidelberg, Germany. Catherine Knobloch, Dr. Andreas Prokein, Dr. Matthias Baer-Beck, and Dr. Michael Grasruck are with Siemens Healthineers AG, Forchheim.
Corresponding author: catherine.knobloch@dkfz.de

II. Methods

A. Application of thermodynamics to tube current adjustments

Newton's law of heating and cooling, a differential equation describing temperature change of an object depending on the ambient temperature, is applied to iteratively determine the tube current at discrete time steps i. The object temperature and ambient temperature are replaced by the tube current I and the filament current I_{F}, respectively, resulting in the following equation

$$I(i+1) = I(i) + \tau(I_{\mathrm{F}}(i+1) - I(i)). \tag{1}$$

with $\tau = \frac{t_{\mathrm{rot}}}{Nk}$ as the heating/cooling rate with a rotation time t_{rot}, number of samples per rotation N, and the heating/cooling constant k. Since this work focuses on the limitation that comes from lowering the tube current, $\tau = 1$

is assumed for all cases in which the tube current is increased, resulting in $I(i{+}1) = I_{\mathrm{F}}(i{+}1)$ if $I_{\mathrm{F}}(i{+}1) > I(i)$. The fastest tube current decrease can be reached by setting $I_{\mathrm{F}} = 0\,\mathrm{mA}$, since negative currents are not feasible. This decrease corresponds to the aforementioned exponential decrease of the output current.

B. Patient model

The TCM methods are demonstrated on data of a whole body CT scan, of which an anterior-posterior projection is shown in figure 1, with previously calculated organ segmentations for a riskTCM study [3]. The patient is sub-sampled into 122 slices and divided into typical anatomical regions, noted in table I with realistic rotation durations t_{rot} and cooling constants τ.

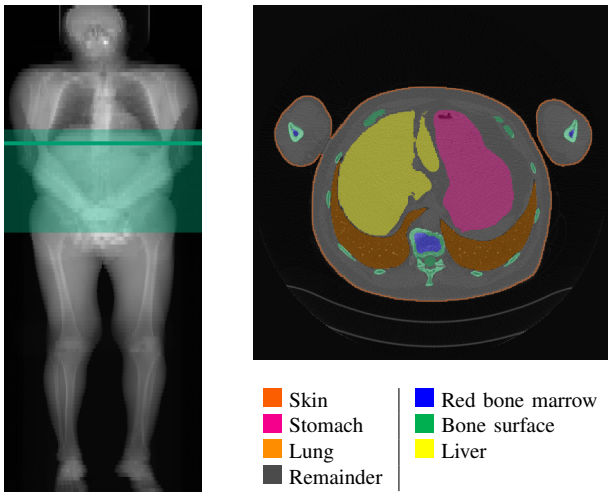


Fig. 1. Anterior-posterior projection of the patient data ($512 \times 512 \times 122$) measured on a Siemens Somatom Flash CT scanner at the DKFZ (left). The green area represents the abdominal scan region (slice 31 to 57). A single slice with segmentation (top right) which corresponds to the horizontal green line in the patient (slice 35) is shown with an organ legend (bottom right).

C. Baseline Methods

This work compares its methods with both the conventional mAsTCM and riskTCM [3] applied to sequential scans, which both do not consider tube current latency. The same noise minimization as in reference [3] was used, except that the limited-memory Boyden-Fletcher-Goldfarb-Shanno method with bounds (L-BFGS-B) [4] was used as an optimizer. This quasi-Newtonian method optimizes smooth functions $f(x)$, typically with many input variables. The noise function, minimized while keeping the effective dose constant, is as follows

$$N_{\mathrm{riskTCM}}(I) = \int dx\, dy\, w(\boldsymbol{r}) \sqrt{\int \frac{e(p)}{I_{\mathrm{D}} + I_{\mathrm{C}}} d\vartheta}$$

with $e(p)$ as the polychromatic exponential function, dependent on the projection value $p(\alpha(\vartheta, \boldsymbol{r}), \beta(\vartheta, \boldsymbol{r}))$, the direct $I_{\mathrm{D}} = I(\alpha_{\mathrm{D}}(\vartheta, \boldsymbol{r}))$ and complementary $I_{\mathrm{C}} = I(\alpha_{\mathrm{C}}(\vartheta, \boldsymbol{r}))$ rays going through the same $\boldsymbol{r}$. The angle $\alpha(\vartheta, \boldsymbol{r})$ is the position angle of the x-ray source for a ray going through voxel $\boldsymbol{r}$ under the angle ϑ w.r.t. the coordinate system. Similarly $\beta(\vartheta, \boldsymbol{r})$ is the angle within the fan for the very same ray. The function $w(\boldsymbol{r})$ is a weighting function that weights voxels within the patient with 1 and voxels outside with 0. The effective dose, which is held constant during the optimization, was previously determined with an in-house Monte Carlo simulation software, based on the Monte Carlo software described in reference [5], by simulating per-angle dose distributions. These were then used to determine the organ doses and weighted with the tissue weighting factors recommended by the International Commission on Radiological Protection (ICRP) [6] to determine per angle effective Dose $D_{\mathrm{eff}}(\alpha)$ values, which had previously been done for a riskTCM study [3].

D. Considering the latent output current

Three methods that incorporate the tube current latency into the riskTCM optimizer are presented. To compare the following methods, the different TCM curves are normalized to the same noise level and the dose reduction with respect to the conventional mAsTCM is compared between methods. The latency between individual sequential slices are neglected.

1) Retrospectively smoothed tube current correction: This method, referred to as smoothed riskTCM, retrospectively smooths the riskTCM current values after an initial riskTCM optimization:

$$I(i+1) = \begin{cases} I(i+1) & \text{if } I(i+1) \geq I(i)e^{-\tau} \\ I(i)e^{-\tau} & \text{else} \end{cases}.$$

If the rate with which the tube current decreases is faster than the allowed exponential decay of the tube current, the values are adjusted.

2) Latent tube current optimization: Since the L-BFGS-B optimization algorithm [4] that is used for riskTCM can implement bounds for each tube current value, these can be set so that the tube current can only reach feasible values (for the case that $I_{\mathrm{F}} = 0$). However, these boundaries can only be defined once at the beginning of an optimization process, which poses a problem, since changing the tube current after each iteration also changes the feasible boundaries. Instead, the optimizer is applied sequentially, with only one iteration per run. After each single iteration, the bounds are updated and the optimization restarted. This method will be referred to as latent riskTCM.

3) Filament current optimization: Since the tube current is a function of the filament current, instead of optimizing the tube current, the filament current can be optimized instead. This is advantageous, as the lower bounds of the filament can be set to $0\,\mathrm{mA}$ for the entire optimization. The

output tube current can then be calculated from equation (1). However, this method uses a different gradient in the riskTCM optimization.

$$\frac{\partial N}{\partial I_{\mathrm{F}}(i)} = \int w(\boldsymbol{r}) \frac{-1}{2\sqrt{\int d\vartheta \frac{e(p)}{I_{\mathrm{D}}+I_{\mathrm{C}}}}} \cdot \cdot \Big(\int \frac{e(p)}{(I_{\mathrm{D}}+I_{\mathrm{C}})^2} \frac{\partial I_{\mathrm{D}}}{\partial I_{\mathrm{F}}(i)} d\vartheta \Big) dx\, dy. \quad (2)$$

Because of its iterative nature, the tube current can also be dependent on previous filament currents, resulting in the following equation for the partial derivative of I_{D}

$$\frac{\partial I_{\mathrm{D}}}{\partial I_{\mathrm{F}}(i)} = \tau_i \prod_{j=i+1}^{k(I_{\mathrm{D}})} (1-\tau_j).$$

with $k(I_{\mathrm{D}})$ being the index of the direct ray. As can be recognized by the equation, this partial derivative becomes zero if any $\tau_j = 1$, which is the case if the tube current does not decrease in step j. Upon receiving initial results, it was observed that a better dose reduction was achieved by omitting the product term, and instead using

$$\frac{\partial I_{\mathrm{D}}}{\partial I_{\mathrm{F}}(i)} = \begin{cases} \tau_i & \text{if } I_{\mathrm{F}}(i) < 0.95 I_{\mathrm{D}} \text{ and } \vartheta(k(I_{\mathrm{D}})) = \vartheta(i) \\ 1 & \text{else} \end{cases}.$$

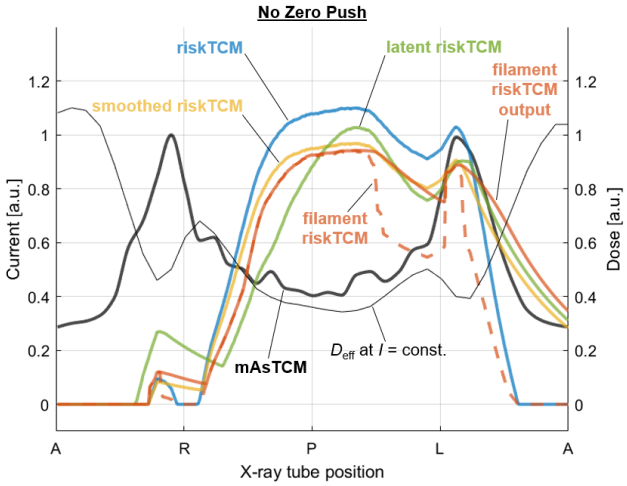
The factor 0.95 allows small filament current noise, without impacting the gradient multiplier.

E. Pushing the current to 0 mA

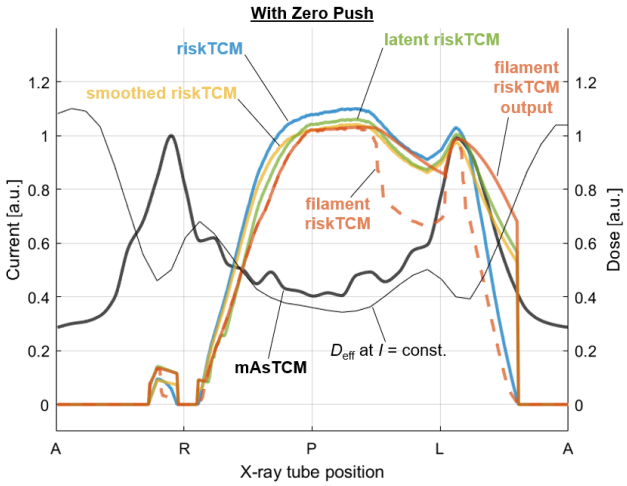
A significant difference between riskTCM and mAsTCM are the regions in which the tube current falls to 0 mA in riskTCM. However, due to tube current latency and typical rotation times and cooling rates, it is difficult to quickly reach low tube current values. Turning off the tube voltage could be a method to almost instantaneously push the tube current to zero. An initial riskTCM optimization without latency is used to determine angular ranges in which the tube current is zero. In these regions, during the optimization of filament and latent riskTCM, the lower and upper bounds of the optimizer are set to force the tube current to nearly zero, which avoids division by zero in equation (2). For the smoothed riskTCM, the zero push is done retrospectively.

III. Results

This section will compare the dose of the three riskTCM methods that take the latency of the tube current into consideration to mAsTCM and riskTCM and the effect of zero pushing for different rotation times and cooling constants (see table I). For comparability, all results are normalized to the same noise level per slice. Results in the abdominal region will be shown in a variety of figures to provide a better representation than merely individual values.



(a)



(b)

Fig. 2. TCM curves without (a) and with (b) zero push in the abdomen and effective dose at constant current for slice 35. All tube current curves are normalized to the same noise.

A. Results in the abdominal region

The investigated region and an example slice, for which TCM curves are provided in figures 2a and 2b, are shown in figure 1. These curves provide a good understanding of the tube current latency when comparing the riskTCM curve (in blue) with the smoothed (yellow), latent (green), and the output of the filament riskTCM (orange), which do not decrease as quickly. Zero pushing can also clearly be seen in figure 2b, in which all curves instantaneously fall to zero for regions in which riskTCM is also zero. Note that after turning the voltage back on, as seen around the R position, the tube current does not begin at zero, because the filament has not fully cooled down. Due to the varying anatomy of the patient in the z-direction, the tube current looks very different for each slice and achieves different effective dose reductions (in

Region	t_{rot}	k	riskTCM	smoothed riskTCM		latent riskTCM		filament riskTCM	
				no zp	zp	no zp	zp	no zp	zp
Abdomen	500 ms	90 ms	75%	82%	**76%**	82%	**76%**	82%	77%
		45 ms	75%	77%	**75%**	78%	**75%**	78%	76%
Head	1000 ms	90 ms	91%	92%	**91%**	92%	**91%**	92%	**91%**
		45 ms	91%	**91%**	**91%**	**91%**	**91%**	**91%**	**91%**
Neck	500 ms	90 ms	76%	85%	**77%**	84%	**77%**	86%	**77%**
		45 ms	76%	80%	**76%**	80%	**76%**	81%	77%
Pelvis	500 ms	90 ms	80%	86%	**81%**	86%	**81%**	87%	82%
		45 ms	80%	82%	**81%**	82%	**81%**	83%	**81%**
Thorax	330 ms	90 ms	81%	88%	84%	88%	**83%**	89%	84%
	250 ms	45 ms	81%	86%	83%	86%	**82%**	87%	83%

TABLE I
DOSE AT THE SAME NOISE RELATIVE TO THE MASTCM DOSE. VALUES ARE ROUNDED TO THE NEAREST INTEGER. THE LOWEST DOSE VALUES USING TCM METHODS THAT ACCOUNT FOR LATENCY ARE SHOWN IN BOLD. NOTE THAT MASTCM DOSE IS ALWAYS AT 100%.

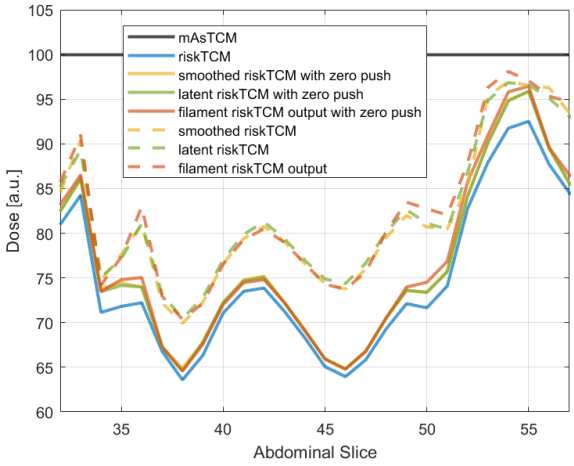


Fig. 3. Effective dose for each abdomen slice ($T = 500\,\text{ms}$ and $k = 90\,\text{ms}$) for the different optimization methods. A level of 100% corresponds to the dose level of mAsTCM at the same noise level.

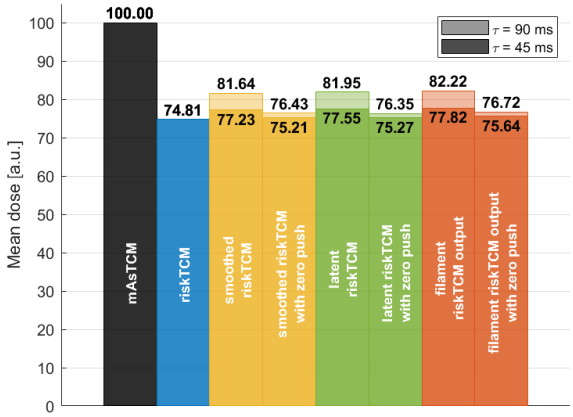


Fig. 4. Comparison of effective dose, in % of the mAsTCM dose, for the different riskTCM methods with rotation times of 500 ms. Saturated bars and lighter bars depict results for $\tau = 45\,\text{ms}$ and $\tau = 90\,\text{ms}$, respectively.

comparison to mAsTCM). Figure 3 shows the effective dose for each riskTCM-based method and mAsTCM for a rotation time $T = 500\,\text{ms}$ and a cooling constant of $\tau = 90\,\text{ms}$. For all slices and methods a significant dose reduction can be observed, especially when introducing a zero push. Figure 4 shows the mean effective dose of all slices in the abdominal region. When comparing the different methods with one another, the same rotation time and cooling constants seem to perform very similarly, unexpectedly smoothed and latent riskTCM slightly outperform filament riskTCM. In the abdomen all methods show a large effective dose reduction compared to mAsTCM, both slice-wise and w.r.t. the mean effective dose reduction.

B. Comparison in all body regions

Table I shows that all methods have a significant advantage compared to mAsTCM (100% effective dose) and the zero push also has a significant effective dose decrease compared to the methods without zero push, especially for slower cooling rates. All methods using zero push achieve similar dose levels, the latent riskTCM slightly outperforming the other methods for most body regions. This is very unexpected, as a direct optimization of the filament current should render the best results and should specifically lead to better results compared to the smoothed riskTCM, which is also not the case. For all regions, except the thorax, using a zero push achieves an effective dose that is higher than that of riskTCM without latency by merely $<1\%$ of the mAsTCM effective dose for both cooling rates of 45 ms and 90 ms. X-ray tubes with slower cooling rates benefit more with a zero push.

IV. Conclusions

The methods with which the tube current latency can be incorporated into the riskTCM optimization were investigated, as well as the incorporation of a zero push. Our results suggest that even with the tube current latency incorporated in the optimizer, riskTCM still can reach a much lower effective dose when compared to mAsTCM. Especially when a zero push is added, all methods attain a similar effective dose as riskTCM without latency. In the future, the practical feasibility will be determined by real-world measurements on CTs.

Acknowledgment

The Monte Carlo dose estimation method RayConStruct-MC was provided by by RayConStruct® GmbH.

References

[1] M. Gies, W. A. Kalender, H. Wolf, C. Suess, and M. T. Madsen, "Dose reduction in CT by anatomically adapted tube current modulation. I. Simulation studies," *Med. Phys.*, vol. 26, no. 11, pp. 2235–2247, Nov. 1999.

[2] W. A. Kalender, H. Wolf, and C. Suess, "Dose reduction in ct by anatomically adapted tube current modulation. ii. phantom measurements," *Medical Physics*, vol. 26, no. 11, pp. 2248–2253, 1999. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.598738

[3] L. Klein, C. Liu, J. Steidel, L. Enzmann, M. Knaup, S. Sawall, A. Maier, M. Lell, J. Maier, and M. Kachelrieß, "Patient–specific radiation risk–based tube current modulation for diagnostic CT," *Med. Phys.*, vol. 49, no. 7, pp. 4391–4403, Jul. 2022.

[4] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[5] M. Baer and M. Kachelrieß, "Hybrid scatter correction for CT imaging," *Phys. Med. Biol.*, vol. 57, pp. 6849–6867, 2012.

[6] "ICRP publication 103: The 2007 recommendations of the international commission on radiological protection," *Ann. ICRP*, vol. 37, no. 2–4, 2007.

Low Contrast Detectability Performance in a Realistic Anthropomorphic 3D Printed Phantom vs standard QA phantoms

Daniel W. Shin, Leening Liu, Jessica Im, Tomoe Hagio, Bernice Hoppel, Amar Dhanantwari, Peter Noel, Kirsten L. Boedeker

Abstract – **Performance evaluation of non-linear CT reconstruction algorithms is inherently influenced by the choice of test object. This study compares trends in low-contrast detectability (LCD) performance across multiple reconstruction algorithms using three phantoms: a realistic anthropomorphic 3D-printed abdominal phantom, the ACR CT accreditation phantom, and the Catphan 600. The anthropomorphic abdominal phantom was produced using PixelPrint, a voxel-based 3D printing technique that converts CT imaging data into printer instructions by modeling density as the ratio of filament to voxel volume, thereby generating sub-resolution partial volume effects that mimic realistic tissue densities and textures. The phantom consisted of insert containing ten low-contrast rods (5 mm diameter, 1 cm length, 30 HU contrast) in native anatomy placed within a water-equivalent ellipse to achieve a total water-equivalent diameter of 34 cm. All phantoms were scanned using a standard abdominal CT protocol at a size-specific dose estimate of 12 mGy. Images were reconstructed using filtered backprojection, two hybrid iterative reconstruction algorithms, two deep learning denoising reconstruction algorithms, and a super-resolution deep learning reconstruction method. LCD performance was quantified using a non-prewhitening model observer with an eye filter. Detectability improvements in advanced reconstructions relative to filtered backprojection varied considerably by phantom. The ACR phantom yielded the smallest relative gains in d′, the PixelPrint liver phantom showed the greatest improvements, and the Catphan demonstrated intermediate behavior. These findings highlight the critical role of phantom realism in task-based LCD evaluation and suggest that protocol optimization based solely on standard ACR phantom measurements may not fully translate to clinically relevant imaging conditions.**

Index Term – **ACR phantom, 3D printed phantom, Catphan, Anthropomorphic phantom, Model Observer, Low Contrast Detectability**

I. INTRODUCTION

Computed tomography (CT) plays a central role in the detection and evaluation of focal liver lesions, where diagnostic accuracy is often constrained by the ability to visualize subtle, low-contrast targets within intrinsically heterogeneous hepatic tissue. To address these challenges, modern reconstruction techniques such as iterative reconstruction (IR) and deep learning-based reconstruction (DLR) have been developed with the goal of improving image quality and lesion conspicuity while supporting radiation dose reduction.

CT image quality has traditionally been characterized using standardized phantoms and metrics such as spatial resolution, image noise, and contrast-to-noise ratio (CNR). The American College of Radiology (ACR) CT accreditation phantom is routinely used for system evaluation, quality assurance, and accreditation, incorporating measurements of these standard metrics as well as a low-contrast detectability (LCD) task module [1][2][3]. The LCD task module consists of uniform cylindrical targets embedded within a homogeneous background. While the ACR phantom enables highly reproducible and standardized assessments, it lacks the size, anatomical texture and structural variability characteristic of clinical abdominal imaging. Consequently, detectability measurements derived from the ACR phantom may not fully reflect performance for clinically relevant tasks such as liver lesion detection.

Task-based image quality assessment provides a framework for addressing this limitation by explicitly linking image quality to a defined detection task. Within this framework, model observers offer objective and repeatable surrogates for human performance. In this study, low-contrast detectability is evaluated using a non-prewhitening model observer with an eye filter (NPWE), which accounts for human visual response [4-7].

Recent advances in additive manufacturing have enabled the development of anthropomorphic phantoms that more accurately replicate patient anatomy and texture. PixelPrint technology enables voxel-wise control of CT attenuation, allowing realistic reproduction of organ-specific textures and contrast distributions. In abdominal imaging applications, PixelPrint phantoms can incorporate liver-mimicking backgrounds with embedded low-contrast lesions, providing a controlled yet anatomically relevant background. Despite the increasing availability of such phantoms, systematic studies examining how phantom realism influences model observer based conclusions remain limited.

The purpose of this study is to investigate how phantom selection affects task-based assessment of low-contrast detectability across different CT reconstruction algorithms. LCD is evaluated using the ACR CT accreditation phantom, the Catphan phantom, and a PixelPrint liver phantom, spanning a range from standardized quality assurance objects to anatomically realistic models. The Catphan phantom serves as a secondary reference, offering a more robust, less scatter-prone LCD module than the ACR phantom while also being

D. Shin, T. Hagio, and K. Boedeker are employees of Canon Medical Systems Corporation, Otawara-shi, Japan.

B. Hoppel and A. Dhanantwari are employees of Canon Medical Systems USA, Tustin, USA.

L. Liu, J. Im, and P. Noel are affiliated with University of Pennsylvania, Philadelphia, USA.

widely used for CT system characterization. It is hypothesized that detectability metrics and reconstruction algorithm rankings derived using the ACR and Catphan phantoms will differ from those obtained with the anthropomorphic PixelPrint phantom, highlighting the influence of anatomical texture and limiting the generalizability of conclusions drawn from conventional QA phantoms to clinically relevant imaging tasks.

II. METHOD AND MATERIALS

A. Phantoms

Three physical phantoms were evaluated including the PixelPrint liver phantom (University of Pennsylvania, Philadelphia, PA), the ACR CT accreditation phantom (Gammex 464, Mirion Technologies, Atlanta, GA), and the Catphan 600 (The Phantom Laboratory, Salem, NY).

The PixelPrint liver phantom consisted of two modules: one containing native anthropomorphic anatomy while the other additionally included ten low-contrast cylindrical rods with a z-length of 1 cm, a diameter of 5 mm, and a contrast of 30 HU. To achieve clinically relevant attenuation conditions, the PixelPrint phantom was placed within a water-equivalent ellipse, resulting in an effective water-equivalent diameter of 34 cm.

The ACR CT accreditation phantom included the low-contrast detectability (LCD) module, which contains uniform cylindrical targets with diameters ranging from 2 mm to 6 mm at approximately 6 HU contrast. For consistency across phantoms in terms of object size, only the 5 mm diameter target was used for analysis. The ACR phantom has a physical diameter of 20 cm and was therefore scanned both with and without an attenuation ellipse. When the attenuation ellipse was applied, the resulting effective water-equivalent diameter was 31 cm, enabling a more comparable attenuation condition relative to the PixelPrint phantom. Note that the contrast of the PixelPrint phantom was intentionally set higher than that of the ACR phantom to compensate for the presence of anatomical background noise.

For the Catphan 600, the CTP515 low contrast module was used. This module includes low contrast cylindrical targets, and the 5 mm diameter object with a contrast of 5 HU was selected for analysis to match object size across phantoms. The Catphan has the same physical diameter as the ACR phantom and was included to allow a direct comparison between the two standardized QA phantoms. Accordingly, the Catphan was scanned only without an attenuation ellipse in this study.

B. Image Acquisition

All the phantoms were scanned with a typical abdominal protocol using a Size Specific Dose Estimate (SSDE) of 12mGy on a CT scanner (Aquilion ONE INSIGHT, Canon Medical Systems). To account for differences in physical phantom size, scan parameters were adjusted as needed to achieve an equivalent SSDE across phantoms. In general, acquisitions were performed at 120 kVp with a 0.5 s rotation time, a standard pitch of 0.813, and ten repeated scans per configuration. Images were reconstructed using filtered backprojection (FBP), two hybrid iterative reconstruction algorithms (HIR1 and HIR2), two deep learning–based reconstruction algorithms (DLR1 and DLR2), and a super-resolution deep learning reconstruction (SR-DLR), all with a slice thickness of 3 mm. For illustrative purposes, representative averaged reconstructed images of the phantoms are shown in Figure 1.

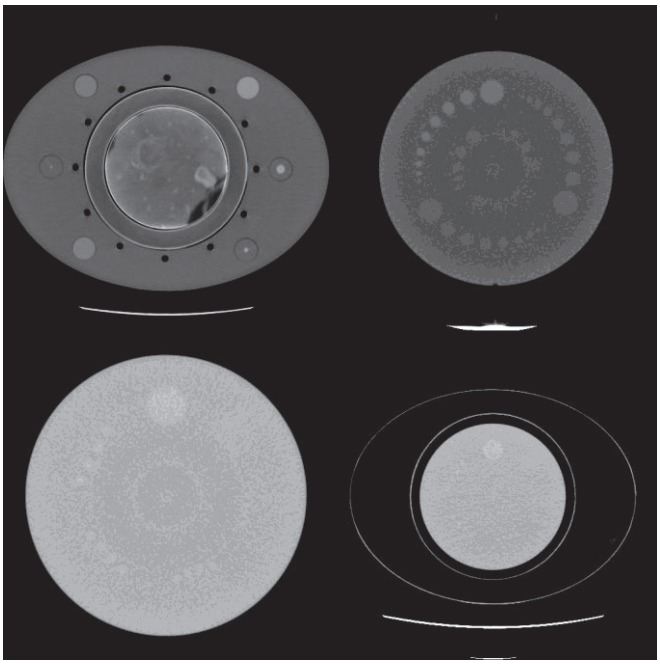


Fig. 1. Representative averaged images of the evaluated phantoms. Top left: PixelPrint liver phantom with the low contrast rods (window/level = 400/40). Top right: Catphan 600 CTP 515 (window/level = 100/80). Bottom left: ACR phantom LCD module (window/level = 100/80). Bottom right: ACR phantom with the ellipse (window/level = 100/80).

C. Data Analysis

Low-contrast detectability (LCD) performance was quantified using a non-prewhitening model observer with an eye filter (NPWE), previously validated against human observer [6]. An observer template was generated using contrast matched to the test object and applied to the phantom images to compute cross-correlation–based test statistics. For each phantom, more than 200 paired realizations of signal-present and signal-absent cases were analyzed. The detectability index (d′) was calculated according to the following equation.

$$d' = \frac{\langle T \rangle_1 - \langle T \rangle_2}{\sqrt{\frac{1}{2}{\sigma_1}^2 + \frac{1}{2}{\sigma_2}^2}}. \quad (1)$$

where T_1 and T_2 denote the mean test statistics for the signal present and signal absent cases, respectively, and σ_1 and σ_2 represent the corresponding standard deviation [4].

Uncertainty in d′ was estimated using a bootstrapping approach. The measured test statistics were resampled with replacement 1000 times, and the resulting distribution of d′ values was used to compute error bars.

III. RESULTS

The detectability index (d′) measured using the standard 20 cm ACR CT accreditation phantom ranged from 0.73 to 1.09 across reconstruction algorithms. For the Catphan phantom, d′ values ranged from 1.24 to 1.84, which were notably higher despite equivalent test object and physical phantom diameter. When the ACR phantom was scanned with the attenuation ellipse applied, d′ values ranged from 0.46 to 0.69. For the PixelPrint liver phantom, detectability values ranged from 1.37 to 2.32. The substantial differences in absolute d′ across phantoms are largely attributable to differences in target contrast, as the ACR phantom contains low-contrast objects at approximately 6 HU, whereas the PixelPrint phantom incorporates rods with a nominal contrast of 30 HU.

To facilitate meaningful comparison across reconstruction algorithms, relative percent improvements in d′ with respect to FBP were evaluated and are summarized in Table 1. Detectability trends exhibited a strong dependence on phantom type, with the PixelPrint phantom demonstrating the largest relative improvements and the ACR phantom showing the smallest gains across advanced reconstruction methods. Visual comparisons of different reconstruction methods for each phantom are demonstrated in Figures 7-10.

Correlations in percent d′ improvement relative to FBP across reconstruction algorithms were assessed between phantoms and are shown in Figures 3–6. The relative detectability gains measured using the Catphan and ACR phantoms exhibited a weak correlation ($R^2 = 0.076$) with each other. Similarly, a weak correlation was observed between the PixelPrint and 20cm ACR phantom ($R^2 = 0.036$). Comparison between PixelPrint and the ACR phantom scanned with the attenuation ellipse yielded a higher correlation ($R^2 = 0.46$) than without the ellipse, although the relationship remained weak. In contrast, strong agreement was observed between PixelPrint and Catphan results, with an R^2 value of 0.84.

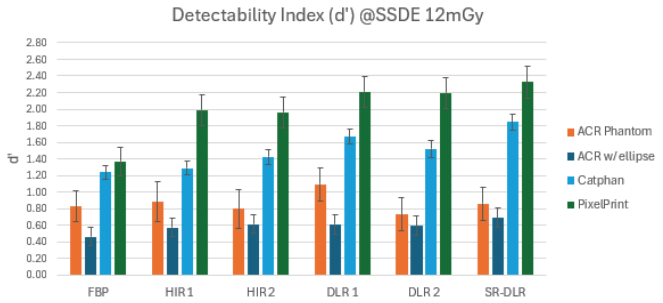


Fig. 2. Detectability index (d′) measured at a SSDE of 12 mGy for the ACR phantom, ACR phantom with attenuation ellipse, Catphan, and PixelPrint liver phantom across different reconstruction algorithms, including FBP, HIR1, HIR2, DLR1, DLR2, and SR-DLR.

TABLE 1
% Improvement in d' Relative to FBP

Recons	ACR	Catphan	ACR w/ Ellipse	PixelPrint
HIR 1	7%	4%	24%	46%
HIR 2	-3%	15%	33%	44%
DLR 1	32%	35%	32%	61%
DLR 2	-11%	23%	28%	61%
SR-DLR	4%	49%	49%	70%

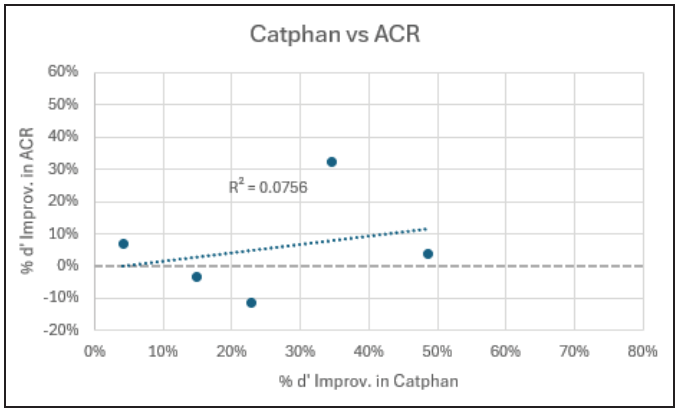


Fig. 3. Comparison of relative improvements in d′ measured using the Catphan and ACR phantoms across reconstruction algorithms. Each data point represents the % improvement in d′ relative to FBP for a given reconstruction method. The dotted line denotes a linear fit, illustrating the weak correlation between detectability gains measured with the two phantoms ($R^2 = 0.076$).

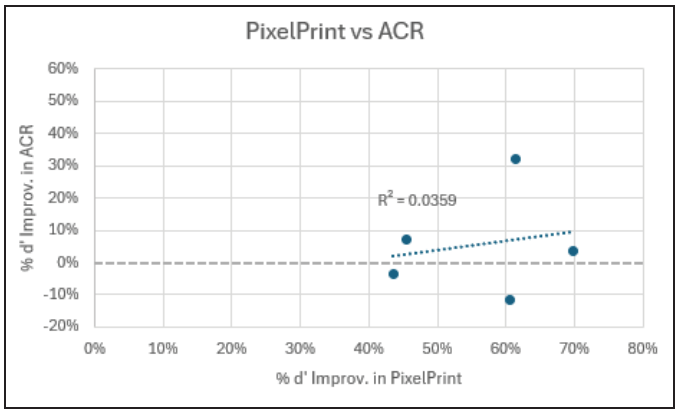


Fig. 4. Comparison of relative improvements in d′ measured using the PixelPrint and ACR phantoms across reconstruction algorithms. The linear fit illustrates weak correlation between detectability gains measured with the two phantoms ($R^2 = 0.036$).

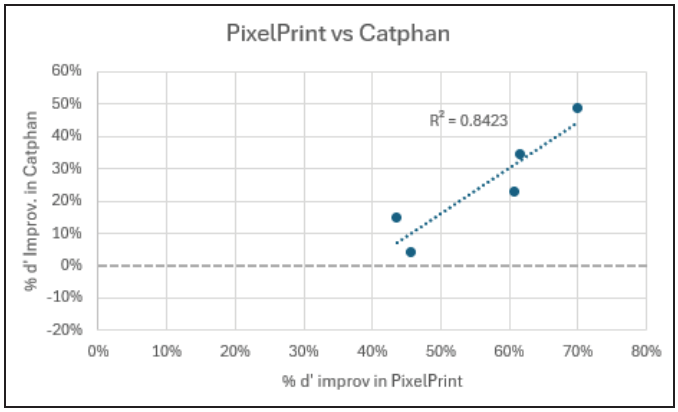


Fig. 5. Comparison of relative improvements in d′ measured using the PixelPrint and Catphan across reconstruction algorithms. The linear fit illustrates strong correlation between detectability gains measured with the two phantoms ($R^2 = 0.84$).

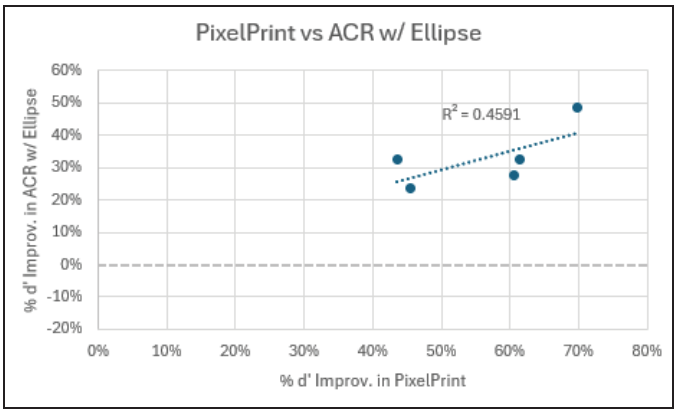


Fig. 6. Comparison of relative improvements in d′ measured using the PixelPrint and ACR phantom with the attenuation ellipse across reconstruction algorithms. The linear fit illustrates weak correlation between detectability gains measured with the two phantoms ($R^2 = 0.46$).

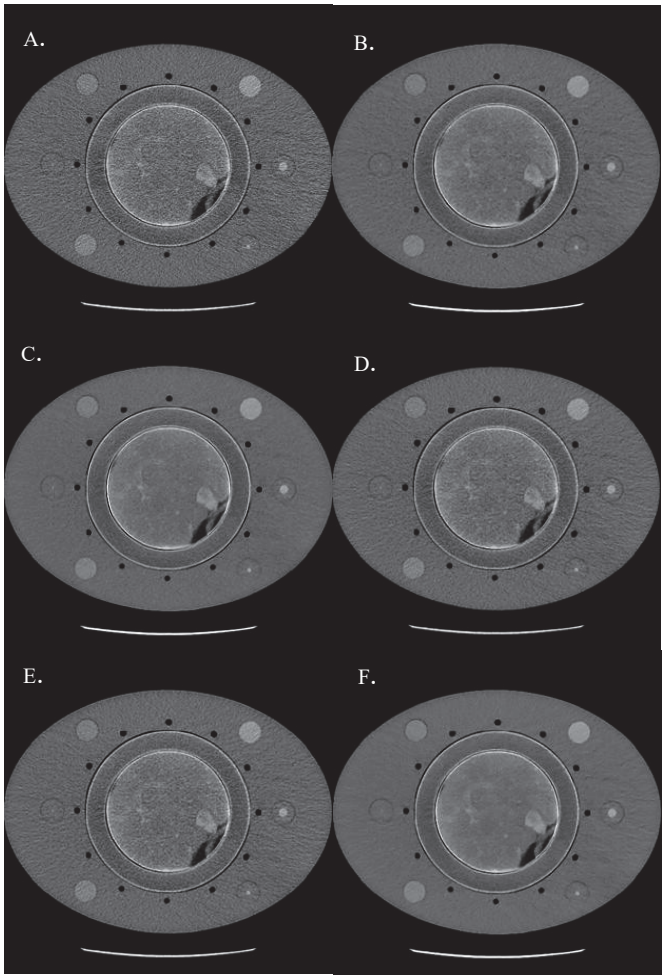


Fig. 7. Representative 3mm slice thickness PixelPrint images reconstructed using different algorithms: (A) FBP, (B) DLR 1, (C) DLR 2, (D) HIR 1, (E) HIR 2, and (F) SR-DLR (window/level = 400/40 HU).

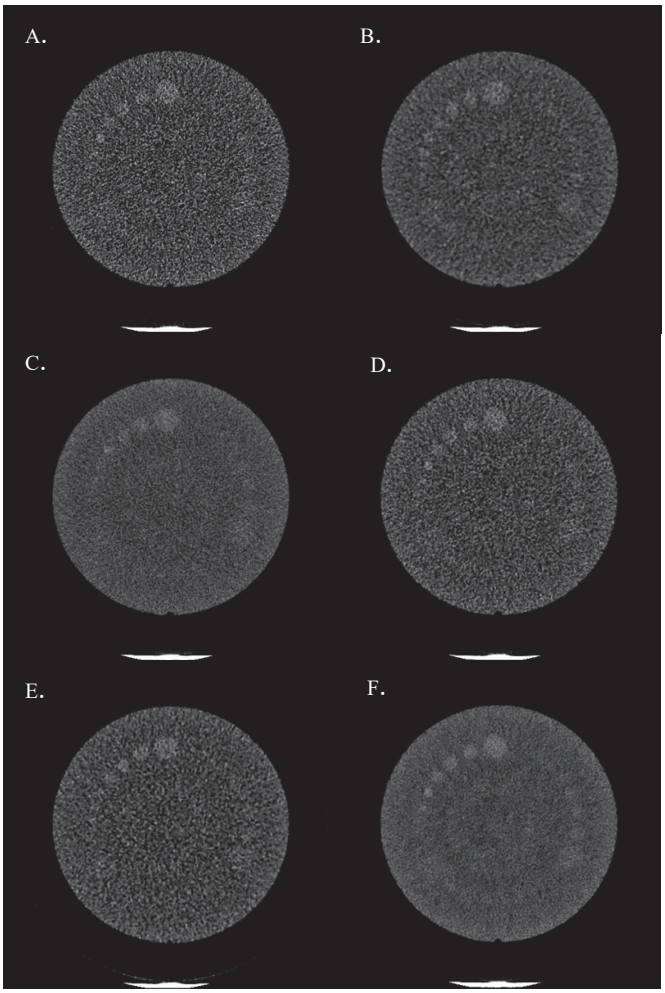


Fig. 8. Representative 3mm slice thickness Catphan images reconstructed using different algorithms: (A) FBP, (B) DLR 1, (C) DLR 2, (D) HIR 1, (E) HIR 2, and (F) SR-DLR (window/level = 100/80 HU).

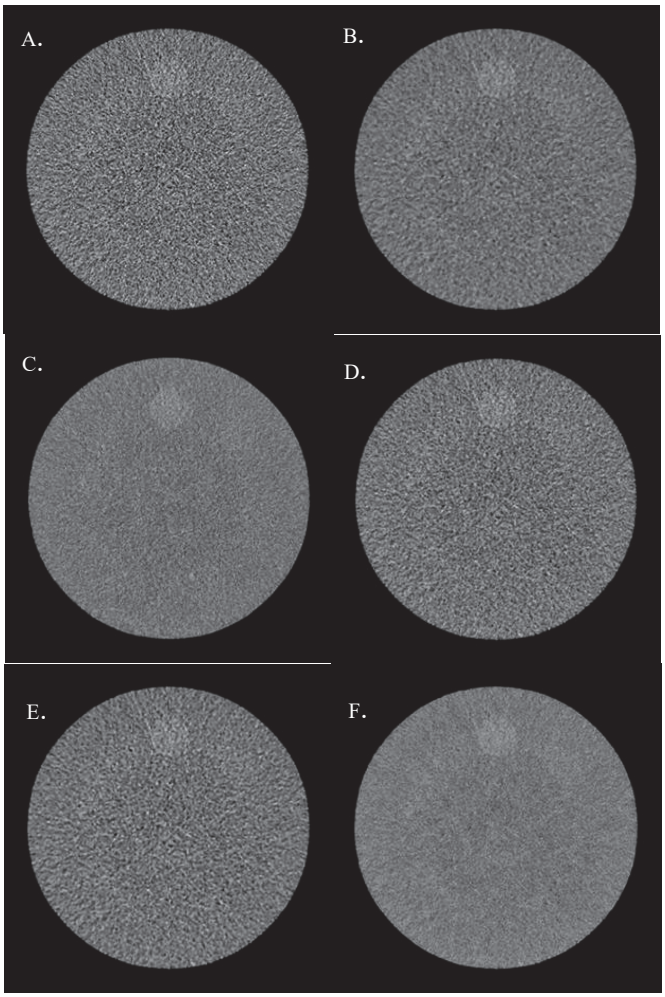
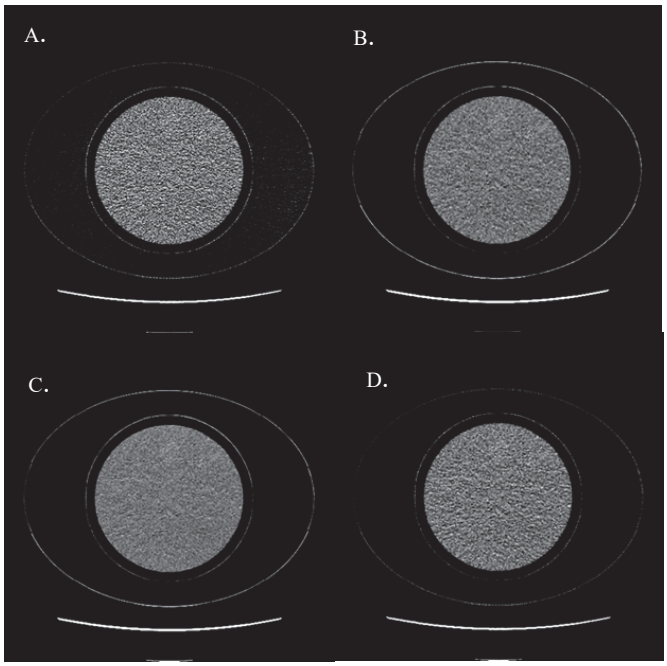


Fig. 9. Representative 3mm slice thickness Catphan images reconstructed using different algorithms: (A) FBP, (B) DLR 1, (C) DLR 2, (D) HIR 1, (E) HIR 2, and (F) SR-DLR (window/level = 100/80 HU).



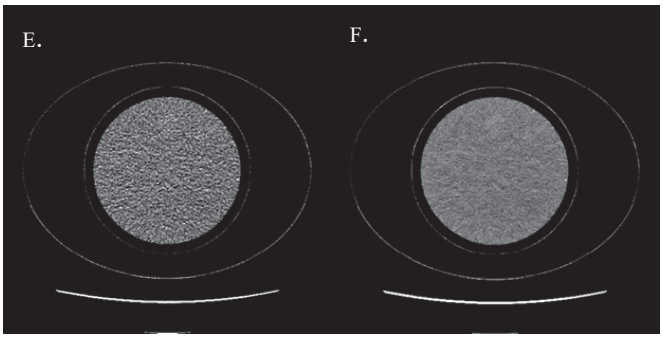


Fig. 10. Representative 3mm slice thickness images of ACR phantom with an attenuation ellipse reconstructed using different algorithms: (A) FBP, (B) DLR 1, (C) DLR 2, (D) HIR 1, (E) HIR 2, and (F) SR-DLR (window/level = 100/100 HU).

IV. DISCUSSION

This study demonstrates that phantom selection has a substantial influence on task-based assessments of low-contrast detectability in CT. When identical acquisition parameters and observer settings were used, both the magnitude of detectability gains and the relative performance ordering of reconstruction algorithms varied considerably across the ACR, Catphan, and PixelPrint phantoms. The most limited differences were observed in the ACR phantom, while the strongest separation between reconstruction methods occurred in the anatomically realistic PixelPrint phantom, with the Catphan phantom exhibiting similar behavior. These findings indicate that phantom design itself strongly modulates measured task performance.

From a task-based imaging perspective, these findings suggest that the ACR phantom may have limited sensitivity to reconstruction-dependent changes in signal background relationships that are important for low contrast lesion detection. Although the ACR phantom is well-suited for evaluating system stability and capturing major errors such as mistakenly using high resolution chest protocol for abdomen, its phantom design may reduce observer sensitivity to improvements that enhance local contrast, edge definition, or texture consistency. In contrast, the PixelPrint phantom introduces spatially varying attenuation patterns that more closely resemble clinical anatomy, allowing reconstruction dependent differences to manifest more clearly in task-based metrics. The observation that algorithm rankings were not preserved with ACR phantom underscores the risk of extrapolating conclusions derived from standardized QA phantoms to clinical detection tasks.

Several limitations should be noted. The analysis was restricted to a single lesion type, size, and dose condition. Although the model observer employed in this study offers a stable and reproducible means of comparing reconstruction performance, it represents only an approximation of clinical image interpretation. Moreover, the PixelPrint phantom used here reflects a single realization of anatomical structure and texture. Future work will extend this investigation to encompass a wider range of lesion sizes and contrasts, and multiple dose conditions with the goal of more comprehensively linking phantom realism and task-based

image quality metrics to clinically meaningful performance.

Overall, these results support the use of anatomically realistic phantoms as a complementary tool for CT image quality evaluation. As the usage of non-linear reconstruction algorithms increases, evaluation tools that are sensitive to texture-dependent performance gains will become increasingly important. PixelPrint phantoms offer a practical and reproducible means to bridge the gap between standardized QA testing and clinically relevant task-based assessment.

V. CONCLUSION

The findings of this study indicate that model-observer–based assessments of low-contrast lesion detectability are highly dependent on phantom design. Although the ACR CT accreditation phantom remains indispensable for standardized quality control, it may limit sensitivity to reconstruction-driven improvements relevant to clinical detection tasks. In contrast, anatomically realistic phantoms such as PixelPrint better capture texture-dependent differences in reconstruction performance, enabling more sensitive evaluation of detectability gains. Incorporating such phantoms into task-based assessment frameworks may enhance the clinical relevance of image quality evaluation and support the advancement of clinical protocol optimization.

REFERENCES

[1] M. A. Hobson, E. T. Soisson, S. D. Davis, and W. Parker, "Using the ACR CT accreditation phantom for routine image quality assurance on both CT and CBCT imaging systems in a radiotherapy environment," *J. Appl. Clin. Med. Phys.*, vol. 15, no. 4, pp. 226–239, Jul. 2014, doi: https://doi.org/10.1120/jacmp.v15i4.4835.

[2] C. H. McCollough *et al.*, "The phantom portion of the American College of Radiology (ACR) Computed Tomography (CT) accreditation program: Practical tips, artifact examples, and pitfalls to avoid," *Med. Phys.*, vol. 31, no. 9, pp. 2423–2442, Sep. 2004, doi: https://doi.org/10.1118/1.1769632.

[3] R. J. Cropp, P. Seslija, D. Tso, and Y. Thakur, "Scanner and kVp dependence of measured CT numbers in the ACR CT phantom," *J. Appl. Clin. Med. Phys.*, vol. 14, no. 6, pp. 338–349, Nov. 2013, doi: https://doi.org/10.1120/jacmp.v14i6.4417.

[4] I. Hernandez-Giron, J. Geleijns, A. Calzado, and W. J. H. Veldkamp, "Automated assessment of low contrast sensitivity for CT systems using a model observer," *Med. Phys.*, vol. 38, no. S1, pp. S25–S35, Jul. 2011, doi: https://doi.org/10.1118/1.3577757.

[5] R. W. Bouwman, R. E. van Engen, D. R. Dance, K. C. Young, and W. J. H. Veldkamp, "Evaluation of Human Contrast Sensitivity Functions Used in the Nonprewhitening Model Observer with Eye Filter BT - Breast Imaging," H. Fujita, T. Hara, and C. Muramatsu, Eds., Cham: Springer International Publishing, 2014, pp. 715–722.

[6] U.S. Food and Drug Administration, 510(k) Premarket Notification K192832: Aquilion Prime SP (TSX-303B8) V10.2 with AiCE-i. Canon Medical Systems Corporation;cleared February 21, 2020.

Risk-Minimizing Tube Current Modulation Compared to Clinical Protocols

Philipp Oelze, Michael Knaup, Daniel Mosig, Aniol Serra Juhé, Andreas Maier, Michael Lell, and Marc Kachelrieß

***Abstract*—To reduce radiation dose in CT, tube current modulation (TCM) is used. Current clinical protocols aim at minimizing the tube output in terms of the total mAs value of the scan, given a certain image quality. Risk-minimizing patient specific tube current modulation (riskTCM) techniques have been proposed recently, which aim to reduce the effective dose that the patient receives, given a certain image quality. We show that we can apply riskTCM tube current curves to clinical data, and use exactly the same reconstruction protocol as the clinical data. This allows us to directly compare clinical scans to riskTCM simulated scans. We also show that the potential dose reduction is in the same order of previous work.**

I. Introduction

AUTOMATIC exposure control (AEC) was an important development in computed tomography (CT), as it plays an important part in CT radiation dose reduction [1, 2]. Common AEC techniques are automatic tube voltage selection and tube current modulation (TCM). Current clinical TCM are based on minimizing the tube current time product (mAs product) for a certain targeted noise level. We refer to this method as mAsTCM. There are also organ-specific TCM available, which aim at reducing the dose to a certain organ, by lowering the tube current anterior to this organ [3]. Recently, Klein et al. proposed a new approach, riskTCM, that aims at minimizing the patient's effective dose [4]. Simulations and an in-house reconstruction software had been used to demonstrate and evaluate riskTCM. To enable assessment by a radiologist, the riskTCM raw data need, however, to be reconstructed with the same parameters and the same reconstruction algorithm as used in clinical routine.

To do so, we obtained access to ReconCT, a CT reconstruction software from Siemens Healthineers that provides images equivalent to those reconstructed at Siemens CT scanners. ReconCT allows us to add noise to the raw data and thereby simulate a different tube current or tube current curve. We used ReconCT to reconstruct CT volumes as if they were scanned on a CT scanner using our riskTCM curves. We compared them to the clinical scan with mAsTCM (aka CARE Dose4D at Siemens scanners) at the same dose level.

Additionally, we improved riskTCM's optimization algorithm, which seeks for the optimal tube current curve given the image quality and effective dose constraints. We did so by switching to a gradient-based approach that now replaces the Simplex optimization algorithm used in Klein et al. [4]. This results in shorter optimization times and allows for much more tube current sample points.

P. Oelze, Dr. M. Knaup and Prof. Dr. M. Kachelrieß are with X-Ray Imaging and Computed Tomography, German Cancer Research Center (DKFZ), Heidelberg, Germany.

P. Oelze and Prof. Dr. M. Kachelrieß are with the Ruprecht-Karls-University, Heidelberg, Germany.

D. Mosig, A. S. Juhé and Prof. Dr.-Ing. A. Maier are with the Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

Prof. Dr. M. Lell is with the Klinikum Nürnberg, Nürnberg, Germany.

Corresponding author: philipp.oelze@dkfz.de

II. Methods

In this work we want to evaluate simulated riskTCM scans to clinical scans with added noise. We have clinical raw data and reconstructions available. Form those we simulated patient dose with a Monte Carlo Simulation. For future implementations in clinical practice, one would need calculate the dose distribution form topograms, or 3D images [5].

A. Risk-Minimizing TCM

For CT imaging, the variance of the monochromatic projection value p is proportional to

$$\sigma^2(p) \propto (P'(q))^2 \frac{e^{Q(p)}}{I} = \frac{e(p)}{I},$$

be, where $e(p) = (P'(q))^2 \exp(Q(p))$ and the tube current I . Hereby are $P(q)$ the water precorrection function that converts a measured attenuation value q into a monochromatic line integral p. Further, let Q(p) be the inverse function that undoes this water precorrection, here the filtered back projection. With that, the noise follows $\sigma \propto \sqrt{I}^{-1}$.

Recently, riskTCM, a TCM algorithm that minimizes the patient's effective dose while keeping the image quality constant, was proposed [4]. The authors showed that using the weighted average of two complementary rays with $\sigma^2(p(\alpha_i, \beta_i)) = e(p(\alpha_i, \beta_i))I^{-1}(\alpha_i)$ and, $p(\alpha_\mathrm{D}, \beta_\mathrm{D}) \approx p(\alpha_\mathrm{C}, \beta_\mathrm{C})$ the minimum variance is proportional to

$$\sigma^2_{\min}(p_{\mathrm{D,C}}) \propto \frac{1}{I(\alpha_\mathrm{D}) + I(\alpha_\mathrm{C})}.$$

The overall noise is the weighted sum over the square–root of all variances that are backprojected into the voxels:

$$N_{\text{riskTCM}}(I) = \int d^3r\, w(\boldsymbol{r}) \sqrt{\int_0^\pi d\vartheta\, \frac{e\,(p(\alpha(\vartheta,\boldsymbol{r}),\beta(\vartheta,\boldsymbol{r})))}{I(\alpha_{\text{D}}) + I(\alpha_{\text{C}})}}, \tag{1}$$

where $w(\boldsymbol{r})$ masks all pixels inside the patient, and $I(\alpha_D)$, $I(\alpha_C)$ being the direct and complementary ray. Thus, the aim of the riskTCM optimization is to find the tube current modulation that generates the minimal noise under the condition that the total risk remains constant:

$$I = \arg\min_{I(\alpha)} N_{\text{riskTCM}}(I), \tag{2}$$

while $\int d\alpha\, D_{\text{eff}}(\alpha) I(\alpha) = \text{const.}$

B. Dose Estimation

We obtained the dose distribution $D(\alpha, \boldsymbol{r})$ by our on in house Monte Carlo simulation [6]. We calculated dose distributions for circular scans with 36 tube positions α and multiple z positions with a spacing of $\Delta z = 10$ mm. From those dose distributions, the organ doses and effective dose curve $D_{\text{eff}}(\alpha, z)$ with the tissue weighting factors as proposed by the ICRP were calculated. We use the semiempirical spectra of reference [7] for our simulations. An additional 6 mm aluminum filter is applied to account for intrinsic prefiltration. Any additional pre- and bowtie filters have been set according to the scan protocol and CT system.

C. Segmentation

To calculate the organ doses, the organ positions must first be determined from the CT volume. To segment the organs from these volumes, we used the automatic segmentation tool TotalSegmentator [8]. This tool is based on deep learning and was developed using methods introduced by Isensee et al. [9]. TotalSegmentator can segment 104 different anatomical structures. Not all of them are needed explicitly, so some are grouped (i.e. kidney, small bowel, muscles and soft tissue) as remainder since they are not specified by the ICRP. As of now, TotalSegmentator does not segment the skin. In order to take the skin into account in the work, the segmentation of the skin is obtained by morphologically eroding the body volume by 1 to 2 mm. From this we then create a segmentation mask, for optimization and the organ segmentation for dose calculation.

D. Optimization

For the numerical optimization of (2) we used the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [10]. To minimize $f(\boldsymbol{x})$, this quasi-Newton method approximates the inverse Hessian matrix by storing the last m positions $\boldsymbol{x}$ and gradients $\nabla f(\boldsymbol{x})$. Unlike the standard Broyden-Fletcher-Goldfarb-Shanno algorithm, which requires $\mathcal{O}(n^2)$ memory to store the full inverse Hessian, L-BFGS achieves a significantly reduced memory complexity of $\mathcal{O}(mn)$, where m is typically chosen between 5 and 20. We use $m = 10$. The update step is then given as

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \boldsymbol{H}^{-1}(\boldsymbol{x}_k) \cdot \boldsymbol{\nabla} f(\boldsymbol{x}_k),$$

with $\boldsymbol{x}$ being the tube current, and $f(\boldsymbol{x})$ the image noise (1). We assigned one tube current value per degree of rotation, resulting in 360 tube current sampling points. Additionally, the tube current is constrained to be non-negative, which introduces an inequality constraint into the optimization problem. We used the NLOpt library for C++ for the optimization, which allows us to address the inequality constraint as box constraints. [11, 12]

E. ReconCT

For image reconstruction, the Siemens Software ReconCT is used. ReconCT enables the reconstruction as if the images were reconstructed by a Siemens CT scanner. This allows us to have the images rated by radiologists in the future instead of relying solely on image noise and dose as metrics. Also, ReconCT can add noise during reconstruction to obtain lower dose images, than the original scan. Our version of ReconCT can further be provided with a tube current curve to simulate TCM. In our case we use the riskTCM curves. Therefore, scans from a real CT setup can be simulated and studied. For this, it is important that the riskTCM curve is lower than the original tube current $I_{\text{risk}}(\alpha) < I_{\text{scan}}(\alpha)\ \forall\ \alpha$. Further, the inverse variance weighting only works for Siemens Somatom X.Cite scanners. For other scanners, there is no noise simulation available to us. There are further limitations to the tube current profile for riskTCM. Due to the implementation of the inverse variance weighting in the software, the tube current can not be set to zero, but rather needs to have a small value (i.e. 10 to 20 mA). Additionally, we are constrained to low pitch scans because we need to have a complementary X-ray beam pair for each voxel.

We further use ReconCT to reduce the original TCM curve, which is a CARE Dose4D curve, by a factor determined to yield exactly the same effective dose as our riskTCM curve.

III. Results

Figure 1 shows the results for different regions for reconstructions with riskTCM and CARE Dose + added noise. From left to right, the columns contain the organ segmentation, the scan reconstruction with added noise, and the riskTCM reconstruction. From left to right, the columns contain the organ segmentation, the scan reconstruction with added noise, and the riskTCM reconstruction. For each slice, the circular plots around the images indicate the effective dose curve $D_{\text{eff}}(\alpha)$ for the segmentation, and the corresponding

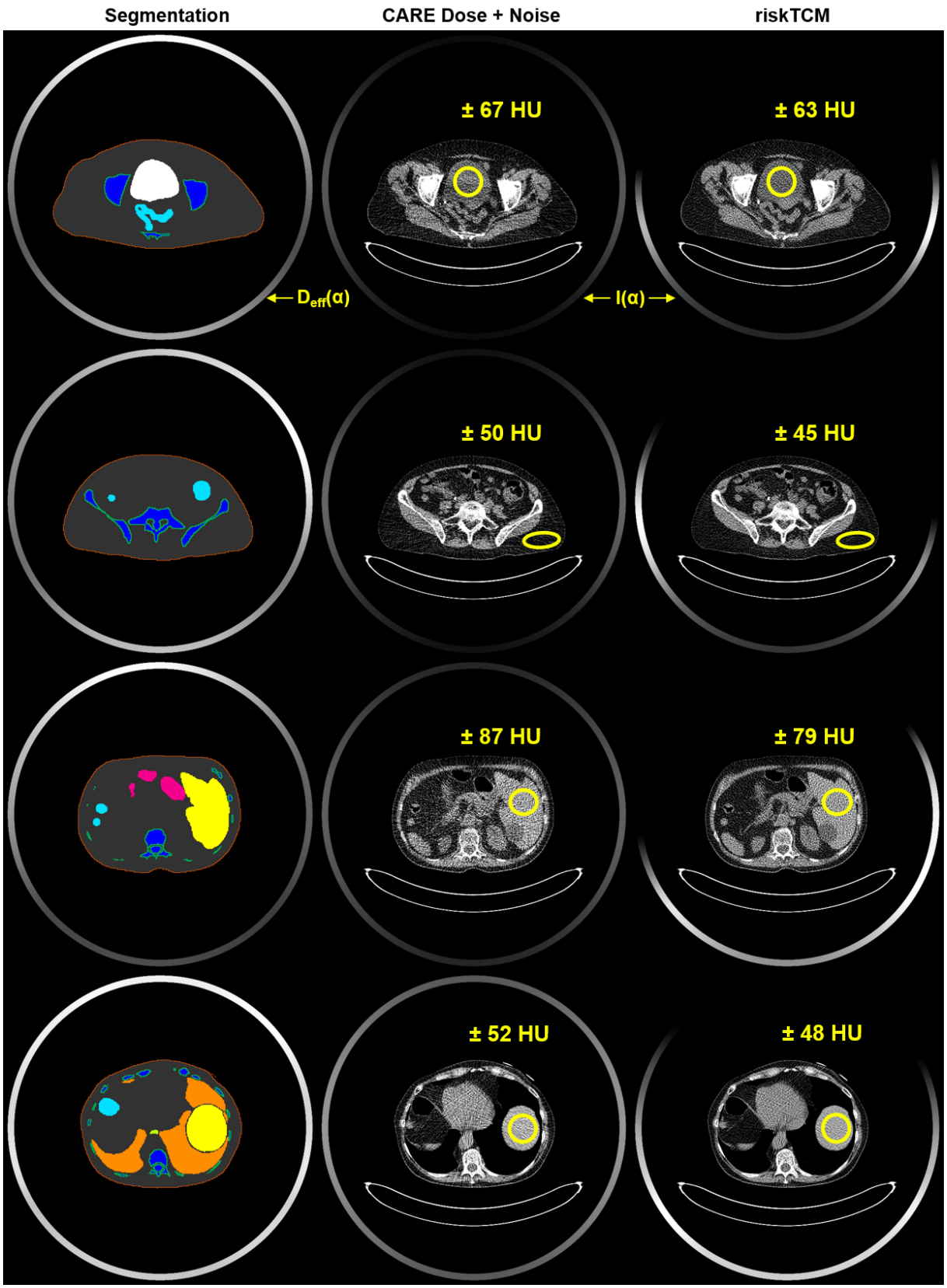


Fig. 1: images with segmented organs, CARE Dose4D image (with added noise), and riskTCM reconstruction. Both reconstructions are at the same effective dose level. The inscribed ring of the organ images represents the effective dose $D_{\text{eff}}(\alpha)$ as a function of the view angle. In the CT images, the inscribed ring's gray values represent the tube current $I(\alpha)$. CT Images are windowed to $C = 40$ HU, $W = 300$ HU.

tube current profiles used for the noise-added and riskTCM reconstructions.

Comparing the effective dose and riskTCM curves across both rows, we observe that riskTCM substantially reduces the tube current in angular regions where the effective dose is high. The scans with added noise were reconstructed using an mAsTCM protocol, which results in tube current modulation following the eccentricity of the respective body region. Both reconstructions are displayed at the same effective dose. Since, image noise and dose are correlated, we can infer equivalent dose reductions for riskTCM, as shown in Table I.

region	CARE Dose + noise	riskTCM	dose reduction
bladder	67 HU	63 HU	11.6 %
pelvis	50 HU	45 HU	19.0 %
abdomen	87 HU	79 HU	17.5 %
lower thorax	52 HU	48 HU	17.8 %

TABLE I: Measured noise in ROIs of Fig. 1. Additionally, the equivalent dose reduction for riskTCM is given.

IV. Discussion

From images (c.f. Fig. 1), we see that riskTCM results in lower image noise, when reconstructed at the same dose level as scan with added noise. The dose reduction varies depending on the region form 11 % to 19 %. Since the limitations of the riskTCM tube current during reconstruction, the tube current needed to be set relatively low, increasing image noise significantly. To reduce overall image noise then each CT volume was averaged over a z length of 5 mm (5 slices). With the ReconCT reconstruction of riskTCM, we achieve similar levels of dose reduction as previously shown by Klein et al. [4].

V. Conclusion

We showed that with ReconCT, we can reconstruct scans as if they were performed with riskTCM and can compare them to a scan with added noise at the same dose level. Ideally, future reconstructions should be done such that riskTCM and CARE dose with added noise, have the same image noise. Until now this is difficult, since the resulting noise of reconstructions with ReconCT can not be determined apriori. Going forward with this project, our goal is to create an evaluation for radiologist with more patient data.

Acknowledgment

This work is funded by the German Research Foundation (DFG) under grant number KA 1678/37, LE 2763/6, and MA 4898/27. The Monte Carlo dose estimation method RayConStruct-MC was provided by RayConStruct®GmBH, Nürnberg, Germany. The data were provided by Klinikum Nürnberg, Nürnberg, Germany. ReconCT was provided by Siemens Healthineers AG, Munich, Germany. We would like to thank Dr. Andreas Prokein and Catherine Knobloch for their support with ReconCT.

References

1. Kalender, W. A., Wolf, H. & Suess, C. Dose reduction in CT by anatomically adapted tube current modulation. II. Phantom measurements. *Med Phys* **26,** 2248–2253. ISSN: 2473-4209. (2025) (1999).
2. Gies, M., Kalender, W. A., Wolf, H., Suess, C. & Madsen, M. T. Dose reduction in CT by anatomically adapted tube current modulation. I. Simulation studies. *Med Phys* **26,** 2235–2247. ISSN: 2473-4209. (2025) (1999).
3. Mussmann, B., Mørup, S., Skov, P., Foley, S., Brenøe, A., Eldahl, F., Jørgensen, G. & Precht, H. Organ-based tube current modulation in chest CT. A comparison of three vendors. *Radiography* **27,** 1–7. ISSN: 1078-8174 (2021).
4. Klein, L., Liu, C., Steidel, J., Enzmann, L., Knaup, M., Sawall, S., Maier, A., Lell, M., Maier, J. & Kachelrieß, M. Patient–Specific Radiation Risk–Based Tube Current Modulation for Diagnostic CT. *Med. Phys.* **49,** 4391–4403 (July 2022).
5. Xu, Y., Yan, H., Ouyang, L., Wang, J., Zhou, L., Cerviño, L., Jiang, S. B. & Jia, X. A method for volumetric imaging in radiotherapy using single x-ray projection. *Medical Physics* **42,** 2498–2509 (2015).
6. Baer, M. & Kachelrieß, M. Hybrid Scatter Correction for CT Imaging. *Phys. Med. Biol.* **57,** 6849–6867 (2012).
7. Tucker, D. M., Barnes, G. T. & Chakraborty, D. P. Semiempirical Model for Generating Tungsten Target X–Ray Spectra. *Med. Phys.* **18,** 211–218 (Mar. 1991).
8. Wasserthal, J., Breit, H.-C., Meyer, M. T., Pradella, M., Hinck, D., Sauter, A. W., Heye, T., Boll, D. T., Cyriac, J., Yang, S., Bach, M. & Segeroth, M. TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. *Radiology: Artificial Intelligence* **5,** e230024. (2025) (Sept. 2023).
9. Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J. & Maier-Hein, K. H. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* **18.** ISSN: 1548-7105. (2025) (Feb. 2021).
10. Nocedal, J. Updating quasi-Newton matrices with limited storage. *Mathematics of Computation* **35,** 773–782 (1980).
11. Johnson, S. G. *The NLopt nonlinear-optimization package* https://github.com/stevengj/nlopt. 2007.
12. Liu, D. C. & Nocedal, J. On the limited memory BFGS method for large scale optimization. *Mathematical Programming* **45,** 503–528 (1989).

The Information Cliff: Testing the Physical Limits of Quantitative Imaging

John M. Hoffman

***Abstract*—Quantitative imaging holds great promise to enhance the value of CT imaging studies, but it has struggled to make the transition into broad, mainstream use. In this work, we explore the fundamental physical limits of quantitative imaging by illustrating the existence of a threshold at which signal degradation leads to the stochastic failure of a simple classification task; an "information cliff." Using simulated computed tomography datasets, we evaluate the impact of varying photon counts and object sizes on the transmission of a binary message through a CT system. Our results demonstrate a non-linear relationship where mutual information remains stable until a critical signal-to-noise ratio is reached, after which performance declines sharply. These findings suggest that quantitative imaging protocols should be optimized not just for visual quality, but for the specific information-theoretic requirements of the intended task. By defining these boundaries, this research provides insight into the interaction between dose reduction and desirable, robust algorithmic design for quantitative imaging in CT imaging systems.**

***Index Terms*—Mutual information, quantitative imaging, information loss.**

I. Introduction

As computed tomography (CT) imaging hardware has reached an exceptionally mature state, the broad and widespread translation of quantitative imaging into daily use has remained stubbornly elusive. Quantitative imaging treats imaging data as input to mathematical models that produce meaningful relationships between the visual image content and meaningful labels. Radiomics and machine learning in medical imaging data-mine large image data sets and train models that seek to accurately predict medically useful output data, such as disease classification, survival prediction, treatment response, etc. [1], [2].

The images used for training most systems include all of the black-box effects of a commercial CT imaging system. Although this approach has uncovered strong and sometimes meaningful correlations between image features and the desired output it also inherently includes unintended, yet real, correlations between parts of the imaging chain including chosen imaging protocol, dose, reconstruction kernel, etc. Although there is now broader awareness of these limitations, the ongoing intense study of harmonization in CT imaging illustrates that this challenge remains unsolved [3].

J. Hoffman is a member of Center from Computer Vision and Imaging Biomarkers (CVIB) in the department of Radiological Sciences, UCLA DGSOM, Los Angeles, CA, 90024 USA. e-mail: jmhoffman@mednet.ucla.edu or contact@johnmarianhoffman.com

Implicitly assumed in this "top-down," data-driven approach is that these abstract, meaningful biological imaging signals (e.g., texture) transit CT imaging systems with equal validity and quality, and that post-hoc methods such as feature selection, and harmonization are capable of compensating for the non-linear black-box effects and detecting signals the human visual system does not see. This is a big assumption. We do not have a general framework to distinguish which signals are real and transit accurately through CT imaging systems, versus those that may be a byproduct of the imaging systems, protocols, and populations reflected in a training dataset.

In this work, we take a "bottom-up", physics-driven approach seeking to understanding how task-specific information flows through a simple imaging system. We treat the full imaging chain, from acquisition through decision as a discrete, memory-less communication channel [4]. We apply information theoretical ideas to argue that there exist non-trivial (i.e., not "zero-dose") imaging parameter configurations where the desired information required to make an accurate decision using our quantitative imaging test may be lost or at least inconsistently available at the decision stage of the imaging chain. Even with a QI test with ground-truth on a well-understood feature (mean attenuation value), results are inherently probabilistic rather than generally robust. Imaging acquired beyond this "information cliff" should generally not be utilized or trusted for the quantitative imaging task since the required amount of physical signal is not guaranteed to be available.

There is a strong history of using information theory in imaging to understand imaging chains and their capabilities, particularly in challenging scenarios. [5] formalizes the connection between task performance and physics by demonstrating that the performance of an ideal model observer is fundamentally constrained by the statistical information content available in the data, thereby providing a rigorous framework for evaluating imaging systems through task-specific information capacity. While prior work has also applied information theory to 2D imaging systems [6], [8], [9], these prior models largely assume system linearity that does not reflect modern non-linear CT reconstruction.

We extend the mutual information approach taken by others to empirically measure the information capacity of our discrete, memory-less communication channel: a simulated parallel-beam CT imaging system, reconstruction, and quan-

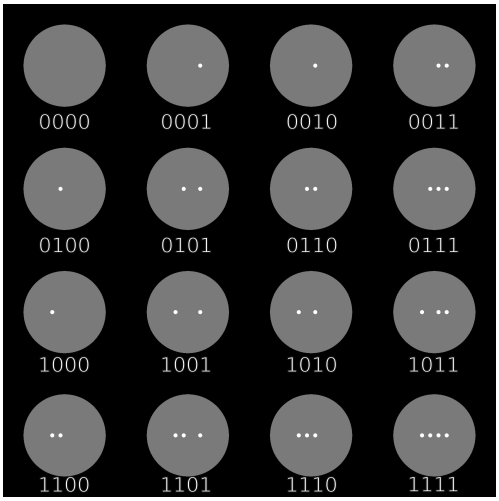


Fig. 1. 4-bit binary signals used in the experiment. 1.0 cm object size shown. Phantom body is 20.0 cm with a simulated attenuation of 0.0192 cm^2/g; bit objects simulated with attenuation 0.04 cm^2/g.

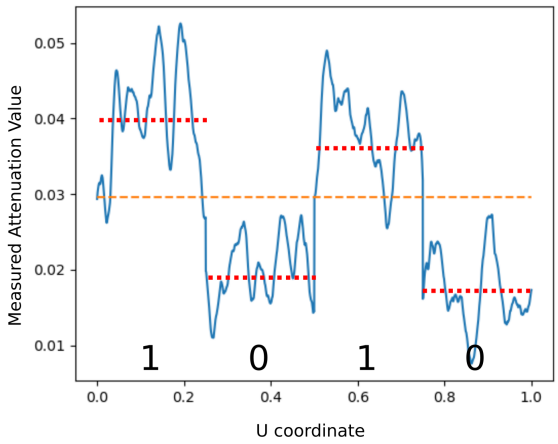


Fig. 2. Sample decoding of a reconstructed binary signal. Mean values (shown with dotted red lines) are calculated per bit from the raw reconstructed values (blue) and then classified as 1 or a 0 if occurring above or below the ground-truth threshold (dashed orange.)

titative feature extraction. We investigate the role of object size, and simulated photon noise on this information capacity and illustrate that there is a threshold where information loss begins to occurs within the channel.

For our well-defined task, these information-loss effects can be measured and quantified rigorously with mutual information. We hypothesize that for a task to be broadly robust, imaging systems and readers, human or machine, must operate in regimes where information loss does not occur. In other words: don't fall off the cliff.

II. Methods

A. Task and Interpretation

We aim to quantify how information loss occurs in our imaging system using the techniques of information theory. To accomplish this, we inject a signal with a ground truth quantity of information entropy (4 bits,) and seek to recover an equal amount of information using the output of our imaging chain. In this information "saturated" regime, the imaging process itself does not compromise the image's information content. To that end, the task was designed as the classification of a series simulated phantoms based on measurable 4-bit patterns appearing at known locations in the phantom (figure 1). The communication channel (i.e., the full input-to-decision imaging chain) is compromised through simulated photon noise and the filtered backprojection reconstruction, which is known to have long-range, correlated effects on reconstructed voxels. The inherent sampling frequency of the simulated scanner adds further challenge to the correct interpretation of the output signal, which we test via simulated object sizes. By choosing this task, one has a known ground-truth amount of information (4 bits) that should always be recovered as mutual information if operating in a "safe" imaging regime.

The simulated circular phantoms have a diameter of 20.0 cm (attenuation of 0.0192 cm^2/g). The bit objects were simulated as individual circles around the center of the phantom with attenuation 0.04 cm^2/g. Object spacing between bits was chosen to be the same as the object size. Bit object sizes were simulated at $d_{obj} = \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0 \text{ cm}\}$. Sixteen total phantoms per bit object size were simulated representing all possible input configurations.

To interpret and classify the output signal, line profiles were reconstructed through each bit object and averaged. Since object attenuation values are known exactly, a fixed threshold of 0.0296 cm^2/g was used then classify each profile as a 1 ($\mu > 0.0296$) or a 0 ($\mu \leq 0.0296$). An illustration of the interpretation of one such profile is illustrated in figure 2. This task is intentionally designed to be high-contrast, with ground-truth (known test-locations, and known object threshold), and based on a CT imaging feature known to be comparatively robust (mean reconstructed attenuation value).

B. Scan simulations

Phantoms were simulated at scanner isocenter. Scans were simulated using a parallel-beam geometry to simplify reconstruction. Geometry details such as scanner acquisition field of view (50 cm; 736 detector channels), source-isocenter (59.5 cm) and source-detector (109.5 cm) spacing, and number of projections per rotation (1152), were chosen to be similar to a standard 64-slice clinical scanner at our institution. This was chosen to establish a basic, intuitive relationship between the phantom object scale and the imaging system, and should not be interpreted as a direct reflection of a real imaging system. Quarter detector offset was simulated; finite focal spot size was not.

Noise-free sinograms were created using ray-marching at a step length of 0.001 cm. Noise addition was performed per-ray assuming $N_{photons} = \{10, 20, 30, 40, 50, 100, 500, 1000\}$. The noise-free attenuation was interpreted as the probability of a successful transmission ($p = exp(-\mu_{noise\ free})$) of a photon along the entire ray path, an assumption that essentially considers the idealized contribution of only the photoelectric effect on the resulting imaging chain. Per-ray photon counts were then generated through a basic Monte-Carlo process and the probabilities derived from the noise-free sinograms. 500 noise realizations were performed per phantom bit pattern (16 patterns), per object-size (10 object sizes), per dose level (8 dose levels).

C. Reconstruction and channel capacity calculation

From each noise realization, line profiles for each bit were directly reconstructed at a sampling distance of 0.001 cm, using parallel-beam filtered backprojection with a band-limited ramp kernel [7]. This is over-satisfies the Nyquist criterion for the simulated native detector resolution and yields $N = d_{obj}/0.001$cm "true" samples per bit; for this work we wish to exclude any effects of undersampling our reconstruction, although undersampling in the final image reconstruction is a likely limitation in many clinical CT systems that use a default 512×512 matrix for reconstruction. Classification of the phantom was then performed according to the task description above and as illustrated in figure 2.

To evaluate channel capacity of our imaging system, we calculate the mutual information according to the equation:

$$I(X;Y) = \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P(x,y)\log\left(\frac{P(x,y)}{P(x)P(y)}\right)$$

This is the joint-probability between the input bit sequences and the bit sequences recovered from the reconstruction. The maximum mutual information for our system is exactly four bits; for a given imaging condition (dose and object size), measuring less than four bits of mutual information between input and output indicates information loss. We observe this practically as bit flips in the decoded sequences relative to the input signal. Intuitively, for an imaging condition to be considered robust, approximately 4 bits of mutual information should be measured between the input and output of our communication channel; fewer than 4-bits measured implies the existence of individual noise realizations where our quantitative test fails stochastically.

III. Results

Figure 3 provides sample image reconstruction for key dose levels using the 1.0 cm object size and spacing. The low dose images combined with the ramp kernel reconstruction make for a highly noisy reconstruction. For a human reader, bit presence or absence is likely discernable starting at extremely doses. Note that this is distinct from our simulated quantitative task where only 1D line profile information is reconstructed and used to discern bit presence or absence. This reinforces the complex distinction between human readers and digital, quantitative imaging.

A. Impact of simulated dose and object size

In figure 4, it is observed that at higher simulated radiation doses, our communication channel "saturates" and incurs no information loss. In particular, this occurs around 500 simulated photons per ray for most object sizes and demonstrates the existence of imaging operating points that support the robust flow of digital information required for our task. At these operating points, the input 4-bits of information are preserved and quantitative imaging is unlikely to be impacted by any individual noise realization. The saturation point also represents the edge of the "cliff," below which noise and other black-box effects of the scanner may impact the intended quantitative evaluation.

Every CT scanner has a native sampling frequency that relates to the detector size, dt. Although reconstructions can be performed at arbitrary sampling frequencies, the highest spatial frequency recoverable under optimal condition is the Nyquist frequency, or $1/2dt \approx 7.35$ lp/cm. For the simulated scanner utilized in this work, this corresponds to an object of size of approximately 0.14 cm. This effect is illustrated by the substantial loss of information for the 0.1 cm object at all simulated doses, since it fails to satisfy the Nyquist criterion for the scanner.

B. The "cliff" of information loss

Figure 4 illustrates the complete exploration performed for this work. It is observed that information recovery for our simulated QI task is in fact, quite robust and only as dose and object size reach extremely challenging visual conditions do we begin to incur meaningful loss of information. For the simulated quantitative imaging chain, we claim that the observed plateau region above the information cliff represents the safe operating point for our quantitative task. In future work, we hope to show that all quantitative tasks have a version of this that can be measured and then prescribed for safe, robust, consistent application of quantitative imaging.

IV. Discussion

We are confident that radiomics and machine learning can function in challenging imaging environments. This work however illustrates that there exists a valid envelope or set of operating points where our tests are making resilient and robust measurements and decisions, as well as other operating points where the model is making confident guesses based on a limited amount of available information. Attempting to develop and deploy robust quantitative imaging in the latter

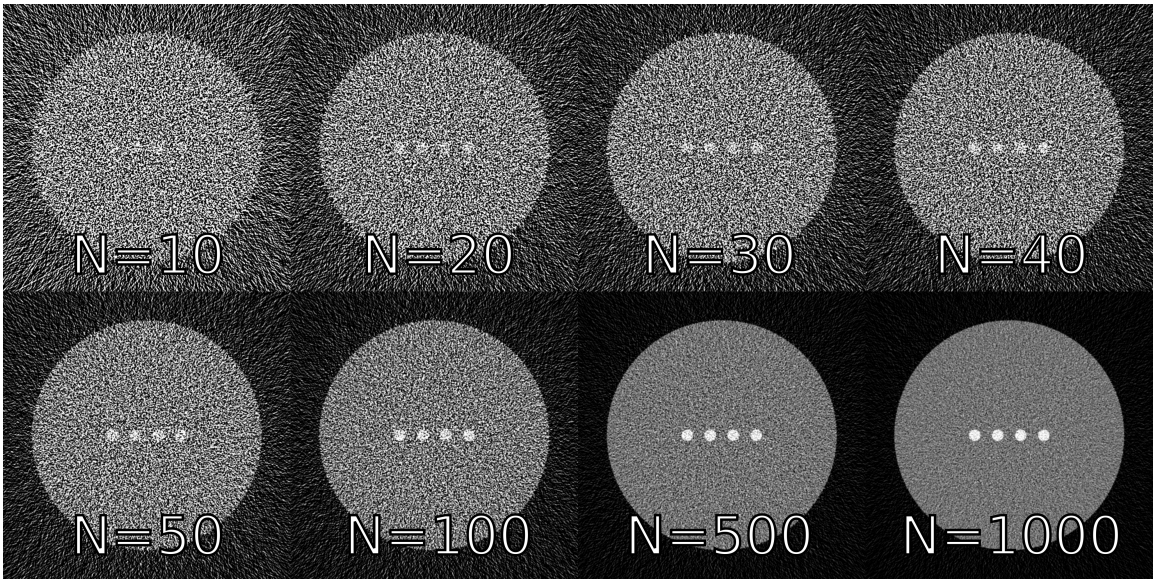


Fig. 3. Sample reconstructions of the "1111" phantom at all simulated doses. N corresponds to the simulated number of "photons" per ray in the noise realization. Only basic attenuation along each ray was simulated; complete photon physics including scatter were not simulated. The communication channel's capacity to accurately transit our required signal starts to vanish below N=500 photos.

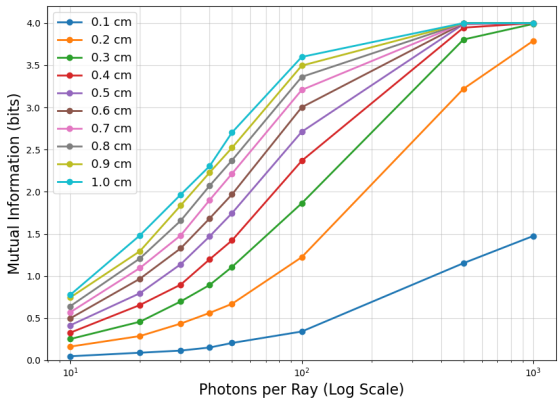


Fig. 4. Mutual information as a function of simulated dose (photons per ray) and object size. Shown on a log plot for clarity; the "cliff" of information loss occurs between 100-500 simulated photons per ray for most objects. The smallest objects fail to accurately capture the binary signal at most simulated doses. In particular, the 0.1 cm objects fail to fulfill the Nyquist criterion for the simulated scanner.

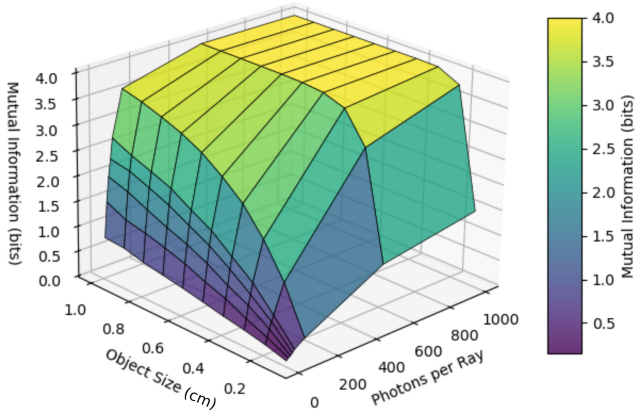


Fig. 5. An alternative presentation of the data from Fig. 4 illustrating the motivation for the "cliff" metaphor. The yellow 4-bit plateau for large object and high doses indicate a safe operating zone for our simulated quantitative test. We anticipate that this framing of quantitative imaging more broadly as a communication channel can be widened to other quantitative tests to better understand how information transits CT imaging.

scenario will always be information-limited, and many low-dose, noisy training data sets likely fall into this category for the complex tasks being developed on them.

The loss of information usable for quantitative imaging should be understood as distinct effect from poor image quality, although they are connected. There is an optimism in the modern era that computer-based observers can "read through" poor image quality; loss of information represents a paradigm shift where there is insufficient physical measurement to support the quantitative task. This is of particular concern in medical imaging where there is substantial pressure to reduce patient radiation dose, while concurrently maximizing the diagnostic utility of scan at the earliest possible discernible disease time-point (i.e., small lesions). These ideas are fundamentally in conflict as we have shown: at low doses and with small lesions we are likely to lose information which will affect even the most optimal observer.

The observations of this work agree well those presented

imaging systems that have found that ultimately, the information capacity of a linear, shift-invariant imaging system depends on the signal-to-noise ratio of the objects and the imaging quanta available for measurement. We have extended the information theoretical approach to include a mostly-linear CT image reconstruction process and empirically measured similar behavior. Importantly, the methodology we have presented requires no such linearity assumption making it suitable for testing non-linear reconstruction and diverse quantitative imaging tasks.

We have utilized a "toy" task in this work to illustrate the existence of information loss for even the simplest quantitative tasks, however we highlight that the chosen approach establishes a *best*-case scenario for quantitative imaging (known object location, known ground-truth attenuation values, non-linear scatter and detector effects excluded). Information loss in more realistic imaging scenarios is likely to occur at higher doses and larger object sizes than those tested here. Our future work will seek to bridge this theoretical experiment into more realistic imaging scenarios including utilizing realistic Monte-Carlo simulation, 3D simulation and reconstruction, the effects of reconstruction kernel and algorithm (i.e., MBIR, AI). Finally, alternative tasks will be explored, substituting our idealized task approach with a more realistic task, including localization of the signal, low-contrast signals, unknown object size, or classifiers based on radiomic feature extraction.

Finally, the preservation of mutual information down into the low dose regime bears a resemblance to human observers reading CT images. In a task-oriented study evaluating the low-dose detection of appendicitis, it was observed that human readers were able to maintain 90% of their clinical-dose performance down to doses equivalent to 20% of clinical doses [10]. If we assume our channel saturation point of 500-1000 photons per ray roughly corresponds to a clinical dose value, then we similarly do not see a meaningful drop in mutual information until 10-20% of our "clinical dose." The similarities between these two results supports the idea that quantitative imaging, like human readers, can perform robustly and reproducibly even in the dose-conscious world of medical CT imaging.

V. Conclusions

By constructing and simulating a binary-signal communication task, we have used information theory to illustrate that CT systems incur information loss that limits our ability to perform quantitative tasks using certain protocols. For this study, the number imaging quanta available for measurement were reflected through object size and photon count, and were shown to exhibit an information "cliff" for our chosen task.

In the era of modern machine learning, it can be tempting to assume that a deep enough network with sufficient training data, or enough features on which to classify, can "learn" through all noise patterns and challenging imaging scenarios to uncover features relating to the desired task, however our experiment shows that even features known to be robust (mean value of measured attenuation) incur loss, past which accurate quantitative imaging become probabilistic, rather than definitive. Although an image may look acceptable to a human reader, the required information content for a given quantitative imaging task may have already disappeared due to dose reduction or image sampling issues. Our human visual system is a poor estimator of these losses for quantitative tasks. Just because we can *calculate* a feature on an image, does not mean we can reliably *measure* a meaningful feature value.

References

[1] H. J. W. L. Aerts et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nat. Commun.*, vol. 5, no. 1, p. 4006, Jun. 2014.

[2] P. Lambin et al., "Radiomics: Extracting more information from medical images using advanced feature analysis," *Eur. J. Cancer*, vol. 48, no. 4, pp. 441–446, Mar. 2012.

[3] K. Kulkarni et al., "CT-Norm: A toolkit to characterize and harmonize variability in CT," in *Proc. SPIE 13411, Medical Imaging 2025: Imaging Informatics*, 134111C, Apr. 2025.

[4] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.

[5] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. Hoboken, NJ, USA: Wiley, 2004.

[6] T. E. Gureyev, Y. I. Nesterets, D. Thompson, S. C. Mayo, C. Hall, and K. M. Pavlov, "Spatial resolution, signal-to-noise and information capacity of linear imaging systems," *Opt. Express*, vol. 24, no. 15, pp. 17130–17142, Jul. 2016.

[7] G. L. Zeng, "Revisit of the ramp filter," *IEEE Trans. Nucl. Sci.*, vol. 62, no. 1, pp. 131–136, Feb. 2015.

[8] R. Narravula, M. M. Hayat, and B. Javidi, "Information theoretic approach for assessing image fidelity in photon-counting arrays," *Opt. Express*, vol. 18, no. 23, pp. 23833–23846, Nov. 2010.

[9] R. F. Wagner, D. G. Brown, and M. S. Pastel, "Application of information theory to the assessment of computed tomography," *Med. Phys.*, vol. 6, no. 2, pp. 83–94, Mar./Apr. 1979.

[10] D. Zhang et al., "Diagnostic accuracy at several reduced radiation dose levels for CT detection of appendicitis," in *Proc. SPIE 8313, Medical Imaging 2012: Image Perception, Observer Performance, and Technology Assessment*, 831347, Feb. 2012.

Mean Squared Error, Structure Similarity, and Channelized Hotelling Observer Study in Machine Learning

Gengsheng L. Zeng

***Abstract*—Machine learning has become the state-of-the-art in many fields including medical imaging. Using image denoising as an example, the training procedure typically involves a large number of noisy/clean image pairs. The mean squared error (MSE) between the noisy and clean images is minimized by a gradient descent algorithm such as Adaptive Moment Estimation (ADAM) algorithm. A routine pattern in a journal manuscript is as follows. Authors propose a new network architecture, the network is trained, and the results are compared in terms of peak signal-to-noise ratio (PSNR) and structure similarity (SSIM). This paper investigates whether it is suitable to train the network by directly optimizing PSNR or SSIM, instead of MSE. We claim that evaluation with PSNR or SSIM may not be proper for lesion detection tasks. We point out that PSNR is equivalent to the mean squared error method. Computer simulation results indicate the feasibility of using SSIM to train a network. The channelized Hotelling observer study shows that if lesions are not used in training the network, the network should not be used for lesion detection.**

***Index Terms*—Machine learning, training, objective function, mean squared error, structure similarity, peak signal-to-noise ratio, optimization, channelized Hotelling observer.**

I. Introduction

In the era of machine learning, neural networks find their applications everywhere, including medical imaging. For a typical setup of supervised learning, input/target pairs are used to train the network parameters. The average discrepancy between the input and the target is usually used as the loss function for training. This discrepancy is commonly measured by the L_2 norm, that is, the mean-squared-error (MSE). For an image denoising task, some other norms are also used, for example, the L_1 loss function, that is, the mean-absolute-error (MAE), which preserves edges better than L_2. The Huber loss function is another example, that has a transition between L_1 and L_2 loss functions.

Many medical imaging papers use the peak signal-to-noise ratio (PSNR) and the structure similarity (SSIM) as metrics to compare different methods [1-3].

We have made this observation: Researchers use one objective function (e.g., MSE) to train a network, but use a different metric (e.g., SSIM) to assess the results. This current paper investigates whether it is possible to use the PSNR or the SSIM as the loss function to train a neural network. This paper also questions whether it makes sense to assess the results with SSIM if the task is lesion detection.

To assess a medical imaging method for lesion detection by physicians, human observer studies are recommended but they are costly [4]. A Channelized Hotelling Observer (CHO) is a computational model observer designed to mimic how a human observer performs a detection or discrimination task on medical images [5, 6]. We will use CHO to evaluate if SSIM is a good choice in assessment of a system/algorithm. We understand that CHO may not fully capture cognitive or perceptual processes of expert radiologists (context, lesion shape reasoning, etc.).

II. Methods

A. PSNR and SSIM

Let us first consider the PSNR, which is defined as

$$PSNR = 10 \times log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20 \times log_{10}(MAX_I) - 10 \times log_{10}(MSE), \quad (1)$$

where MAX_I is the maximum possible pixel value of the true image and MSE is mean squared error between the true image $I(i,j)$ and the processed image $K(i,j)$ defined as

$$MSE = \sum_{i,j}[I(i,j) - K(i,j)]^2. \quad (2)$$

Since $\mathrm{MAX_I}$ is a constant, maximization of the PSNR is equivalent to minimization of the MSE. Therefore, there is no gain in optimizing the PSNR over optimizing the MSE in training a neural network.

The SSIM is perception-based and is used to measure the similarity between two images $X(i,j)$ and $Y(i,j)$. The local SSIM is defined as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (3)$$

where x is a patch in image X and y is a same-size, same-location patch in image Y, with

μ_x the pixel sample mean of x;

μ_y the pixel sample mean of y;

σ_x^2 the sample variance of x;

σ_y^2 the sample variance of y;

σ_{xy} the sample covariance of x and y;

c_1 and c_2 are two small constants to stabilize division.

The default value for the patch size of x and y is 11×11. A

This work was supported in part by the National Institutes of Health under Grant 2R15EB024283.

G. L. Zeng is with the Department of Computer Science, Utah Valley University, Orem, UT 84058 USA (e-mail: larry.zeng@uvu.edu) and the Department of Radiology and Imaging Sciences, University of Utah, Salt Lake City, UT 84108 USA (larry.zeng@hsc.utah.edu).

patch is a sliding window, that slides over the entire image. This means that x and y are 11×11 patches (or sliding windows) from the two images being compared. The *total* SSIM is computed over these local windows and then averaged over the whole image.

The SSIM can take values in [-1, 1], and the value 1 indicates perfect similarity. Value 0 means no structural similarity. A negative value implies reversed contrast.

We will investigate the feasibility of using

$$L = 1 - SSIM \tag{4}$$

as the loss function to train a neural network. Here, the *total* SSIM is calculated between the input image X and the target image Y. Obviously, L defined in (4) is not a norm of X-Y.

We now explain what the SSIM metric really means. Since the constants c_1 and c_2 in (3) are user-determined to prevent division by zero. Pushing $SSIM(x, y)$ close to 1 can be achieved by pushing the following two expressions (5) and (6) to 0. When the ratio is approximately 1, the difference between the numerator and the denominator is approximately 0. That is,

$$\mu_x^2 - 2\mu_x\mu_y + \mu_y^2 \approx 0, \tag{5}$$

$$\sigma_x^2 - 2\sigma_{xy} + \sigma_y^2 \approx 0. \tag{6}$$

Expression (5) is

$$\left(\mu_x - \mu_y\right)^2 \approx 0, \tag{7}$$

which implies

$$\mu_x \approx \mu_y \approx \mu. \tag{8}$$

Therefore, SSIM encourages the mean values of corresponding patches in both images close to each other.

Expression (6) can be re-written as

$$\frac{1}{n-1}\sum_{i=1}^{n}\left[(x_i-\mu_x)^2 - 2(x_i-\mu_x)\left(y_i-\mu_y\right) + \left(y_i-\mu_y\right)^2\right]$$

$$= \sum_{i=1}^{n}\frac{(x_i-\mu_x)^2 - 2(x_i-\mu_x)\left(y_i-\mu_y\right) + \left(y_i-\mu_y\right)^2}{n-1} \approx 0. \tag{9}$$

In the situation where (8) is approximately satisfied, (9) can be approximated as

$$\sum_{i=1}^{n}\frac{(x_i-\mu)^2 - 2(x_i-\mu)(y_i-\mu) + (y_i-\mu)^2}{n-1}$$

$$= \sum_{i=1}^{n}\frac{x_i^2 - 2x_iy_i + y_i^2}{n-1}$$

$$= \sum_{i=1}^{n}\frac{(x_i-y_i)^2}{n-1} \approx 0\,. \tag{10}$$

Therefore, (5) and (6) imply

$$\sum_{i=1}^{n}(x_i-y_i)^2 \approx 0, \tag{11}$$

which is the mean squared error (MSE). Clearly, pixel-wise condition (11) is much stronger than patch-wise conditions (5) and (6).

The above discussion indicates the equivalence between SSIM and MSE when $SSIM \approx 1$ and $MSE \approx 0$. Other than in this extreme situation, SSIM optimization and MSE optimization are different.

When SSIM was introduced [1], the goal was to compare luminance (brightness), contrast, and structure between two images. For a small image patch in each image, the brightness was measured and compared by the mean value (8); the contrast was measured and compared by the standard deviation

$$\sigma_x^2 \approx \sigma_y^2 \approx \sigma^2 \quad \text{or} \quad 2\sigma_x\sigma_y \approx \sigma_x^2 + \sigma_y^2 \tag{12}$$

The structure similarity of the two patches was represented by the correlation of the standard deviation between the two patches, that is,

$$\sigma_{xy} \approx \sigma_x\sigma_y. \tag{13}$$

Combing (12) an (13), we have

$$2\sigma_{xy} \approx \sigma_x^2 + \sigma_y^2, \tag{14}$$

which is exactly the same as (6).

B. Training with Loss Function of (1- *SSIM*)

We used a small U-net to denoise computer simulated 64 × 64 images. The Images consisted of random ellipses. There were no lesions in the images. Zero-mean Gaussian noise with random variance was added. One thousand images were generated, with 900 pairs for training and 100 pairs for testing.

The small arbitrarily designed U-net had 13 layers as follows in tensorflow notation.

```
from tensorflow.keras.layers import Conv2DTranspose, Conv2D, Input, InputLayer
model = keras.Sequential([
 layers.Conv2D(15, 3, strides=2, activation="relu", padding="same"),
 layers.Conv2D(15, 3, activation="relu", padding="same"),
 layers.Conv2D(20, 3, strides=2, activation="relu", padding="same"),
 layers.Conv2D(20, 3, activation="relu", padding="same"),
 layers.Conv2D(25, 3, strides=2, padding="same", activation="relu"),
 layers.Conv2D(25, 3, activation="relu", padding="same"),
 layers.Conv2DTranspose(25, 3, activation="relu", padding="same"),
 layers.Conv2DTranspose(25, 3, activation="relu", padding="same", strides=2),
 layers.Conv2DTranspose(20, 3, activation="relu", padding="same"),
 layers.Conv2DTranspose(20, 3, activation="relu", padding="same", strides=2),
 layers.Conv2DTranspose(15, 3, activation="relu", padding="same"),
 layers.Conv2DTranspose(15, 3, activation="relu", padding="same", strides=2),
 layers.Conv2D(1, 3, activation="relu", padding="same")
])
```

The loss function of this U-net was defined in (4). The number of epochs was 4000. The optimizer was Adaptive Moment Estimation (ADAM), and the learning rate was 0.0004.

C. Channelized Hotelling Observer

The Channelized Hotelling Observer (CHO) is a computational surrogate for human detection performance, for optimizing and benchmarking medical imaging systems for lesion detection.

Let the mean vector from the lesion-absent images be $\boldsymbol{f_0}$ and the mean vector from the lesion-present images be $\boldsymbol{f_1}$. Let the covariance matrix of $\boldsymbol{f_0}$ be $R_0 = cov(\boldsymbol{f_0})$, and the covariance matrix of $\boldsymbol{f_1}$ be $R_1 = cov(\boldsymbol{f_1})$. The average covariance matrix is calculated as

$$R = \frac{R_0 + R_1}{2}. \tag{15}$$

An observer template is a column vector defined as

$$\boldsymbol{w} = R^{-1}(\boldsymbol{f_1} - \boldsymbol{f_0}). \tag{16}$$

For an arbitrary image, we can compute its feature vector $\boldsymbol{f}$ and then the decision variable (a scalar) as

$$t = \boldsymbol{w}^T\boldsymbol{f}. \tag{17}$$

A larger t tends to indicate that this is a lesion-present image. This decision variable t can be used to find the Receiver Operating Characteristic (ROC) curve and compute the area under the ROC curve (AUC). An AUC of 1.0 indicates a perfect model. An AUC of 0.5 indicates performance no better than random chance.

To simulate a decision-making process, the CHO chooses a threshold,τ. If $t > \tau$, the decision is lesion present; otherwise, lesion absent.

By sweeping the threshold, τ, across a range of possible values, we can calculate the *sensitivity* (i.e., the *true positive* rate, TPR) and the 1-*specificity* (i.e., the *false positive* rate, FPR). Each pair (FPR(τ), TPR(τ)) corresponds to one point on the ROC curve.

The CHO in our implementation had 8 bandpass filters; the passing bands were uniformly divided from 0.1 to 0.5 in normalized frequency. A dot-product was evaluated between each image and a spatial-domain filter kernel. Thus, each image became 8 scalar values, forming a column feature vector of 8 elements.

In our CHO study, independently generated 1000 lesion-free images and 1000 images with a small lesion were used. A small lesion was a circular point source at the location (44, 44) with a radius of 1.92 pixels.

III. Results

After our small U-net was trained with the loss function 1 – *SSIM* (4), the network was tested with noisy images that the network was not trained on. Some denoising results are shown in Fig. 1. Therefore, we have demonstrated that it is practical and feasible to use 1 – *SSIM* as a loss function to train a denoising network.

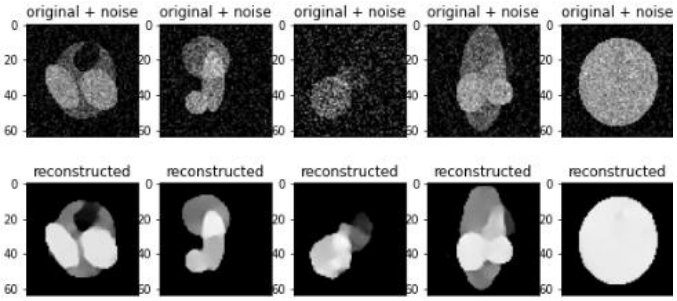


Fig. 1. Top: Noisy images. Bottom: Denoised images with the network trained by the loss function 1-*SSIM*.

During lesion detection evaluation, equal numbers of images with and without lesions were simulated. Zero-mean Gaussian noise of different variances were added to the images. Figs 2 to 4 show some examples of images with lesions at different noise levels. Fig. 2 shows the images with a high noise level with standard deviation of 0.25. Fig. 3 shows the images with a noise level with standard deviation of 0.15. Fig. 4 shows the images with a noise level with standard deviation of 0.1.

The noisy images (with and without lesions) were processed by the denoising network that was trained earlier. Some examples are shown in Fig. 5 for the noise standard deviation of 0.25.

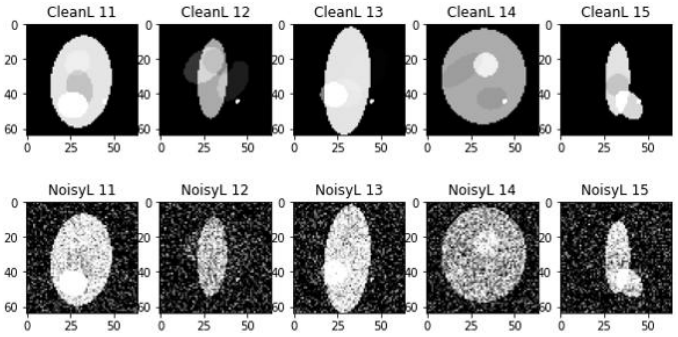


Fig. 2 Top: noise-less images. Bottom: corresponding noisy images with noise standard deviation of 0.25.

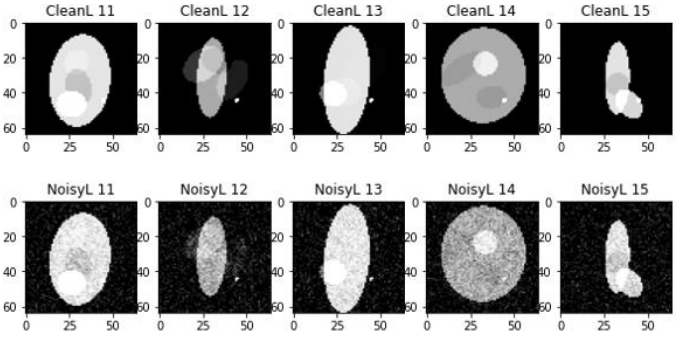


Fig. 3. Fig. 2 Top: noise-less images. Bottom: corresponding noisy images with noise standard deviation of 0.15.

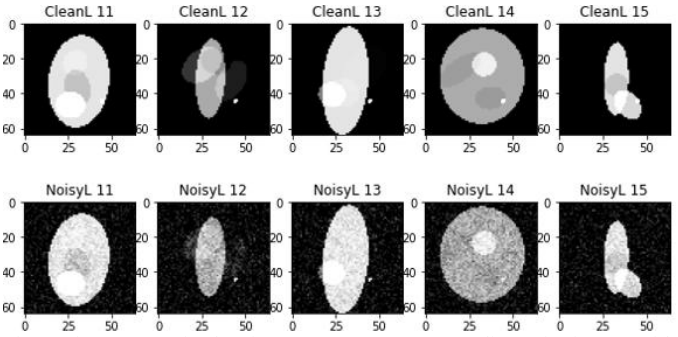
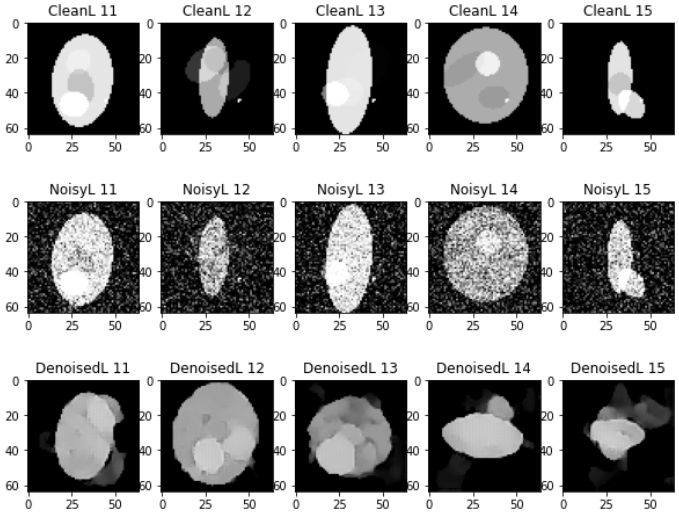


Fig. 4. Fig. 2 Top: noise-less images. Bottom: corresponding noisy images with noise standard deviation of 0.1.



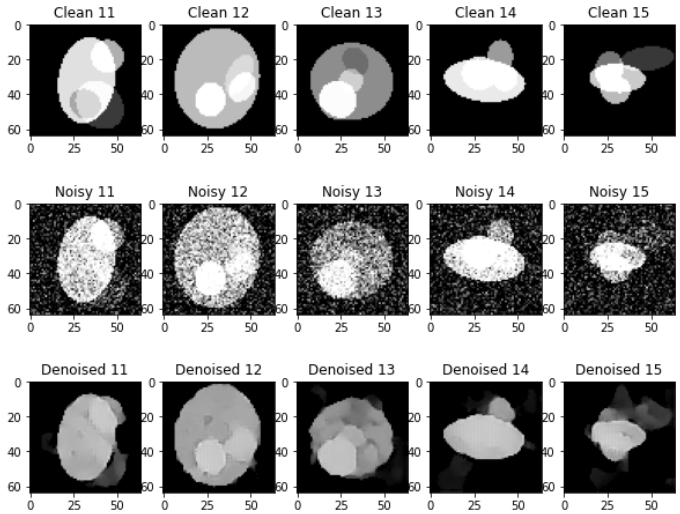


Fig. 5. Some results for noise standard deviation of 0.25. Row 1: Noiseless images with lesion; Row 2: Noisy images with lesion; Row 3: Denoised images with lesion; Row 4: Noiseless images without lesion; Row 5: Noisy images without lesion; Row 6: Denoised images without lesion.

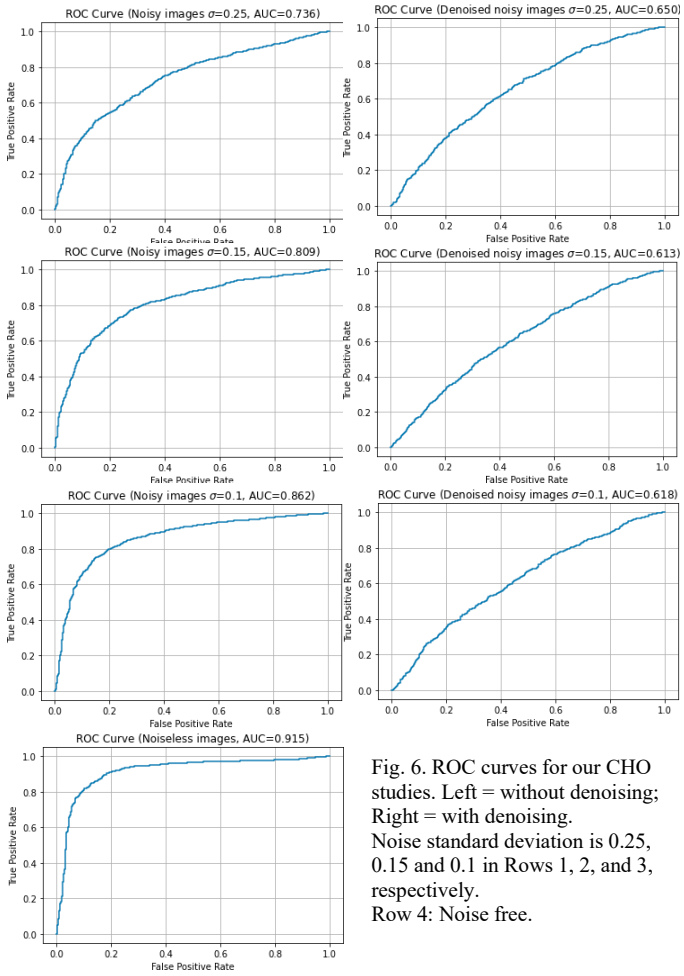


Fig. 6. ROC curves for our CHO studies. Left = without denoising; Right = with denoising.
Noise standard deviation is 0.25, 0.15 and 0.1 in Rows 1, 2, and 3, respectively.
Row 4: Noise free.

Some Channel Hotelling Observer (CHO) study curves and associated AUCs are shown in Fig. 6 and Table 1, respectively. PSNR and SSIM results are also calculated and listed in Table 2. Table 2 implies that better PSNR and SSIM results do not always provide better lesion detection.

Table 1. Area under the ROC curve

Noise standard deviation	Without denoising	With network denoising
0.25	0.736	0.650
0.15	0.809	0.613
0.1	0.862	0.618
0	0.915	NA

Table 2. PSNR and SSIM

Noise standard deviation	Without denoising PSNR/SSIM	With network denoising PSNR/SSM
0.25	10.94/0.44	14.92/0.75
0.15	15.21/0.53	15.22/0.89
0.1	18.67/0.59	15.63/0.91
0	∞/1	NA

IV. Discussion and Conclusions

Maximizing the PSNR is almost the same as minimizing the MSE. Using the MSE as the loss function during network training will take care of the PSNR requirement.

If the images are eventually assessed by SSIM, as shown in this paper, it is feasible to train the network with a loss function of (1 – *SSIM*). However, we cannot conclude that a network trained by the (1 – *SSIM*) loss function will have better SSIM performance than the same network trained by the MSE loss function, because (1-*SSIM*) is defined locally and may not have the nice properties that MSE has (e.g., the MSE is a convex function and may lead to better convergence).

If the medical imaging task is lesion detection, the assessment with the SSIM metrics may be misleading. If a general network is trained for denoising purposes, the network may have excellent SSIM values but poor lesion detection performance.

Human observer studies are believed to be most reliable to predict physician's ability to detect lesions. However, human observer studies are too costly, while CHO is a cheaper substitute. This paper shows an extreme example, in which the denoising network is not trained with lesion detection in mind. The denoised images have much worse AUCs than the original noisy images.

References

[1] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," in *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004, doi: 10.1109/TIP.2003.819861.

[2] Rehman, Abdul, Mohammad Rostami, Zhou Wang, Dominique Brunet, and Edward R. Vrscay. "SSIM-inspired image restoration using sparse representation." *EURASIP Journal on Advances in Signal Processing* 2012, no. 1 (2012): 16.

[3] Zhao, Hang, Orazio Gallo, Iuri Frosio, and Jan Kautz. "Loss functions for image restoration with neural networks." *IEEE Transactions on computational imaging* 3, no. 1 (2016): 47-57.

[4] Metz, Charles E. "Some practical issues of experimental design and data analysis in radiological ROC studies." *Investigative Radiology* 24, no. 3 (1989): 234-245.

[5] Yao, Jie, and Harrison H. Barrett. "Predicting human performance by a channelized Hotelling observer model." In *Mathematical Methods in Medical Imaging*, vol. 1768, pp. 161-168. SPIE, 1992.

[6] Zeng, Gengsheng L. and Grant T. Gullberg. "A channelized-hotelling-trace collimator design method based on reconstruction rather than projections." IEEE Transactions on Nuclear Science 49, no. 5 (2002): 2155-2158.

Task-Based Modeling of Pulmonary Nodule Detectability in CT

Elsa B. Pimenta, Gisell R. Boiset, Luuk J. Oostveen, Ioannis Sechopoulos, Alessandra Tomal, Paulo R. Costa

***Abstract* — Pulmonary nodule detectability in CT imaging can be evaluated through the detectability index, d′, linking physical image properties to observer performance. In this study, the detectability of pulmonary nodules was evaluated using a hybrid anthropomorphic lung phantom containing a synthetic ground-glass opacity nodule. Phantom images were acquired with a wide-area detector CT scanner at a fixed low dose ($CTDI_{vol}$ = 1.4 mGy), and multiple reconstruction strategies were investigated, including hybrid iterative reconstruction (HIR), model-based iterative reconstruction (MBIR), and a deep learning–based algorithm. Various reconstruction kernels were applied, and spatial resolution and noise characteristics were assessed using the Task Transfer Function (TTF) and Noise Power Spectrum (NPS), respectively. A log-linear regression model was developed to predict detectability index (d′) as a function of these image quality metrics and reconstruction parameters. The model demonstrated strong explanatory and predictive performance (adjusted R^2 = 0.80; test set R^2 = 0.79), revealing that higher average noise frequency and advanced reconstruction methods significantly improve nodule detectability, while higher spatial resolution alone may reduce detectability in low-contrast tasks. These findings highlight the interplay between noise texture, spatial resolution, and reconstruction strategy in determining CT detectability and underscore the utility of task-based image quality metrics for protocol optimization.**

***Index Terms* — Computed tomography, Pulmonary nodule, Phantom study, Task-based image quality, Detectability index**

I. Introduction

The detection of pulmonary nodules in chest computed tomography (CT) is a critical component of lung cancer screening and diagnosis. Detection performance is influenced by a combination of patient-related factors, lesion characteristics, and image acquisition and reconstruction parameters. Previous studies have demonstrated that nodule size, density, and morphology strongly affect detectability, while acquisition parameters such as slice thickness, tube current, tube potential, and reconstruction kernel modulate spatial resolution, image noise, and contrast, ultimately impacting observer performance [1, 2, 3].

Numerous investigations have shown that thinner slice thickness improves detection sensitivity and volume accuracy by reducing partial volume effects, particularly for small solid and subsolid nodules [3, 4]. Reconstruction kernel selection represents a trade-off between spatial resolution and noise, with sharper kernels enhancing edge definition for solid nodules but degrading detectability of ground glass opacities due to increased noise, while smoother kernels improve contrast to noise ratio at the expense of spatial resolution [5, 6, 7]. Tube current reduction leads to increased noise and reduced detectability unless advanced reconstruction techniques are applied [8].

The introduction of iterative reconstruction and, more recently, deep learning based reconstruction has substantially altered the relationship between dose, noise, and image texture. Compared with filtered back projection, iterative reconstruction improves detectability at reduced dose levels by suppressing noise while maintaining spatial resolution. Deep learning based reconstruction further enhances contrast to noise ratio and modifies noise texture, resulting in improved detectability of small and subsolid nodules, particularly under low and ultra-low dose conditions [8, 9, 10].

Most studies evaluate acquisition and reconstruction parameters independently. As a result, there is a lack of quantitative models capable of predicting overall detection performance when multiple reconstruction parameters are varied simultaneously. From a task-based image quality perspective, the detectability index, d′, provides a framework for linking physical image quality metrics to observer performance and has been widely used to evaluate CT systems in detection tasks [11, 12, 13]. Predicting d′ as a function of reconstruction parameters provides a unified, task-based framework grounded in second-order image statistics for understanding how reconstruction strategies affect pulmonary nodule detectability.

The goal of this study was to develop and validate a predictive model of pulmonary nodule detectability based on the

This study was financed by São Paulo Research Foundation-FAPESP (processes 2022/11457-0, 2018/05982-0 and 2023/03945-8) and by National Council for Scientific and Technological Development - CNPq (process 141335/2021-1 and 302986/2023-5).

Paulo R. Costa, Elsa B. Pimenta, and Gisell R. Boiset are with the Institute of Physics, University of São Paulo, São Paulo, Brazil. (e-mail: pcosta@if.usp.br).

Ioannis Sechopoulos and Luuk J. Oostveen are with the Department of Medical Imaging, Radboud University Medical Center, Nijmegen, The Netherlands.

Alessandra Tomal is with the Institute of Physics Gleb Wataghin, University of Campinas (UNICAMP), Campinas, Brazil.

detectability index, quantifying the combined effects of CT reconstruction method, kernel, noise texture, and spatial resolution on detection performance.

II. Material and Methods

A. The anthropomorphic lung phantom

A hybrid anthropomorphic lung phantom (Figure 1) was employed to simulate realistic anatomy and provide a clinically relevant imaging background. The phantom features a 3D-printed bronchial tree and lung walls using a photopolymer (VeroBlue, Stratasys Inc.), and a parenchymal region filled with fractionated foam chips to mimic lung tissue densities. The background lung parenchyma exhibited (-909 ± 30) HU. This indicates a density of (0.09 ± 0.03) g/cm³ [14]. Synthetic lung nodules of varying sizes and morphologies, including both solid and GGO types, were integrated into the phantom. A synthetic GGO nodule with a volume of 142 mm³ was hand-molded using ethylene vinyl acetate (EVA), with an approximate diameter of 8 mm. This study used one of these GGO nodules, which was embedded at the right upper lobe of the 370 mm diameter phantom module. The choice of a GGO nodule reflects its clinical relevance and the difficulty of detecting and characterizing such low-contrast lesions. This setup allows for a quantitative evaluation of CT system performance in identifying subtle findings and estimating lesion volume under challenging imaging conditions.

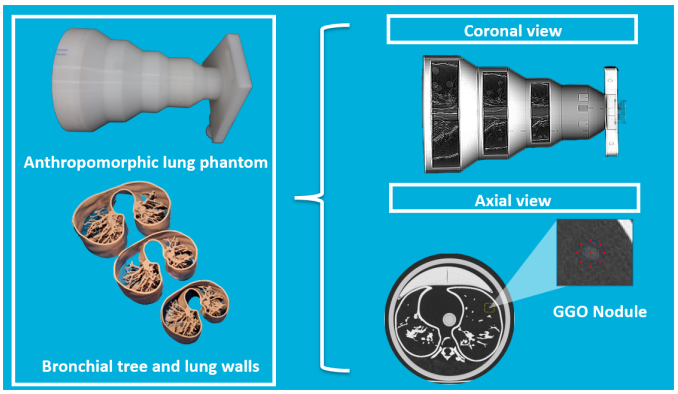


Figure 1. Hybrid phantom containing anatomically realistic lung walls and bronchial tree structures. A synthetic ground-glass opacity (GGO) nodule was embedded in the right upper lobe of the 370 mm module of the phantom. Axial and coronal CT images illustrate the internal configuration used for the evaluation of detectability and volumetric precision.

B. CT Acquisition Protocols

An energy-integrating CT system (Aquilion One PRISM Edition) from Canon Medical Systems were used in this study. All scans were acquired at a fixed $CTDI_{vol}$ of 1.4 mGy. The tube potential was set to 120 kV, and the tube current was fixed at 50 mA, a rotation time of 0.275 s, and a pitch of 0.813.

Six reconstruction protocols were evaluated, differing in reconstruction algorithms and kernels, all having a slice thickness of 0.5 mm. Three protocols used hybrid iterative reconstruction (HIR) with the Adaptive Iterative Dose Reduction 3D (AIDR 3D) algorithm and kernels FC51, FC52, and FC86, while the other three used deep learning–based reconstruction (DLR) via the Advanced Intelligent Clear-IQ Engine (AiCE) with the LUNG kernel. Model-based iterative reconstruction (MBIR) using FIRST (Forward projected model-based Iterative Reconstruction SoluTion) was also applied using the LUNG kernel. The evaluated reconstruction settings were investigated under a low-dose condition ($CTDI_{vol}$ = 1.4 mGy) and do not necessarily correspond to a standardized clinical protocol. However, they are representative of clinically relevant low-dose imaging scenarios for pulmonary nodule detection.

C. Image Analysis

The detectability index values were computed using imQuest software based on a non-prewhitening model observer with an eye filter (NPWE), following the methodology of the AAPM TG-233 report. For each imaging protocol, the median d′ was estimated from 90 repeated measurements within a single CT acquisition, using a fixed region of interest (ROI, 195 × 195 pixels) centered on the synthetic GGO nodule at the slice corresponding to its largest axial diameter. NPWE model parameters included a zoom factor of 1.5 and a viewing distance of 500 mm, and the Eckstein eye filter function $E(f)$ [13, 15 - 18]:

$$d'^2_{NPWE} = \frac{\left[\int W(f)^2 TTF^2(f) E^2(f) df\right]^2}{\int W(f)^2 TTF^2(f) E^4(f) NPS(f) df} \quad (1)$$

where f is the radial spatial frequency. The $W(f)$ represents the detection task in the frequency domain and characterizes the signal to be identified by the observer. It is defined as the magnitude of the Fourier transform of the difference between the two competing hypotheses,

$$W = |FT\{h_1(x,y) - h_2(x,y)\}| \quad (2)$$

where FT denotes the Fourier transform, $h_1(x,y)$ corresponds to the hypothesis in which the object is present, and $h_2(x,y)$. represents the background-only hypothesis.

Given that this study focuses on a lung imaging protocol, the task function was derived from a contrast profile designed to mimic a non-spiculated pulmonary nodule. The radial contrast distribution was modeled as [17]

$$c(r) = C\left[1 - \left(\frac{2r}{D}\right)^2\right]^n \quad (3)$$

where $c(r)$ is the signal amplitude at radial distance r, C represents the peak contrast relative to the background in Hounsfield units, D is the effective diameter of the lesion, and $n$controls the edge sharpness of the contrast profile. In this study, $D = 8$ mm was used. The task contrast C was estimated from the mean difference in HU between the insert and the surrounding background, resulting in values on the order of 50–100 HU.

Noise power spectra (NPS) were calculated using four ROIs placed around the nodule. NPS-derived metrics included the peak frequency (f_{peak}) and the average frequency (f_{ave}). Spatial

resolution was characterized using the task transfer function (TTF). The TTF was computed using the circular edge method, based on radial averaging of the edge spread function from a circular ROI centered on the nodule at the slice corresponding to its largest axial diameter, followed by Fourier transformation to obtain the corresponding frequency response. The TTFf_{50} was defined as the spatial frequency at which the TTF magnitude dropped to 50% of its zero-frequency value.

D. Statistical Analysis

To investigate the combined influence of noise, spatial resolution, and reconstruction strategy on nodule detectability, a log-linear regression model was developed using $\log(d')$ as the dependent variable:

$$\log(d') = \beta_0 + \beta_1 f_{\text{peak}} + \beta_2 f_{\text{ave}} + \beta_3 \text{TTF}f_{50} + \sum_k \beta_k R_k + \varepsilon \quad (4)$$

where β_0 is the intercept, β_i are the regression coefficients for the continuous predictors, β_k are the regression coefficients associated with the reconstruction method–kernel indicator variables R_k, and ε is the random error term.

Because reconstruction algorithm and convolution kernel were highly collinear (e.g., DLR was only available with the LUNG kernel), these variables were combined into a single categorical factor to avoid multicollinearity and ensure model stability. This approach allowed quantification of the relative contributions of each image quality metric and reconstruction method to detectability.

The dataset was divided into independent training (70%) and test (30%) subsets to distinguish between model fitting and model evaluation. The training set was used to estimate the relationship between detectability (d′), image quality metrics, and reconstruction settings, while the test set was reserved exclusively to assess how well this relationship generalizes to unseen data. This procedure prevents overly optimistic performance estimates that may arise when a model is evaluated using the same data employed for parameter estimation.

Model predictions of d′ on the test set were compared with measured values using the coefficient of determination (R^2), the root mean squared error (RMSE), and the mean absolute error (MAE). These metrics respectively quantify the fraction of variance explained by the model, the typical magnitude of prediction errors, and the average absolute deviation between predicted and observed detectability values.

The RMSE and MAE were defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(d_i' - \hat{d}_i')^2}, \quad (5)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} | d_i' - \hat{d}_i' |, \quad (6)$$

where d_i' are the observed detectability values, $\hat{d}_i'$ are the model-predicted values, and n is the number of observations in the test set.

III. RESULTS

The log-linear regression model was used to investigate the association between image quality metrics and the detectability index, where higher values indicate improved detectability, using log(d′) as the dependent variable. The log-linear regression model showed strong explanatory power (adjusted $R^2 = 0.80$), with the overall model being statistically significant ($p < 0.001$). Among the noise-related predictors, both f_{peak} and f_{ave} were positively and significantly associated with detectability. Specifically, f_{peak} presented a regression coefficient of 0.025 ($p < 0.01$), indicating that increases in f_{peak} noise power spectrum frequency are associated with higher detectability. f_{ave} had an even stronger effect, with a coefficient of 0.072 ($p < 0.001$), confirming that average noise frequency content plays a major role in influencing detection performance.

Spatial resolution, represented by TTFf_{50}, was negatively associated with detectability, with a coefficient of −0.203 ($p < 0.001$), indicating that higher cutoff frequencies (i.e., sharper images) are associated with reduced d′ in this dataset, likely reflecting the trade-off between noise and resolution in low-contrast detection tasks.

Reconstruction method and kernel combinations also had a significant impact on detectability. Compared to the reference reconstruction (HIR/FC86), the DLR/LUNG and MBIR/LUNG combinations showed the strongest positive effects, with coefficients of 0.203 and 0.220, respectively (both $p < 0.001$), indicating substantially improved detectability. The HIR/FC52 combination was associated with a significant reduction in detectability (coefficient −0.124, $p < 0.001$), while HIR/FC51 showed a smaller but still statistically significant positive effect (coefficient 0.043, $p < 0.05$).

The model also demonstrated strong predictive performance when applied to an independent test set, achieving an R^2 of 0.79, an RMSE of 1.34, and an MAE of 1.02. This indicates that the relationship identified between detectability, image quality metrics, and reconstruction settings is not specific to the data used to fit the model, but remains valid when evaluated on independent measurements. In practical terms, the model was able to accurately reproduce observed variations in detectability across different noise characteristics, spatial resolution levels, and reconstruction strategies. Inter-acquisition variability was not evaluated, as only a single scan was performed for each protocol; however, 90 repeated measurements per acquisition provided an estimate of intra-acquisition stability.

These findings demonstrate that detectability is strongly influenced by both frequency-domain noise characteristics and reconstruction strategy, with advanced reconstruction methods yielding the highest gains in performance.

IV. CONCLUSIONS

The present work demonstrates that detectability in anthropomorphic phantom imaging is governed by a combination of noise texture, spatial resolution, and

reconstruction strategy. Among these, the average noise frequency and reconstruction method exert particularly strong effects. The positive and significant impact of deep learning–based reconstruction underscores the importance of incorporating task-based performance metrics into protocol optimization and system evaluation. This task-based study was conducted using a single acquisition and a single GGO nodule, enabling a systematic evaluation under well-defined conditions. Detectability was assessed using a NPWE model observer, providing a task-based framework for comparative evaluation across reconstruction methods. These results are specific to the acquisition conditions and system investigated, and their generalization to other scanners or dose levels remains to be established. The absence of a human observer study implies that the relationship between model observer detectability (d′) and clinical performance remains indirect. Future work will extend this framework to multiple acquisitions, a broader range of nodules, and human observer validation.

V. ACKNOWLEDGMENT

The authors wish to thank the São Paulo Research Foundation-FAPESP (processes 2022/11457-0, 2018/05982-0 and 2023/03945-8) and by National Council for Scientific and Technological Development - CNPq (process 141335/2021-1 and 302986/2023-5).

REFERENCES

[1] Guedes Pinto E, Penha D, Ravara S, Monaghan C, Hochhegger B, Marchiori E, et al. Factors influencing the outcome of volumetry tools for pulmonary nodule analysis: a systematic review and attempted meta-analysis. *Insights Imaging*. 2023;14:152.

[2] Wood DE, Kazerooni EA, Baum SL. NCCN Clinical Practice Guidelines in Oncology: Lung Cancer Screening. *J Natl Compr Canc Netw*. 2026.

[3] Kim D, Park JH, Lee CH, Kim YJ, Kim JH. Improved consistency of lung nodule categorization in CT scans with heterogeneous slice thickness by deep learning-based 3D super-resolution. *Diagnostics*. 2024;15(1):50.

[4] Jeong J, Park D, Kang JH, Kim M, Kim HY, Choi W, et al. Deep learning-based slice thickness reduction for computer-aided detection of lung nodules in thick-slice CT. *Diagnostics*. 2024;14:2558.

[5] Denzler S, Vuong D, Bogowicz M, Pavic M, Frauenfelder T, Thierstein S, et al. Impact of CT convolution kernel on robustness of radiomic features for different lung diseases and tissue types. *Br J Radiol*. 2021;94.

[6] Tanabe N, Kaji S, Shima H, Shiraishi Y, Maetani T, Oguma T, et al. Kernel conversion for robust quantitative measurements of archived chest computed tomography using deep learning-based image-to-image translation. *Front Artif Intell*. 2022;4:769557.

[7] Mao Y, Heuvelmans MA, van Tuinen M, Yu D, Yi J, Oudkerk M, et al. Artificial intelligence as an independent reader of risk-dominant lung nodules: influence of CT reconstruction parameters. *Eur Radiol*. 2025.

[8] Tsuchiya N, Kobayashi S, Nakachi R, Tomori Y, Yogi A, Iida G, et al. Application of a pulmonary nodule detection program using AI technology to ultra-low-dose CT: differences in detection ability among various image reconstruction methods. *[Journal not specified]*.

[9] Jiang B, Li N, Shi X, Zhang S, Li J, de Bock GH, et al. Deep learning reconstruction shows better lung nodule detection for ultra–low-dose chest CT. *Radiology*. 2022;303:202–212.

[10] Wang J, Zhu Z, Pan Z, Tan W, Han W, Zhou Z, et al. Deep learning reconstruction improves computer-aided pulmonary nodule detection and measurement accuracy for ultra-low-dose chest CT. *BMC Med Imaging*. 2025;25:200.

[11] Burgess AE, Li X, Abbey CK. Visual signal detectability with two noise components: anomalous masking effects. *J Opt Soc Am A*. 1997;14:2420–2442.

[12] Barrett HH, Myers KJ. *Foundations of Image Science*. Hoboken: John Wiley & Sons; 2013.

[13] Pimenta EB, Costa PR. Model observers and detectability index in x-ray imaging: historical review, applications and future trends. Phys Med Biol. 2025;70:07TR2.

[14] Costa PR, Boiset GR, Pimenta EB, Rocha RMV, Moura RAS, Marques WH, et al. Hybrid phantom for lung CT: design and validation. *Med Phys*. 2025;52:e17990.

[15] Ishida M, Doi K, Loo LN, Metz CE, Lehr JL. Digital image processing: effect on detectability of simulated low-contrast radiographic patterns. *Radiology*. 1984;150(2):569–575.

[16] Abadi E, Harrawood B, Sharma S, Kapadia A, Segars WP, Samei E. DukeSim: a realistic, rapid, and scanner-specific simulation framework in computed tomography. *IEEE Trans Med Imaging*. 2019;38(6):1457–1465.

[17] Ria F, Solomon JB, Wilson JM, Samei E. Technical note: validation of TG 233 phantom methodology to characterize noise and dose in patient CT data. *Med Phys*. 2020;47(4):1633–1639.

[18] Eckstein MP, Bartroff JL, Abbey CK, Whiting JS, Bochud FO. Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal-known-exactly detection tasks. *Opt Express*. 2003;11(5):460–475.

Deep Learning Enhancement of Ultralow-Dose 3D CT Scouts for Organ Localization and Dose

Liyan Sun, Sen Wang, Maria Jose Medrano, Kaylee W. Fang, Grant M. Stevens, Justin R. Tse, and Adam S. Wang

***Abstract*—Conventional CT scout images are 2D X-ray radiographs used to position patients and plan CT scans. Recently, 3D CT scout techniques have been developed to reveal 3D anatomical information of patients at dose levels similar to conventional 2D scouts. In this study, we propose to enhance the image quality of ultralow-dose 3D CT scouts using deep learning in a simulation study that emulated 1% dose scans from clinical body CT scans. The deep learning image quality enhancer is a modified 2.5D Nonlinear Activation Free Network (NAFNet) to better exploit inter-plane correlation. The 3D scouts were then used to estimate CT organ dose via automated organ segmentation and Monte Carlo-based dose calculation. The resulting enhanced image quality, organ segmentation, and organ dose were compared against that of the clinical CT scan. The experimental results demonstrate the improved image quality after enhancement both quantitatively and qualitatively. The enhanced 3D scouts offer accurate automatic organ segmentation and organ dose estimates, thus presenting a pathway to patient-specific organ-based dose optimization.**

***Index Terms*—CT scout, CT dose, Deep learning, CT image quality improvement.**

I. Introduction

COMPUTED tomography is a non-invasive medical imaging technique widely used in clinical practice. Before acquiring CT scans, scout images, which are also called topograms, scanograms, or localizer radiographs, are routinely acquired to position patients, define the CT scan range, and adapt the tube current to control radiation exposure to patients. Although some body landmarks like lungs and bones can be identified from conventional 2D scouts, other low-contrast organs are less visible, making it challenging to perform organ-based dose estimation and optimization. However, patient-specific organ-based dose optimization is critical to achieve the "As Low As Reasonably Achievable" (ALARA) principle by adapting CT dose to each individual and enabling task-specific acquisitions where CT dose is applied to regions of interest and reduced in other regions.

Recently, ultralow-dose 3D CT scouts were proposed to acquire low-SNR yet volumetric information in contrast to 2D information from conventional scouts, where the 3D scout dose is kept similar to that of conventional 2D scouts. Yin et al. demonstrated the organ segmentation capability of ultra-low dose 3D scouts at 1.25% of the dose of a traditional CT scan by suitable acquisition, preprocessing, and reconstruction [1]. In their study, continuous-exposure and pulsed acquisitions were investigated, and filtered-back-projection reconstruction with a smoothing kernel to suppress noise and sparse-angle artifacts were applied. However, the potential of deep learning approaches to improve image quality of ultra-low dose 3D scouts was still underexplored. The acquisition of low-dose 3D scout scans for cone-beam computed tomography, whose dose level could go as low as around 5%, was also investigated [2]. Alternatively, Golbus et al. studied the use of 3D landmark scans, which are reduced-dose helical CTs at around 6% of the dose of a traditional scan, which also reduces the acquisition range of the main CT scan [3]. Dose reduction was also achieved by SilverBeam [4], which is a beam-shaping energy filter that selectively removes low energies with a silver filter to harden the beam and reduce overall exposure. This approach is vendor-specific and requires ad hoc hardware design. Bortnikov et al. developed a deep-learning-based image denoising in the projection domain for ultralow-dose 3D scout denoising [5]. The low-dose emulation was simulated by a known Poisson noise model. Although the image quality of 3D scouts was improved through projection-domain enhancement, the downstream application such as organ segmentation and dose calculation remains to be studied.

In this paper, we propose an image quality improvement approach to ultralow-dose 3D CT scouts using a 2.5D Nonlinear Activation Free Network (NAFNet) based on a simulation study. The image quality improvement of 3D scouts is validated by qualitative and quantitative assessments. We also demonstrated the improved automatic segmentation accuracy

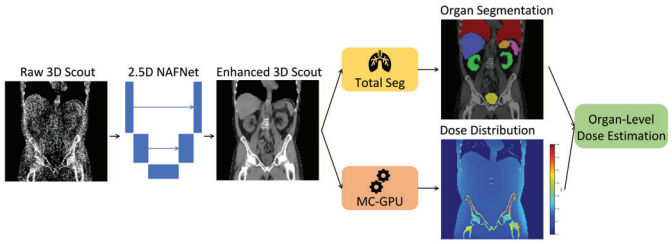


Fig. 1. Flowchart of image quality enhancement and application of ultralow-dose 3D CT scouts. TotalSegmentator and MC-GPU use the enhanced 3D scout to predict abdominal organ segmentation and dose distribution to estimate organ-level CT dose.

Liyan Sun, Sen Wang, Maria Jose Medrano, Kaylee W. Fang, and Justin R. Tse, are with the Department of Radiology, Stanford University, Stanford, CA USA 94305 USA (e-mail: sunly92@stanford.edu; senwang@stanford.edu; medranom@stanford.edu; kwfang@stanford.edu; jrtse@stanford.edu).

Grant M Stevens is with GE HealthCare, Waukesha, WI USA 53188 (email:grant.stevens@gehealthcare.com)

Adam S. Wang is with the Departments of Radiology and Electrical Engineering, Stanford University, Stanford, CA USA 94305 (email: adamwang@stanford.edu).

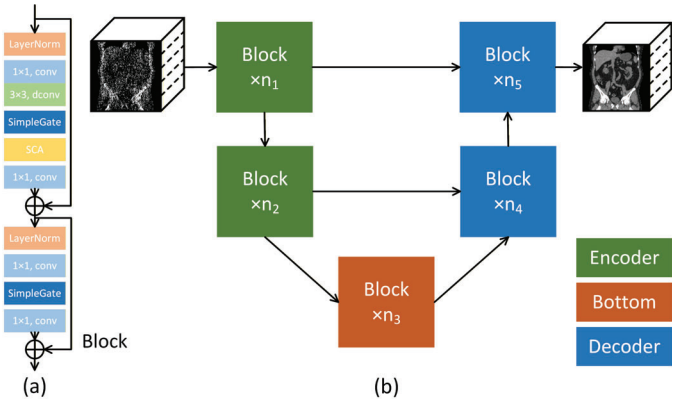


Fig. 2. Architecture of a 3-layer 2.5D NAFNet taking axial-plane sliced subvolumes as input. (a) Structure of the block (b) Network architecture.

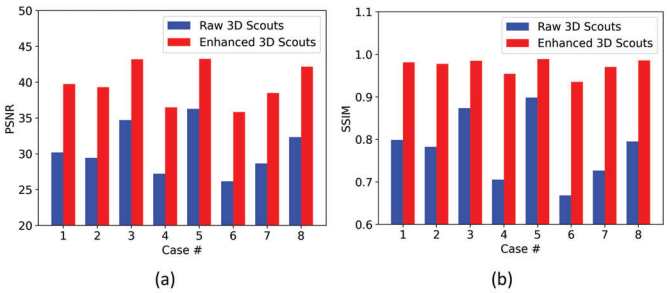


Fig. 3. Barplots of PSNR and SSIM results of raw 3D scouts and enhanced 3D scouts. (a) PSNR (b) SSIM.

and organ dose estimation of enhanced 3D scouts. Our study lays the foundation for optimized organ-based tube current modulation based on 3D scouts.

II. Method

We present the framework for enhancing ultralow-dose 3D CT scouts, as shown in Fig 1. The 2.5D NAFNet takes raw 3D scouts and predicts enhanced 3D scouts in a supervised learning manner where the references are full-dose body CT examinations. Then, the open-source TotalSegmentator model automatically segments the main abdominal organs from our enhanced 3D scouts. We also infer the body attenuation from the enhanced 3D scouts and feed it into the open-source Monte Carlo simulation software, MC-GPU, to calculate the body dose distribution of a CT scan. With organ segmentation and body dose distribution, we can then estimate organ-level CT dose.

A. 2.5D Nonlinear Activation Free Network

Nonlinear Activation Free Network (NAFNet) [6] breaks down the design techniques of state-of-the-art attention-based deep learning architectures and condenses them into a low-computational complexity activation-free network, which is still nonlinear and retains expressivity. The NAFNet is promising for processing high-resolution CT images with low complexity. However, the NAFNet is a 2D deep learning approach and was originally proposed for natural images. The architecture of a 3-layer NAFNet is shown in Fig. 2. The SimpleGate module and Simple Channel Attention (SCA) achieve the spatial and channel attention. Their details could be found in [6]. The layer normalization helps stabilize training, and the separable convolution, composed of 1×1 convolution and depth-wise 3×3 convolution, reduces network complexity. The feature dimension heads is 32. The 2-stepsize convolution and pixel shuffle are used to undersample and upsample feature maps. Shortcuts are used to propagate shallow-layer features to symmetric deep layers directly. Here, we design a 5-layer 2.5D NAFNet, where the configuration for the encoder is $(n_1, n_2, n_3, n_4) = (4, 4, 4, 4)$, for the bottom $(n_5) = (4)$, and for the decoder $(n_6, n_7, n_8, n_9) = (4, 4, 4, 4)$. Here, we adapt the NAFNet to its 2.5D version to better exploit the inter-plane information in CT volumes, and we show the volume slicing in Fig. 2 with dashed lines. In training, we randomly crop 8 consecutive axial slices from raw 3D scouts, and the network is trained to predict their full-dose counterparts. In testing, a sliding-window operation crops 3D scouts into raw subvolumes, and the enhanced subvolumes are aggregated back with average weighting of a step size of 1.

B. Abdominal Organ Segmentation

Organ-level dose estimation requires organ segmentation. We use the open-source TotalSegmentator [7] to segment the major abdominal organs from CT scans. TotalSegmentator has shown robustness to CT dose as low as 20% dose in some studies [8], but its performance has not been evaluated on CT dose levels as low as 1% nor on enhanced scouts at this dose level. In addition to organ-level dose estimation, the organ segmentation can also be used to determine the main CT scan range. With TotalSegmentator, organ segmentation maps S_i are generated, where $i = \{1, 2, 3, 4, 5, 6, 7\}$ denotes the lungs, kidneys, liver, bladder, spleen, pancreas, and stomach.

C. Body Dose Estimation

We use the open-source MC-GPU [9] for personalized CT dose estimation through CUDA-accelerated Monte Carlo simulation. We simulate a helical trajectory that is consistent with the real data acquisition. To model realistic CT systems, we modified the original MC-GPU by adding the bowtie filter and anode heel effect [10].

The 3D scouts were used to produce patient-specific voxelized phantoms, which contain both material type and volumetric mass density. To reduce computational burden, we resampled 3D scouts to isotropic voxels of $4 \times 4 \times 4\ mm^3$. The density mapping was done by following a piece-wise linear curve that is used to define densities of mixtures of water and bone. The MC-GPU dose simulation was repeated 4 times with uniformly spaced start angles (0°, 90°, 180°, and 270°), since the actual start angles cannot be controlled prospectively. Then, the 4 dose maps were averaged to produce the final CT dose map. In this study, we set a constant tube current of 400 mAs. With the final dose map D and organ segmentation maps S_i, the average CT dose of the i_{th} organ can be calculated as

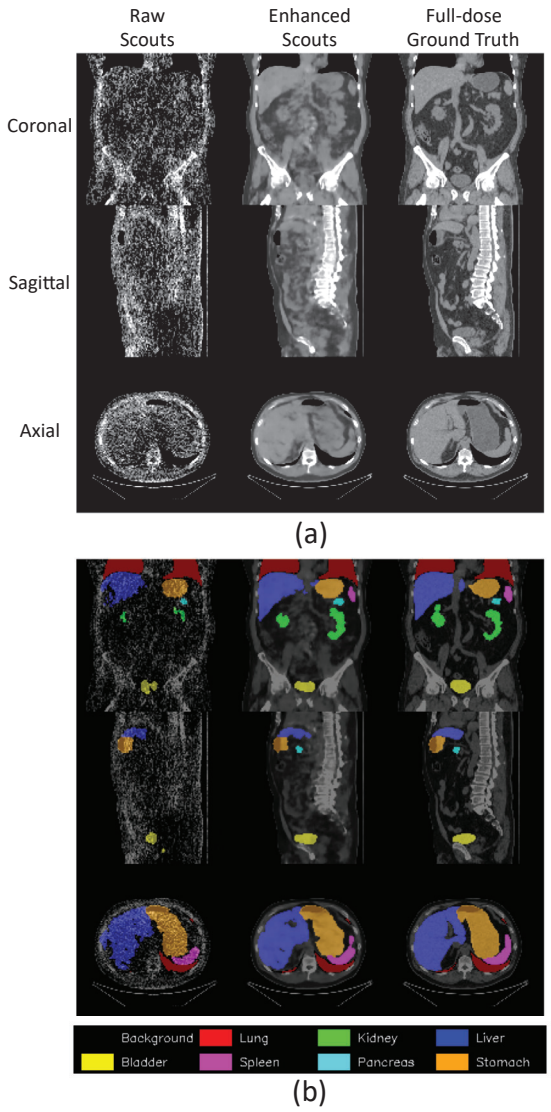


Fig. 4. Multi-view display of the CT image and segmentation of raw 3D scout, enhanced 3D scout, and full-dose ground truth of one test case. (a) CT images, (b) abdominal segmentation.

$$d_i = \frac{\sum_{x,y,z} D\left(x,y,z\right) \cdot S_i\left(x,y,z\right)}{\sum_{x,y,z} S_i\left(x,y,z\right)}, \tag{1}$$

where (x, y, z) represents the position in the dose map and organ segmentation.

III. Experimental Setting

We retrospectively collected non-contrast-enhanced Intravenous Pyelogram (IVP) abdominal CT images from 38 patients, where we intentionally selected a non-contrast protocol since scouts are typically non-contrast. The CT scans were acquired using a GE HealthCare Revolution Apex scanner with various tube potentials. The in-house data were collected under IRB approval from Stanford Health Care. The original resolution of the CT scans is $0.9765 \times 0.9765 \times 2.5\ mm^3$, with an in-plane matrix size of 512×512. The ultralow-dose 3D scouts at 1% dose level were emulated by noise insertion

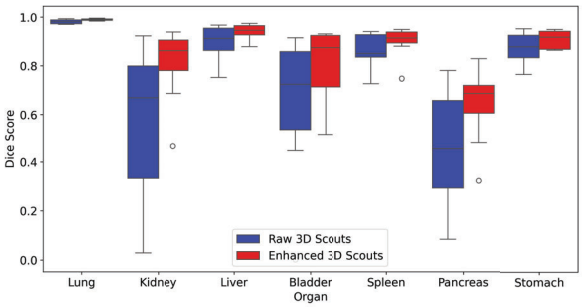


Fig. 5. Boxplots of the abdominal organ segmentation of raw 3D scouts and enhanced 3D scouts.

[11]. We randomly split the dataset into a training set of 23 cases, validation set of 7 cases, and test set of 8 cases.

In training, the mini-batch size was set to 1 with two randomly cropped subvolumes in each batch. The model was trained for 10,000 epochs, with an initial learning rate of $5e{-}5$. The learning rate was halved after the first 7,000 epochs. We adopt Mean Squared Error (MSE) against the full-dose CT as the loss function, and the model with the minimal MSE loss on the validation set was used for testing.

IV. Results

A. Assessment of CT Image Quality

We first evaluated the image-quality improvement of 3D scouts from the deep learning approach on 8 test cases. The Peak Signal-to-Noise Ratio (PSNR) results are shown in Fig. 3(a). The average PSNR for raw 3D scouts is 30.61 dB, while the average PSNR for enhanced 3D scouts is improved to 39.80 dB. The Structural Similarity Index Measure (SSIM) results are shown in Fig. 3(b). The average SSIM for raw 3D scouts is 0.7809, and the average SSIM for enhanced 3D scouts is improved to 0.9721. We show the raw 3D scouts, enhanced 3D scouts, and full-dose ground-truth CT images of one test case in Fig. 4(a) with display window [-150, 250] HU. The CT volume is sliced in the center along the three orthogonal-view planes. Due to noise and artifacts, the liver boundary is invisible, and other organs like the kidneys, bladder, and pancreas are unidentifiable from the emulated 1% dose scan. After enhancement, we observe much better agreement between the CT scout and the full-dose ground truth: the liver boundary is sharper, and other abdominal organs are restored. The consistency between the quantitative and qualitative assessments demonstrates the effectiveness of the deep learning approach to improve the image quality of ultralow-dose 3D CT scouts.

B. Performance of Organ Segmentation

With enhanced 3D scouts, we apply TotalSegmentator to generate organ segmentations and compare them with that of raw 3D scouts. We evaluate segmentation with the Dice score and show the boxplots in Fig. 5. TotalSegmentator is surprisingly robust to an extreme 1% CT dose level when segmenting lungs, liver, and spleen, with an average Dice score of 0.9811, 0.8959, and 0.8609, while the deep learning enhancement

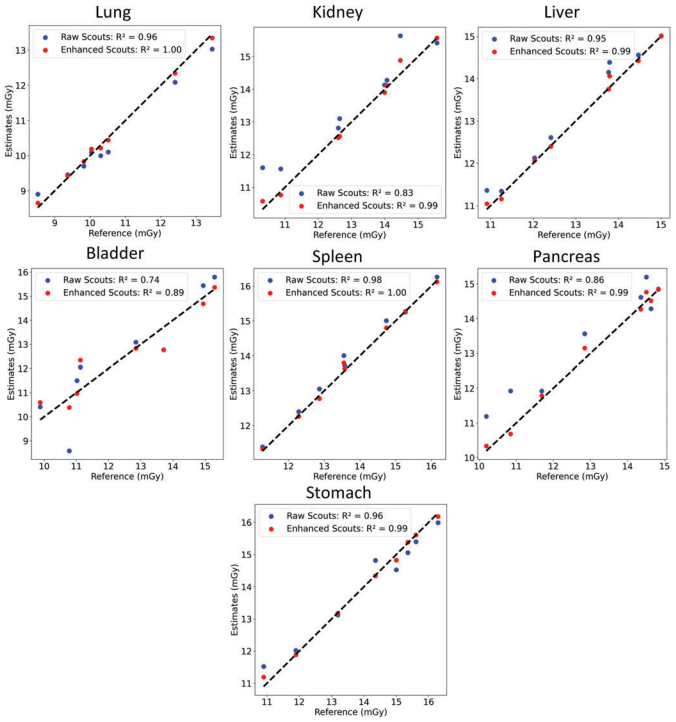


Fig. 6. Scatter plots of dose estimation of 7 abdominal organs.

still gives better segmentation accuracy at 0.9897, 0.9410, and 0.8986. The performance of TotalSegmentator on raw 3D scouts deteriorates to 0.5638, 0.6389, and 0.4614 when segmenting kidneys, bladder, and pancreas. After enhancement, the segmentation accuracy of the 3 challenging organs rises to 0.8070, 0.7994, and 0.6415. In Fig. 4(b), TotalSegmentator failed to segment the kidneys and spleen in the coronal image, and failed to segment the pancreas in the sagittal image. In contrast, the deep learning enhancement consistently improves the segmentation quality across all organs.

C. Performance of Organ-Level Dose Estimation

The downstream application of organ segmentation and dose distribution estimation is organ dose estimation. We calculate the average CT dose in mGy of 7 abdominal organs using Eq. 1 on 8 test cases and report the results in Fig. 6. The reference CT doses are calculated based on the full-dose ground truth CT images. We observe that the enhanced 3D scouts better approximate the CT dose reference in terms of the R^2 score. Notably, the R^2 scores of the kidneys, bladder, and pancreas are improved from 0.83, 0.74, and 0.86 to 0.99, 0.89, and 0.99. In an optimization framework for organ-level tube current modulation, the precision of dose estimation of all organs, including these three, is critical.

V. Discussion and Conclusions

In this paper, we present a deep learning method, a 2.5D nonlinear activation-free network, to enhance ultra-low dose CT scouts (1% dose) toward the image quality of full-dose counterparts in a simulation study on a clinical dataset. The automatic segmentation of abdominal organs is performed using TotalSegmentator to extract organ masks. The enhanced image quality from the deep learning model brings improved accuracy of organ masks. We also calculated dose distribution using GPU-accelerated Monte Carlo simulation based on the voxelized phantom from 3D scouts. Then, we estimated the average organ doses based on organ masks and dose distribution, and demonstrate that more precise organ dose estimation can be achieved via deep-learning image quality enhancement. Note that the image quality enhancement can be achieved by other general deep learning models, and we use the 2.5D NAFNet as the representative model in this paper and will explore other advanced models for future research. Direct prediction of organ mask and body dose distribution from ultralow 3D scouts is an alternative to the proposed image quality enhancement approach and will be studied in the future research. Besides, the raw 3D scouts can also be corrected Low Signal Correction (LSC).

The noise emulation in this study is based on uniform dose reduction (e.g., 1%) of the full-dose CT profile. In future studies, we will investigate the dynamics of the accuracy of organ segmentation and organ dose estimation with different ultra-low dose levels or modulation for 3D scouts. We will also explore more data acquisition strategies, like pulsed acquisition. Then, we will validate our simulations and approach on real ultra-low dose 3D CT scouts. In addition, image noise is directly related to the CT dose, so we will develop a noise estimation approach to produce CT noise maps as a function of CT dose, with ultra-low dose 3D scouts as input. Both the prospective organ-level dose and noise estimations are essential to optimizing patient-specific organ-based tube current modulation.

Acknowledgments

This work was supported by GE HealthCare.

References

[1] Z. Yin, Y. Yao, A. Montillo, M. Wu, P. M. Edic, M. Kalra, and B. De Man, "Acquisition, preprocessing, and reconstruction of ultralow dose volumetric CT scout for organ-based CT scan planning," *Medical Physics*, vol. 42, no. 5, pp. 2730–2739, 2015.

[2] J. Gomes, G. J. Gang, A. Mathews, and J. W. Stayman, "An investigation of low-dose 3D scout scans for computed tomography," in *Proceedings of SPIE–the International Society for Optical Engineering*, vol. 10132, 2017, p. 101322M.

[3] A. E. Golbus, J. L. Schuzer, C. Steveson, S. F. Rollison, J. Matthews, J. Henry-Ellis, M. Razeto, and M. Y. Chen, "Reduced dose helical CT scout imaging on next generation wide volume CT system decreases scan length and overall radiation exposure," *European Journal of Radiology Open*, vol. 13, p. 100578, 2024.

[4] A. E. Golbus, C. Steveson, J. L. Schuzer, S. F. Rollison, T. Worthy, A. M. Jones, P. Julien-Williams, J. Moss, and M. Y. Chen, "Ultra-low dose chest CT with silver filter and deep learning reconstruction significantly reduces radiation dose and retains quantitative information in the investigation and monitoring of lymphangioleiomyomatosis (LAM)," *European Radiology*, vol. 34, no. 9, pp. 5613–5620, 2024.

[5] M. Bortnikov, F. Bergner, A. Chernyavskiy, K. M. Brown, N. Black, and M. Grass, "3D scout scans using projection domain denoising," in *Medical Imaging with Deep Learning*, 2021.

[6] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple baselines for image restoration," in *European Conference on Computer Vision*. Springer, 2022, pp. 17–33.

[7] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang *et al.*, "TotalSegmentator: robust segmentation of 104 anatomic structures in CT images," *Radiology: Artificial Intelligence*, vol. 5, no. 5, p. e230024, 2023.

[8] A. Tsanda, H. Nickisch, T. Wissel, T. Klinder, T. Knopp, and M. Graß, "On TotalSegmentator's performance on low-dose CT images," in *Medical Imaging 2024: Image Processing*, vol. 12926. SPIE, 2024, pp. 86–89.

[9] A. Badal and A. Badano, "Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit," *Medical Physics*, vol. 36, no. 11, pp. 4878–4880, 2009.

[10] S. Wang, M. J. Medrano, A. A. Z. Imran, W. Lee, J. J. Cao, G. M. Stevens, J. R. Tse, and A. S. Wang, "Automated estimation of individualized organ-specific dose and noise from clinical CT scans," *Physics in Medicine & Biology*, vol. 70, no. 3, p. 035014, 2025.

[11] T. M. Benson and B. K. De Man, "Synthetic CT noise emulation in the raw data domain," in *IEEE Nuclear Science Symposium & Medical Imaging Conference*. IEEE, 2010, pp. 3169–3171.

Physics-Informed Scalable CycleGAN for CT Reconstruction Kernel Conversion and Harmonization

Seoyoung Lee, Aeden K. Davis, Shravani A. Kharat, Wilson I. Gonsalves, Shuai Leng, Lifeng Yu, Cynthia H. McCollough, Francis I. Baffour* and Hao Gong*

Abstract*— CT reconstruction kernel variation compromises the robustness of quantitative analysis in retrospective studies and the generalizability of deep learning models. Prior data-driven methods exhibit strong performance on kernels seen in training, while they are oftentimes limited in interpretability and generalizability on unseen kernels. To tackle these challenges, we develop a physics-informed, scalable CycleGAN that explicitly employs kernel properties and training-free calibration scheme. Our model uses a two-stage single-generator architecture: (1) a deblurring stage using a U-Net with learnable Wiener deconvolution layers, and (2) a target-kernel imposing stage, employing an Image Fusion Module that synthesizes images via a learnable parameter α. Our method extends to unseen kernels by calibrating this parameter. For proof-of-concept validation, our method was compared against two baseline models over abdominal CT scans (n=13 cases; n=2 kernels) via visual inspection, structural similarity (SSIM), modulation transfer function (MTF), and CT number accuracy. Our method outperformed the baseline, achieving higher total SSIMs (0.949 and 0.987) for bidirectional conversions between medium-smooth and -sharp kernels, high CT number accuracy (HU bias < 10), and high MTF consistency (e.g., error at 10% MTF < 1.20%). Scalability tests exhibited high SSIM (> 0.946), strong MTF agreement (e.g., error at 10% MTF < 2.20%), and high noise textural similarity (spectral correlation mapper; SCM ≥ 0.922) on unseen kernels. In summary, our method can provide accurate, scalable CT kernel conversion, thereby facilitating effective data harmonization, robust quantitative analyses, and enhancing deep-learning model performance across diverse down-stream clinical tasks.

***Index Terms*—Reconstruction Kernel Conversion, Harmonization, Deep Learning, Physics-Informed, Scalability**

I. Introduction

The task-specific selection of reconstruction kernels leads to inconsistent CT image characteristics, substantially degrading the performance of downstream quantitative and radiomic analyses. Similar challenges arise in AI-driven workflows, as domain-specific variations in texture, sharpness, and CT numbers contribute to performance variability in pre-trained AI models. Since raw projections are typically discarded due to storage limitations, kernel variability remains a primary obstacle in longitudinal, multicenter, and opportunistic CT studies. Retrospective kernel conversion offers a viable solution by harmonizing images to a unified kernel. Such harmonization improves the consistency of image analysis, reduces potential biases during AI training, and facilitates robust and reliable performance during clinical deployment.

Recent advances in deep learning have demonstrated potential in CT kernel conversion. Early studies employed supervised learning frameworks with paired datasets [1]. However, since obtaining adequate paired data is often impractical in clinical settings, unsupervised learning approaches have been explored. Representative examples include CycleGAN-based methods [2, 3]. To bridge the gap between data-driven models and imaging physics, some studies have attempted to incorporate kernel information as an auxiliary input [4]. Nonetheless, the mechanism oftentimes remains implicit in prior methods, resulting in limited model interpretability and a lack of physical inductive bias. This further affects model generalizability and scalability across clinical CT tasks using new kernels unseen in training.

We address this by developing SPIKE-GAN (Scalable Physics-informed Kernel-Explicit GAN) which utilizes single-generator CycleGAN architecture. Our method differs from prior methods in the following way: the network is conditioned on patient-specific, kernel-dependent point spread functions (PSFs) of input images and target reconstruction settings, enabling explicit integration of kernel characteristics during conversion. This physical conditioning provides a training-free, tunable framework that can be re-calibrated to new kernels unseen in training, ensuring high scalability across diverse clinical protocols.

II. Materials and Methods

A. CT data preparation

The proposed model was trained and validated using patient cases from NIH-AAPM-Mayo Low Dose CT Grand Challenge [5]. In this pilot study, we randomly selected 13 routine-dose scans to generate training (n=6), validation (n=2), and testing (n=5) groups, comprising 3,469, 1,085, and 2,727 images, respectively. The original dataset involved paired medium-smooth (B30) and medium-sharp (D45) kernels. New images were reconstructed using commercial filtered-back-projection with 3 additional kernels (B20, B40, and B50). Reconstruction field-of-view ranged from 340 to 420 mm. Slice thickness and increment were 1.0 mm and 0.8 mm, respectively.

B. MTF and PSF estimation

We characterized reconstruction kernels by modulation transfer function (MTF), which was estimated in a patient-specific manner. First, the tissue region was isolated from axial cross-sections of the B30 images by applying cropping and

All authors are affiliated with Department of Radiology, Mayo Clinic, Rochester, MN 55901, USA. ***Corresponding authors**: F. I. Baffour (Baffour.Francis@mayo.edu) and H. Gong (Gong.Hao@mayo.edu)

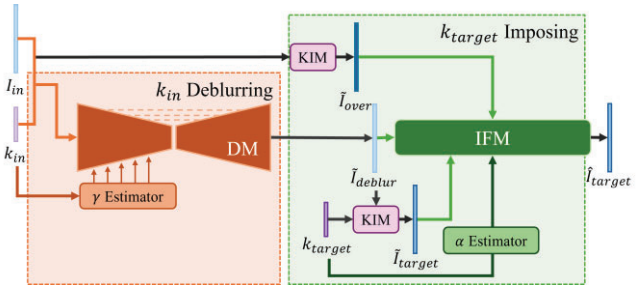


Fig. 1. Overall diagram of SPIKE-GAN generator. It has two sequential stages: input kernel deblurring and target kernel imposing. DM: Deblurring Module, IFM: Image Fusion Module, KIM: Kernel Imposing Module.

intensity clipping. Binary Otsu thresholding was then used to further refine the segmentation, after which the largest contour corresponding to the patient body was identified. The edge spread functions (ESFs) were measured from the patient-air boundary. The line spread functions (LSFs) were subsequently calculated after oversampling and fitted with a Gaussian function to derive the MTF. The PSF was calculated from the MTF under the assumption of an isotropic response and saved as a 15×15 image.

C. The Network

Fig. 1 illustrates the overall architecture of SPIKE-GAN. The proposed generator has two sequential stages. The first stage serves as a non-blind deblurring network, designed to restore a sharp, non-blur image from the input. U-Net was selected as a backbone. This sub-network receives dual input: the input CT image (I_{in}) and its corresponding estimated PSF (k_{in}). To provide explicit guidance for the deblurring task, we integrated a Wiener deconvolution layer to the contracting path of the U-Net. In this layer, a learnable regularization parameter (γ) adaptively controls the strength of the filtering at each feature level. γ is estimated via a Multi-Layer Perceptron (MLP) module from the input PSF.

The second stage involves imposing the target kernel into the deblurred image. Ideally, since the output of the first stage approximates the underlying un-blurred image, simply convolving it with the target PSF should yield the desired kernel conversion result. However, given the inherent uncertainties in both kernel estimation and the deblurring performance, a simple convolution is likely to fail. To mitigate this, we designed an Image Fusion Module (IFM) which takes three candidate basis images:

$$\tilde{I}_{deblur} = \mathrm{Deblur}(I_{in}, k_{in}), \quad (1)$$

$$\tilde{I}_{target} = \mathrm{KIM}(\tilde{I}_{deblur}, k_{target}), \quad (2)$$

$$\tilde{I}_{over} = \mathrm{KIM}(I_{in}, k_{in}), \quad (3)$$

where I denotes the image, k represents the PSF, $\mathrm{Deblur}(\cdot)$ refers to the first stage deblurring module. $\mathrm{KIM}(\cdot)$ represents a Kernel Imposing Module, implemented as a 2D convolution. Of note, $\tilde{I}_{over}$ denotes an oversmoothed image, which is used to broaden the interpolation range. The contribution of these basis images to the final output is governed by a trainable parameter α, which is defined as a function of the target PSF (k_{target}):

$$\alpha = f_{\alpha}(k_{target}). \quad (4)$$

In this study, the estimator $f_{\alpha}(\cdot)$ is implemented as a MLP consisting of two linear layers followed by a tanh activation function. Consequently, the final target image is computed as:

$$\hat{I}_{target} = \mathrm{IFM}(\tilde{I}_{deblur}, \tilde{I}_{target}, \tilde{I}_{over} \mid \alpha), \quad (5)$$

where $\mathrm{IFM}(\cdot)$ is implemented as a weighted linear combination of the basis images on α:

$$\mathrm{IFM}(\tilde{I}_{deblur}, \tilde{I}_{target}, \tilde{I}_{over} \mid \alpha) = \begin{cases} \alpha\tilde{I}_{deblur} + (1-\alpha)\tilde{I}_{target}, & \alpha \geq 0 \\ -\alpha\tilde{I}_{over} + (1+\alpha)\tilde{I}_{target}, & \alpha < 0 \end{cases}. \quad (6)$$

The network is trained using a bidirectional LS-GAN framework, inspired by CycleGAN architecture. The main difference from vanilla CycleGAN is that our proposed model utilizes a single generator. By conditioning the generator on the PSF of both input and target domains, a single generator can perform bidirectional kernel conversion.

D. Loss functions

In addition to the standard CycleGAN losses, we introduced fidelity losses to ensure that the first stage deblurring module performs effectively:

$$L_{fidelity} = \left|\left|I_{in} - \tilde{I}_{deblur} * k_{in}\right|\right|_1, \quad (7)$$

$$L_{fidelity,FT} = \left|\left|\log\left(1 + FFT(I_{in})\right) - \log\left(1 + FFT(\tilde{I}_{deblur} * k_{in})\right)\right|\right|_1. \quad (8)$$

The Fourier domain loss was added to mitigate the blurring effects typically associated with the use of L1 loss in the spatial domain. The total generator loss is then defined as:

$$L_G = \lambda_{GAN}L_{GAN} + \lambda_{cycle}L_{cycle} + \lambda_{identity}L_{identity} + \lambda_{fidelity}L_{fidelity} + \lambda_{fidelity,FT}L_{fidelity,FT}, \quad (9)$$

where λ denotes the corresponding weights determined empirically. The discriminator was implemented and trained following the PatchGAN architecture.

E. Training setup

Horizontal flipping and random cropping to a size of 256×256 were applied for data augmentation. We used 200 training epochs with Adam optimizer (learning rate fixed at 1×10^{-6}). The fidelity, cycle-consistency, and identity loss were monitored, and an early stopping strategy was employed to prevent overfitting. The training only involved bidirectional translation between B30 and D45 kernels. The network was implemented using PyTorch. Network training was conducted using a single-GPU workstation (Nvidia RTX 6000Ada).

F. Post-hoc, training-free calibration

As established in (6), the output image is determined by the interpolation between $\tilde{I}_{deblur}, \tilde{I}_{target}$ and $\tilde{I}_{over}$, governed by α. Therefore, any image within this interpolation manifold can be synthesized by adjusting α. Leveraging this property, our proposed network can be scaled to synthesize unseen kernels by tuning α. Ideally, the optimal α for an unseen kernel conversion can be identified through a post-hoc *calibration* using a single image pair. The Wasserstein distance between the MTFs of the synthesized and target images was utilized as the optimization criterion.

G. Experimental Evaluation Methods

For performance benchmarking, we compared our method against vanilla CycleGAN [6] and an AdaIN-CycleGAN. The latter incorporates AdaIN for style-conditioned translation, motivated by previous work [2]. Since target images inherently contain noise, conventional pixel-wise metrics such as MSE

TABLE I
QUANTITATIVE COMPARISON OF SSIM VALUES FOR WHOLE IMAGES AND ORGAN-SPECIFIC ROIS. BOLD INDICATES THE BEST PERFORMANCE.

	Total		Liver ROI		Lung ROI	
	D45	B30	D45	B30	D45	B30
CycleGAN	0.861±0.048	0.956±0.016	0.597±0.015	0.824±0.014	0.943±0.006	0.983±0.001
AdaIN-CycleGAN	0.940±0.017	0.974±0.010	0.811±0.066	0.915±0.010	0.948±0.007	0.984±0.003
Proposed *($p < 0.05$*)*	**0.949±0.013**	**0.986±0.005**	**0.896±0.010**	**0.972±0.004**	**0.968±0.003**	**0.995±0.001**

* Compared to AdaIN-CycleGAN.

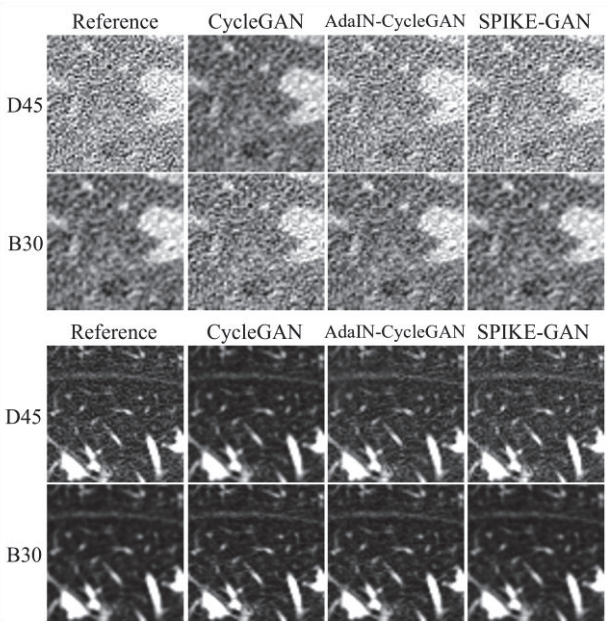


Fig. 2. Representative results of kernel conversion tasks showing (upper panel) a liver ROI and (lower panel) a lung ROI. Compared to the baseline models, our proposed method accurately reproduces the target kernel characteristics, such as sharper boundaries in D45 and smoother textures in B30. The ground truth images are displayed in the leftmost column.

and PSNR would over-penalize stochastic noise distribution. Therefore, structural similarity index measure (SSIM) and CT number accuracy served as the Figure of Merit (FoM) in quantitative evaluation. We evaluated SSIM across full FOV and 80×80 pixel regions of interest (ROIs) at livers and lungs. CT number accuracy was assessed by calculating the mean HU difference between synthesized and reference images using 16×16 pixel ROIs cropped from relatively uniform regions in left ventricles, livers, and spleens. Furthermore, patient-specific MTF curves were compared to evaluate the similarity in spatial frequency components. Mean-Absolute-Percentage-Error (MAPE) at 10% and 50% MTF was quantified. The observed performance difference was tested using Wilcoxon signed-rank test with 95% confidence level.

To demonstrate the scalability over unseen kernels, we defined four additional conversion tasks: B30→{B20, B40} and D45→{B40, B50}. A single patient case from validation set was designated as the calibration subject. The optimal α was determined through an exhaustive grid search within the range of [-1, 1] with a step size of 0.025. Once the optimal α is determined, it was applied to the five patients in the test set to demonstrate the scalability of our framework. Noise textural similarity was further validated in Fourier domain. ROIs (32×32 pixels) were cropped from a uniform region at spleen. After subtracting mean signal, Spectral Correlation Mapper (SCM)[7] was calculated to quantify spectral similarities, using

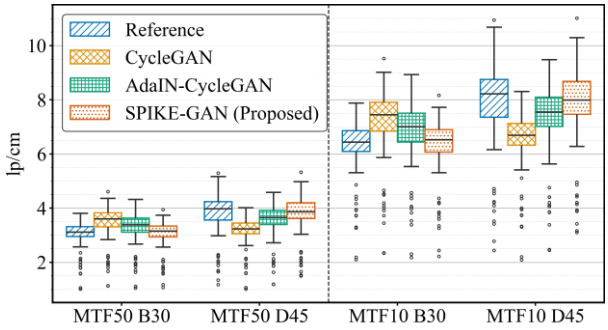


Fig. 3. Box-and-whisker plots of patient-specific Modulation Transfer Function (MTF) distributions for the kernel conversion tasks. MTF50: Spatial frequency at 50% MTF, MTF10: Spatial frequency at 10% MTF.

form radial mean profiles of 2D Fourier magnitude spectra.

III. RESULTS

A. Performance on Kernels Seen in Training

SPIKE-GAN exhibited improved performance compared to the baseline, successfully reproducing both the sharper and smoother appearance of D45 and B30, respectively (Fig. 2). The outputs of vanilla CycleGAN failed to replicate image characteristics in B30-to-D45 and D45-to-B30 tasks, showing obvious bias toward the input domain. As for AdaIN-CycleGAN, there existed noticeable blurring in certain structures (e.g., lung fissures) in B30-to-D45 task and more residual high-frequency textures in D45-to-B30 task.

In quantitative comparison, SPIKE-GAN achieved the highest SSIM across all scenarios (TABLE I, $p < 0.05$). Regarding CT number accuracy, both AdaIN-CycleGAN (< 5.22 HU for B30-to-D45; < 2.42 HU for D45-to-B30) and SPIKE-GAN (< 4.45 HU for B30-to-D45; < 5.24 HU for D45-to-B30) consistently satisfied the IAEA recommendation of 10 HU [8]. Although CycleGAN met this threshold for the D45-to-B30 task (< 2.23 HU), it failed to maintain accuracy in B30-to-D45 task, with a maximum HU bias of 18.5 HU.

The distributions of patient-specific MTFs were compared using box plots (Fig. 3). The overall distributions of the proposed method were highly consistent with those of reference, with no statistically significant difference for the D45-to-B30 task ($p > 0.05$). The B30-to-D45 task also exhibited strong agreement (MAPE 1.20%).

B. Scalability to Unseen Kernels

Across all tested unseen kernel conversions, SPIKE-GAN-synthesized images demonstrated a good alignment with the reference (Fig. 4). In quantitative evaluation, SPIKE-GAN achieved high structural similarity, yielding SSIM > 0.946, across all conversion tasks and ROIs (TABLE II). In MTF comparison, the MTFs from SPIKE-GAN align well with the reference ones across different unseen kernel conversion tasks (Fig. 5; MAPE at MTF10 < 2.20%).

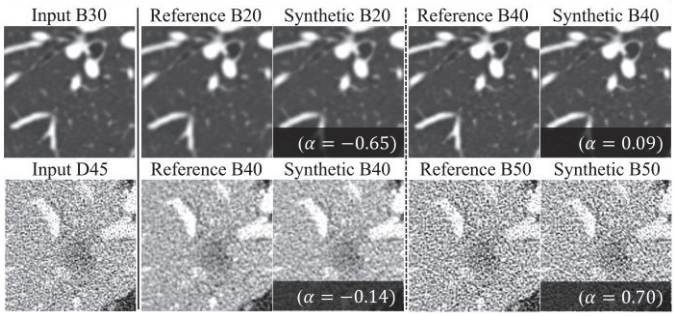


Fig. 4. Example images from unseen kernel conversion task through the calibration strategy. (Upper panel) B30-to-B20 and B30-to-B40 task, (lower panel) D45-to-B40 and D45-to-B50 task.

TABLE II
SSIM FOR UNSEEN KERNEL CONVERSION TASKS ACROSS DIFFERENT REGIONS OF INTEREST (ROIS).

	B30-to-B20	B30-to-B40	D45-to-B40	D45-to-B50
Total	0.988±0.004	0.981±0.005	0.984±0.008	0.969±0.011
Liver	0.958±0.021	0.950±0.027	0.946±0.007	0.969±0.005
Lung	0.995±0.004	0.990±0.009	0.993±0.002	0.975±0.005

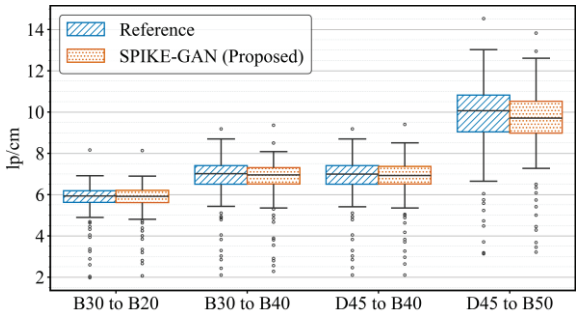


Fig. 5. Box-and-whisker plots illustrating the distributions of spatial frequency at 10% Modulation Transfer Function (MTF10) for various unseen kernel conversion tasks.

The synthesized Fourier magnitude spectra and the reference exhibit highly consistent spectral content (Fig. 6). Quantitative evaluation yielded SCM ≥ 0.922 across all test cases (TABLE III), indicating high similarity in noise texture.

IV. DISCUSSIONS

We developed and validated a deep learning framework for CT reconstruction kernel conversion that explicitly integrates kernel properties. We decoupled deblurring and target kernel imposition. By controlling the latter via parameter calibration, the proposed method achieves scalability without excessive retraining required by conventional approaches [3].

Unlike existing physics-informed models [4] that rely on implicit learning from auxiliary kernel inputs, SPIKE-GAN explicitly integrates analytic and interpretable operations, such as Wiener deconvolution and linear interpolation. These analytic modules ensure robustness across unseen kernels.

While this proof-of-concept study was limited to small cohort and two scanner models (SOMATOM Definition AS+ and Flash, Siemens), the generic framework is readily extensible to other scanners, including photon-counting-detector CT. Furthermore, although task-driven assessments were not performed, we conducted a systematic, quantitative evaluation across SSIM, HU bias, MTF and spectral analysis, where SPIKE-GAN exhibited performance comparable or superior to the baseline. Future studies are warranted to include cross-scanner validation, radiomics analyses, and evaluation of down-stream deep learning model performance.

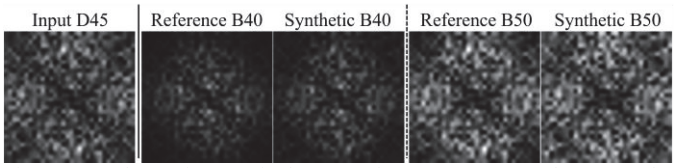


Fig. 6. Example Fourier magnitude spectra from D45-to-B40 and to-B50 tasks.

TABLE III
SPECTRAL CORRELATION MAPPER (SCM) ANALYSIS OF THE SPLEEN REGION.

B30-to-B20	B30-to-B40	D45-to-B40	D45-to-B50
0.974±0.013	0.922±0.043	0.989±0.008	0.981±0.016

V. CONCLUSION

We developed a Scalable Physics-informed Kernel-Explicit GAN (SPIKE-GAN) for CT kernel conversion. SPIKE-GAN is a two-phase, single generator CycleGAN that sequentially performed deblurring via a Wiener-deconvolution-integrated U-Net and kernel imposition through a learnable blending module. It outperforms the baseline methods in both qualitative and quantitative evaluation. Furthermore, SPIKE-GAN exhibited promising scalability over unseen reconstruction kernels without requesting re-training. In summary, our method has potential to facilitate large-scale data harmonization, thereby enhancing the consistency of quantitative analyses and the generalizability of deep learning models in various down-stream tasks, from longitudinal and multicenter studies to opportunistic CT applications.

ACKNOWLEDGMENT

This work was supported in part by the NIH under Grant **R01 EB036987** and **R01 EB017095**.

REFERENCES

[1] J. Choe *et al.*, "Deep Learning-based Image Conversion of CT Reconstruction Kernels Improves Radiomics Reproducibility for Pulmonary Nodules or Masses," *Radiology,* vol. 292, no. 2, pp. 365-373, Aug 2019, doi: 10.1148/radiol.2019181960.

[2] S. Yang, E. Y. Kim, and J. C. Ye, "Continuous Conversion of CT Kernel Using Switchable CycleGAN With AdaIN," *IEEE Trans Med Imaging,* vol. 40, no. 11, pp. 3015-3029, Nov 2021, doi: 10.1109/TMI.2021.3077615.

[3] A. R. Krishnan *et al.*, "Multipath cycleGAN for harmonization of paired and unpaired low-dose lung computed tomography reconstruction kernels," *Med Phys,* vol. 52, no. 11, p. e70120, Nov 2025, doi: 10.1002/mp.70120.

[4] M. Zarei, S. Sotoudeh-Paima, C. McCabe, E. Abadi, and E. Samei, "A Physics-Informed Deep Neural Network for Harmonization of CT Images," *IEEE Trans Biomed Eng,* vol. 71, no. 12, pp. 3494-3504, Dec 2024, doi: 10.1109/TBME.2024.3428399.

[5] C. H. McCollough *et al.*, "Low-dose CT for the detection and classification of metastatic liver lesions: Results of the 2016 Low Dose CT Grand Challenge," *Med Phys,* vol. 44, no. 10, pp. e339-e352, Oct 2017, doi: 10.1002/mp.12345.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223-2232.

[7] O. A. De Carvalho and P. R. Meneses, "Spectral correlation mapper (SCM): an improvement on the spectral angle mapper (SAM)," in *Summaries of the 9th JPL Airborne Earth Science Workshop, JPL Publication 00-18*, 2000, vol. 9: JPL publication Pasadena, CA, USA, p. 2.

[8] International Atomic Energy Agency, *Quality Assurance Programme for Computed Tomography: Diagnostic and Therapy Applications* (IAEA Human Health Series). Vienna: IAEA, 2012.

A practical model observer for a task-based image quality assessment based on a local hypothesis test

Okkyun Lee

***Abstract*—This study introduced a practical model observer that mimics the characteristics of human observers. We proposed a local hypothesis test under the assumption that the background image pixels are independent and identically distributed Gaussian random variables. It can be extended to a local likelihood ratio test by exploiting neighboring pixels. As a result, a local likelihood ratio map is obtained on which one can apply a threshold scan to calculate the ROC curve and the corresponding AUC value, even for a given single image. The simulation studies demonstrated that the proposed method is more stable than conventional model observers for task-based image quality assessment when a small number of sample images are provided.**

***Index Terms*—Image quality assessment, model observer, likelihood ratio test, receiver operating characteristic curve**

I. Introduction

TASK-based image quality assessment is a promising method for evaluating the performance of imaging systems and reconstruction methods, using mathematical model observers, and it can optimize system parameters and radiation exposures [1]. However, the conventional model observers require a sufficiently large number of images to train, i.e., to estimate the first- and/or second-order statistics of the data [1]. Furthermore, they often outperform human observers, so one needs to adjust several regularization parameters to match their performance [2, 3].

This study aims to develop a practical model observer that mimics characteristics of the human observer: (C1) The human observer decides the presence of lesions for a given image without knowledge of statistical information from a sufficient number of training images. It may be the main reason that the conventional model observer usually outperforms the human observers. (C2) Assuming a homogeneous background, a human observer approximates where the background is and how much the noise level is in there. The observer then scans for a potential signal, noting any difference between the *background* and the region under observation.

Here, we simplified the case to a homogeneous background and signal, corresponding to the usual CT test phantom used for CT image quality assessment [4]. We also assumed that the background image pixels are independent and identically distributed (i.i.d.) Gaussian random variables. The proposed model observer performs the following steps: (1) It assumes a local background region is known in a given image so that one can estimate the mean and variance of the background. This assumption is reasonable when the target is sparse relative to the background region. (2) For a given image pixel, it performs a likelihood ratio test, assuming the neighborhood pixels came from the same population as that of the center pixel. It performs the test throughout the image. (3) Under the null hypothesis, all the test statistics of the likelihood ratio asymptotically follow a chi-squared distribution with one degree of freedom. Thus, we apply the same threshold value to all the test statistics to determine the presence of a signal. (4) We calculate the receiver operating characteristic (ROC) curve by changing the threshold value, using the ground truth of the signal location, and then calculate the area under the curve (AUC).

The proposed model observer assesses a given image without training images and performs a local-region-based test, mimicking human observer properties in (C1) and (C2). We compared the proposed method with the conventional non-prewhitening matched filter (NPW) and the channelized Hotelling observer (CHO) using white Gaussian noise (WGN)-added images and CT images.

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00249223) and the InnoCORE program of the Ministry of Science and ICT(26-InnoCORE-02(2026040042)).

O. Lee is with the Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Republic of Korea (email: oklee@dgist.ac.kr).

II. Methods

A. Notations

We use $\mathbf{g} = [g_1, g_2, \cdots, g_n] \in \mathbb{R}^n$ to denote a given $N \times N$ image, represented as a column vector of size $n = N^2$. For the sake of simplicity, the symbol $\mathbf{g}$ refers to an image or column vector in this study. The specific meaning will be clear in the context. For the CHO, $\mathbf{g}$ denotes the dimension-reduced vector obtained by applying the channels to the image. One may treat $\mathbf{g}$ as a random vector and a given image as a realization from a population. The upper bar symbol (e.g., $\bar{\mathbf{g}}$) indicates the theoretical expectation (or sample mean of given images) of the population. H_0 is the null hypothesis that indicates the signal is absent, and H_1 is the alternative hypothesis that the signal is present.

B. Proposed Method

The main assumption is that the background elements in $\mathbf{g}$ are i.i.d. Gaussian random variables. Let the mean and variance for the background be μ_0 and σ_0^2, respectively, and assume that the variance of the signal is also σ_0^2. An example of the ground truth image ($\bar{\mathbf{g}}$) is given in Fig. 1, which consists of three signals, with low, middle, and high contrasts. Assume that a noisy image is given, see Fig. 1 (WGN 20dB). First, we select a background ROI to estimate sample mean ($\hat{\mu}_0$) and variance

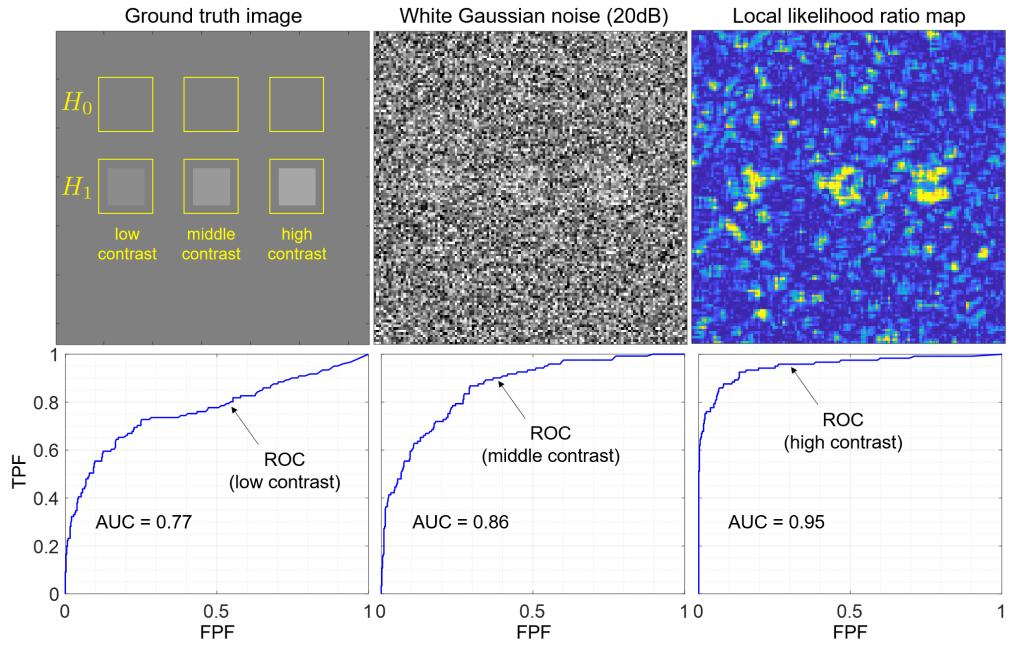


Fig. 1. An example of the local likelihood ratio map for the WGN case and the ROC curves for three different signals. A 5×5 window was used as $\mathcal{N}_j$. (WL/WW) [unitless]: (1.0/0.4) for the ground truth and WGN images and (3.5/7) for the local likelihood ratio map.

$(\hat{\sigma}_0^2)$, treating them as the ground truth population parameters. We used the yellow box above the signal as the ROI (H_0).

The proposed local hypothesis test for each pixel is given by,

$$H_0: \ g_j \in \text{background} \quad vs. \quad H_1: \ g_j \in \text{signal}, \tag{1}$$

which is equivalent to

$$H_0: \ \bar{g}_j = \hat{\mu}_0 \quad vs. \quad H_1: \ \bar{g}_j \neq \hat{\mu}_0. \tag{2}$$

We called it the local hypothesis test because it performs the test at each pixel in the image (for $j = 1, 2, \cdots, n$). Then, the associated log-likelihood ratio test is as follows:

$$\lambda(g_j) = \log \frac{\sup_{\theta \in \mathbb{R}} \mathcal{L}(\theta | g_j, \hat{\sigma}_0^2)}{\mathcal{L}(\theta = \hat{\mu}_0 | g_j, \hat{\sigma}_0^2)}, \tag{3}$$

where $\mathcal{L}(\theta|g_j, \hat{\sigma}_0^2)$ is the likelihood of the parameter θ for given values of g_j and $\hat{\sigma}_0^2$. Here, θ indicates the expected value of the population. The test in (3) is unstable due to insufficient sample data. Thus, under the main assumption, we generalized the local hypothesis test by exploiting the neighborhood pixels as follows:

$$\begin{aligned} \lambda_j(\mathbf{g}, \mathcal{N}_j) &= 2 \log \frac{\sup_{\theta \in \mathbb{R}} \mathcal{L}(\theta | \mathbf{g}, \mathcal{N}_j, \hat{\sigma}_0^2)}{\mathcal{L}(\theta = \hat{\mu}_0 | \mathbf{g}, \mathcal{N}_j, \hat{\sigma}_0^2)} \\ &= \sum_{i \in \mathcal{N}_j} \frac{(g_i - \hat{\mu}_0)^2}{\hat{\sigma}_0^2} - \sum_{i \in \mathcal{N}_j} \frac{\left(g_i - \hat{\mu}_j\right)^2}{\hat{\sigma}_0^2} \\ &\sim \chi_1^2, \end{aligned} \tag{4}$$

where j indicates the center pixel in the test, $\mathcal{N}_j$ is the indices for neighborhood pixels of g_j, and $\hat{\mu}_j$ is the sample mean in $\mathcal{N}_j$, the solution of $\sup_{\theta \in \mathbb{R}} \mathcal{L}(\theta|\mathbf{g}, \mathcal{N}_j, \hat{\sigma}_0^2)$. Note that we multiplied the factor 2 so that the test statistic $(\lambda_j(\mathbf{g}, \mathcal{N}_j))$ asymptotically follows the chi-squared distribution with one degree of freedom (the dimension of parameter search space; $\theta \in \mathbb{R}$) when H_0 is true [5].

One can use a rectangular window, such as 3×3 or 5×5, or an arbitrary shape for $\mathcal{N}_j$. An example of $\lambda_j(\mathbf{g}, \mathcal{N}_j)$ using a 5×5 window as $\mathcal{N}_j$ is shown in Fig. 1 (color map). As a summary index for the proposed model observer, we suggest calculating the ROC curve under the assumption that the signal location is known. Note that, when H_0 is true, the local test statistic in (4) asymptotically follows χ_1^2 regardless of pixel index. Thus, one can apply a threshold scan for the likelihood ratio map, $\left\{\lambda_j(\mathbf{g}, \mathcal{N}_j)\right\}_{j=1}^n$. For example, in Fig. 1 (ground truth; low contrast signal), one can calculate the true positive fraction (TPF) within the low contrast signal and the false positive fraction (FPF) within the yellow box above the signal (the ROI used to calculate background parameters). Examples of ROC curves and their corresponding AUC values are shown in Fig. 1 (second row).

C. Evaluations

The overall evaluation workflow for a single trial is illustrated in Fig. 2. We implemented the conventional NPW and CHO and compared them to the proposed method. For CHO, we used the Gabor filters [6] with 6 channels, as shown in Fig. 2. We used the same number of training and test images $(N_{train} = N_{test})$. Note that the train images are not necessary for the proposed method and that even a single test image is sufficient to calculate the AUC value. When multiple test images were used for the proposed method, we calculated

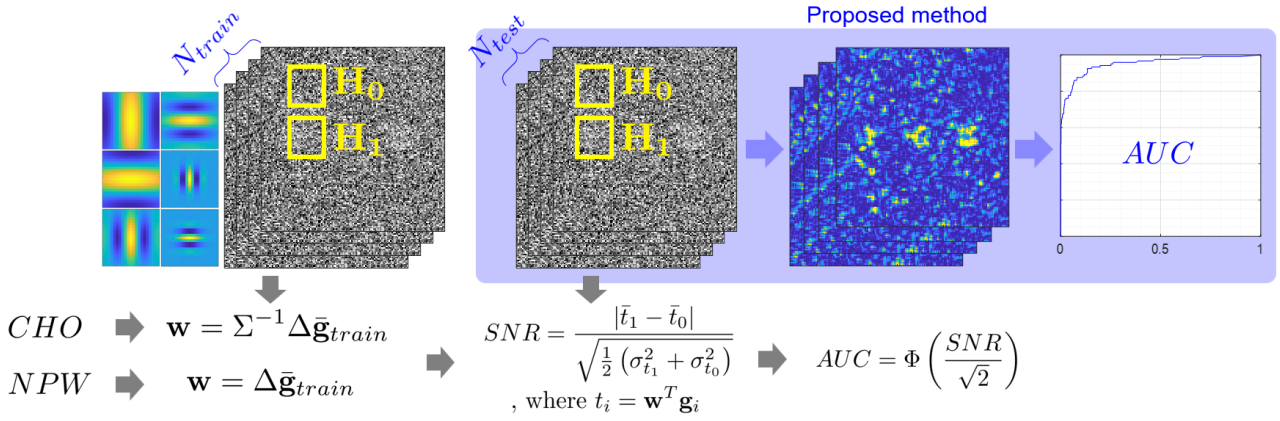


Fig. 2. The overall process of the evaluation for a single trial. It describes the train and test phases of the conventional NPW and CHO. The dimensions of $\mathbf{w}$ and $\mathbf{g}$ for NPW and CHO differ depending on channel usage. The proposed method only requires test images and works even with a single test image. The ROI images for H_0 and H_1 depict the low contrast case. It is straightforward to apply to other contrasts. We set $N_{train} = N_{test}$ and performed 100 trials for statistical analysis. (Σ: a covariance matrix, $\Delta\bar{\mathbf{g}} = \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_0$, Φ: a probability distribution function for a standard Gaussian random variable.)

the AUC for each image and then averaged them. We varied N_{test} from 1 to 20 and performed 100 trials per case for statistical analysis. We prepared 1000 training and test images and randomly sampled them for each evaluation.

We performed two studies. The first one is as shown in Fig. 1, the WGN (20dB) case with three different contrasts. It is the ideal case in which the i.i.d. assumption is valid. The second case is the CT image study, as shown in Fig. 3 with examples of the local likelihood map and corresponding ROC curves. We used a numerical PMMA phantom with three square signals with different densities. 2D fan-beam geometry was used with 500 detectors and 1000 views per rotation. 140kV with an aluminium prefilter (4mm) was used. The source-to-isocenter and isocenter-to-detector distances were 60cm and 40cm, respectively. The detector pitch was 0.1cm and the image pixel size was $0.06\,\text{cm} \times 0.06\,\text{cm}$. We used the ASTRA toolbox for implementing the fan-beam projection and filtered backprojection (FBP; Shepp-Logan filter) [7].

For both studies (WGN and CT image), the signal size was 15×15, and the two ROIs (H_0 and H_1) were both 21×21, on which all the methods were applied. A 7×7 window was used as $\mathcal{N}_j$ in the proposed method. The ROI (H_0) was used to calculate $\hat{\mu}_0$ and $\hat{\sigma}^2_0$ of the background.

III. Results

The main results of the calculated AUC for various methods are shown in Fig. 4. Overall, the standard deviation decreases while the average AUC increases as the contrast increases. In the WGN case with a low contrast signal, the proposed method exhibits the smallest standard deviation across the number of

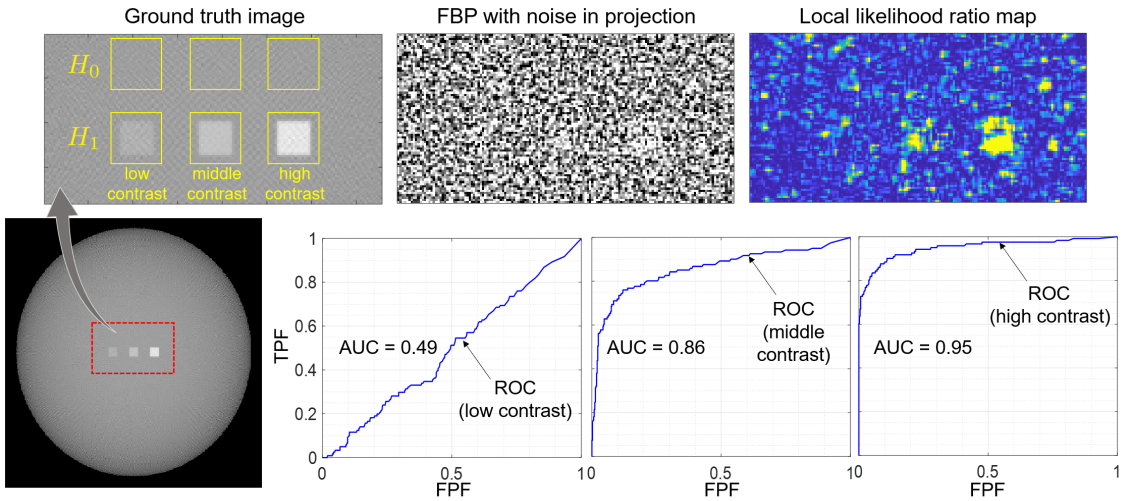


Fig. 3. An example of the local likelihood ratio map for the CT image case and the ROC curves for three different signals. A 5×5 window was used as $\mathcal{N}_j$. (WL/WW) [HU]: (125/70) for the ground truth and FBP images and (3.5/7) for the local likelihood ratio map. The differences between the signal and background are around 5, 10, and 20 [HU] for the low, middle, and high contrasts, respectively. The circular PMMA phantom has a diameter of 24cm, and a water beam hardening correction was applied. Each target size is $0.9 \times 0.9\,\text{cm}^2$.

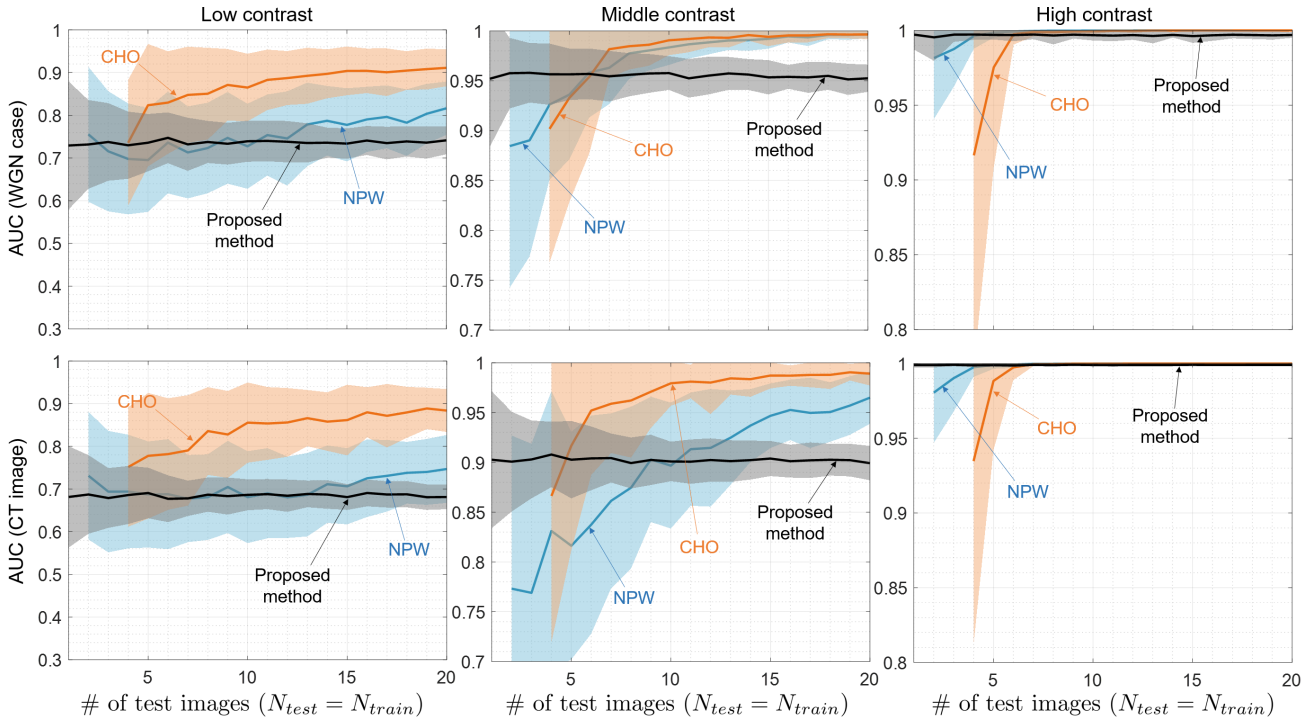


Fig. 4. Calculated AUC values along with the number of sample (test) images using various methods. A 7×7 window was used as $\mathcal{N}_j$. The shaded area indicates one standard deviation. (First row): WGN case and (second row): CT image case.

test images (x-axis). NPW needs at least 2 images to calculate the SNR, and CHO needs at least 4 images to calculate the inverse covariance matrix (no AUC values are available for fewer than the minimum required number of images). For the middle contrast signal, the proposed method exhibits the smallest deviation over the first 10 test images, and for the high contrast signal, roughly up to 7 images. The tendency is similar to the CT image case.

IV. Discussions and Conclusions

We have applied a 7×7 window for $\mathcal{N}_j$ during the evaluation in Fig. 4. We have also tested 3×3 and 5×5 windows and found that the overall mean AUC increases with increasing window size. However, the size or shape of $\mathcal{N}_j$ should be determined based on the signal size and the human observer's performance.

The potential benefits of the proposed model observer are as follows: (1) It presents more stable AUC values than conventional NPW and CHO when the sample images are small. (2) It shows consistent AUC values regardless of the number of sample images. (3) It has minimal hyperparameters to control the model performance: the size of the window for $\mathcal{N}_j$ (or the shape of the $\mathcal{N}_j$). (4) The i.i.d. condition in the background may not be valid in the CT image. However, the tendency was similar to that in the i.i.d. (WGN) case in Fig. 4, demonstrating its potential applicability in practice.

We are aware of robust model observers for small-sample images [8, 9]. However, the former uses prior knowledge of the signal in both training and test, and controls model performance by subsampling within the signal region, which may limit the applicable signal size. The latter assumes a radially symmetric homogeneous signal.

The limit of the proposed model is the assumption of a homogeneous background. Extending it to the inhomogeneous case and the comparison with human observer performance are left for future study.

References

[1] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. Hoboken, New Jersey: Wiley-Interscience, 2004.

[2] C. K. Abbey and H. H. Barrett, "Human- and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A*, vol. 18, no. 3, pp. 473–488, Mar 2001.

[3] L. Yu, B. Chen, J. M. Kofler, C. P. Favazza, S. Leng, M. A. Kupinski, and C. H. McCollough, "Correlation between a 2D channelized Hotelling observer and human observers in a low-contrast detection task with multislice reading in CT," *Med. Phys.*, vol. 44, no. 8, pp. 3990–3999, 2017.

[4] M. Fan, T. Thayib, C. McCollough, and L. Yu, "Accurate and efficient measurement of channelized Hotelling observer-based low-contrast detectability on the ACR CT accreditation phantom," *Med. Phys.*, vol. 50, no. 2, pp. 737–749, 2023.

[5] G. Casella and R. L. Berger, *Statistical Inference*, 2nd ed. Pacific Grove, CA, USA: Duxbury, 2001.

[6] A. Wunderlich, F. Noo, B. D. Gallas, and M. E. Heilbrun, "Exact confidence intervals for channelized Hotelling observer performance in image quality studies," *IEEE Trans. Med. Imag.*, vol. 34, no. 2, pp. 453–464, 2015.

[7] W. van Aarle *et al.*, "Fast and flexible x-ray tomography using the ASTRA toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, October 2016.

[8] M. Anton, A. Khanin, T. Kretz, M. Reginatto, and C. Elster, "A simple parametric model observer for quality assurance in computer tomography," *Phys. Med. Biol.*, vol. 63, no. 7, p. 075011, mar 2018.

[9] M. Anton, W. J. H. Veldkamp, I. Hernandez-Giron, and C. Elster, "RDI-a regression detectability index for quality assurance in: x-ray imaging," *Phys. Med. Biol.*, vol. 65, no. 8, p. 085017, apr 2020.

GrainNet: An Unsupervised Framework for Frequency-Dependent Higher-Order Noise Modeling in CT

Ou Li, Kirsten Boedeker, Daniel Shin, Tomoe Hagio, Koen Michielsen, Luuk J. Oostveen, and Ioannis Sechopoulos

***Abstract*—Visual appearance in CT images is strongly influenced by noise texture, yet most noise simulation pipelines address only the second-order statistics such as variance. While global kurtosis, a fourth-order statistic, has been proposed as a descriptor of non-Gaussian noise behavior, it does not capture its frequency-dependent nature observed in measured CT data. In preparation for an observer study, this work proposes GrainNet, an unsupervised framework for injecting frequency-dependent kurtosis noise statistics into CT images. GrainNet works by modulating band-limited stochastic residuals with a learned spatial magnitude field, with higher-order statistics evaluated using frequency- and orientation-selective Gabor filters. The proposed method is unsupervised, driven by statistics from measured CT images without needing paired training data. Results demonstrated visually realistic noise texture beyond conventional NPS-matched models while statistically matching frequency-dependent non-Gaussian noise characteristics. GrainNet was useful in solving an ill-posed inverse problem and provides a practical and physically motivated tool for CT observer studies, where systematic variation of higher-order noise statistics may be required to assess perceptual and task-based image quality.**

***Index Terms*—noise texture, computed tomography, perception, kurtosis, inverse problem.**

I. Introduction

THE visual appearance of medical images is influenced by both noise magnitude and noise texture. Both attributes can affect the detectability of small and low-contrast lesions [4]. Many factors affect noise magnitude; however, the main factor affecting noise texture in CT imaging is the reconstruction method and additional denoising techniques. This includes for example the reconstruction kernel in FBP or regularization methods such as total variation minimization which is often combined with iterative CT reconstruction methods such as model-based iterative reconstruction (MBIR) and can cause reconstructed images to appear plasticky or blotchy for radiologists when not properly tuned [10], [11], [13], [15].

Noise texture in CT images is commonly characterized by the noise power spectrum (NPS), which describes the distribution of variance across spatial frequencies. [12]. However, NPS is limited to variance information and is often summarized by scalar descriptors such as the peak or average frequency, which do not uniquely define the full spectral shape. As demonstrated by Oostveen et al., multiple NPS shapes with identical peak or average frequency can correspond to perceptually distinct noise textures, indicating that higher-dimensional characteristics of the NPS shape are required to capture visual appearance [2], [10]. The kurtosis of a distribution, a fourth order statistic, was proposed to address this limitation, and is able to show deviations from Gaussian distribution [14]. Even so, kurtosis is an aggregated measure, which prevents it from capturing "Non-Gaussianess" within the various noise frequencies. To amend the above problem, frequency-dependent kurtosis was proposed to quantify kurtosis as a function of spatial frequency. Abbey et al. demonstrated that the kurtosis of Gabor filter response in breast CT varies strongly with spatial frequency, suggesting that this higher-order statistics is distributed non-uniformly across frequency bands and maybe linked to visual sensitivity [1]. Findings presented in Abbey et al. motivated the use of the Gabor filter, for realistic CT noise modeling, especially in the context of observer studies.

Despite recognition of the importance of higher-order and frequency-dependent noise statistics influencing image texture and observer performance, no framework exists for simulating such noise properties in a controlled manner. Prior studies have analyzed frequency-dependent kurtosis in medical images; however, these analyses have not translated into an injection-based modeling framework for observer studies. Moreover, synthesizing noise with prescribed higher-order statistics is an ill-posed inverse problem. Many stochastic fields can satisfy the same frequency-resolved statistical targets, while only a small subset produce visually plausible textures. This non-uniqueness is particularly visible when higher order statistics are defined as the Gabor response, making noise synthesis challenging.

O. Li, K. Michielsen, L.J. Oostveen, and I. Sechopoulos are with the Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands.

K. Boedeker, D. Shin, and T. Hagio are with Canon Medical Systems Corporation, Los Angeles, California, United States

I. Sechopoulos is with the Dutch Expert Centre for Screening (LRCB), Nijmegen, The Netherlands.

I. Sechopoulos is with the Technical Medicine Centre, University of Twente, Enschede, The Netherlands, e-mail: Ioannis.sechopoulos@radboudumc.nl.

To address this problem, this work proposes GrainNet, an unsupervised framework for injecting frequency-dependent higher-order noise statistics into white noise images under additional physical and perceptual constraints, enabling realistic noise generation.

II. Methods and Materials

A. Gabor filter

Image statistics were analyzed using a bank of 2D Gabor filters. A Gabor filter is a band-pass linear filter obtained by modulating a sinusoidal carrier with a localized Gaussian envelope. This approach allows joint localization in spatial and frequency domains. For image $g(\boldsymbol{r})$, a 2D Gabor filter at spatial frequency f_0 and orientation θ can be expressed as

$$h_{f_0,\theta} = \exp\left(-\frac{||r||^2}{2\sigma^2}\right)\sin(2\pi f_0 \boldsymbol{u}_{\boldsymbol{\theta}}^{\top}\boldsymbol{r}) \tag{1}$$

where $\boldsymbol{u}_\theta = (\cos(\theta), \sin(\theta))$ and σ, the envelope width, controls the filter bandwidth [5]. In the experiment with octave-bandwidth filters, σ is inversely proportional to f_0, which yields approximately constant relative bandwidth across scales [6]. As a result, filters are simultaneously localized in space and frequency [7].

To obtain the Gabor response in the image domain, the input image g is convolved with a Gabor kernel $h_{f_0,\theta}$, yielding

$$x_{f_0,\theta}(\boldsymbol{r}) = (g * h_{f_0,\theta})(\boldsymbol{r}), \tag{2}$$

where $x_{f_0,\theta}(\boldsymbol{r})$ encodes the spatially localized response of the image content around the center frequency f_0 and orientation θ. The statistical distribution $x_{f_0,\theta}$ is estimated by sampling over spatial locations within a region of interest (ROI) [6].

While the second-order statistics of $x_{f_0,\theta}$ are fully determined by the NPS, higher-order moments depend on statistical information beyond variance. The kurtosis of the response distribution

$$\gamma_{f_0,\theta} = \frac{E[(x_{f_0,\theta} - \mu)^4]}{\sigma^4} \tag{3}$$

quantifies the deviations from normality within a localized frequency band, measuring how much the distribution shifts towards the tail [6].

Kurtosis measured in the Gabor domain corresponds to a frequency-localized projection of trispectrum [8], [9], a fourth-order structure. However, this mapping is many-to-one, many stochastic fields may satisfy identical frequency-dependent Gabor-domain kurtosis constraints while differing substantially in visual appearance, rendering noise synthesis under such constraints an ill-posed inverse problem.

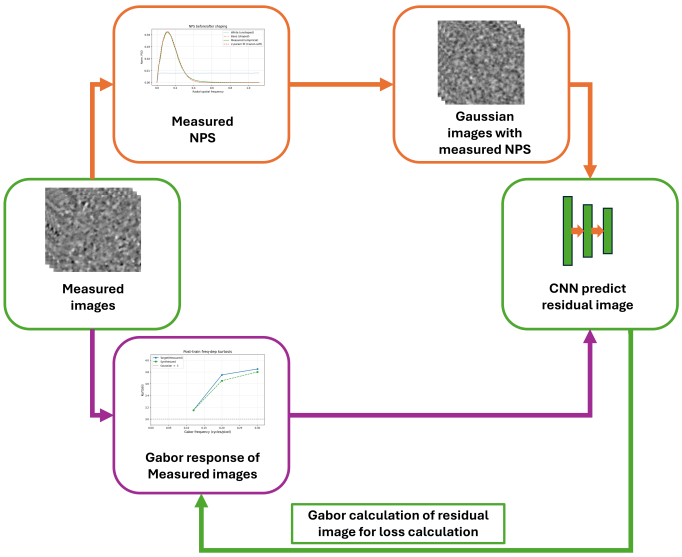


Fig. 1: Diagram of the image synthesis pipeline. Orange path: Calculate and apply measured NPS onto images with Gaussian white-noise for input into CNN. Purple path: Calculate Gabor response corresponding to the measured images for input into CNN. Green path: Calculate loss from residual image predicted by CNN for optimization. (e.g. Kurtosis loss calculated by comparing Gabor response of synthesized image with that of measured image)

B. Neural network

GrainNet is a lightweight convolutional model that predicts a noise magnitude field from an NPS-matched base image rather than directly synthesizing a noise texture.The architecture contains three convolutional layers and LeakyReLU activations followed by a linear output layer.

The base image b is generated by shaping Gaussian white noise with an NPS from reconstructed measured images.

Given base image $b(x, y)$ the network produces

$$m(x, y) = \mathcal{K}(b; \theta) \tag{4}$$

where $\mathcal{K}$ is a learned mapping with parameter θ. Furthermore, m gates a Gaussian noise field $\eta(x, y) \sim \mathcal{N}(0, 1)$ to form $r(x, y) = m(x, y)\eta(x, y)$, which is then added to the base Gaussian image. Then the Gabor response of the residual image is calculated for comparison with the Gabor response of the measured image, forming the main component of the optimization loss function.

C. Training and losses

The training of GrainNet is formulated as a constrained optimization problem that addresses the intrinsically ill-posed nature of synthesizing visually plausible CT noise textures

under higher-order statistical constraints. Because frequency-dependent kurtosis alone does not uniquely determine a noise realization, GrainNet is trained in an unsupervised manner using a composite loss function in which multiple objectives act in a competitive, counterbalancing fashion to restrict the solution space to physically and perceptually valid outcomes.

The primary objective enforces agreement between the orientation-resolved Gabor responses of the synthesized images and the target frequency-dependent kurtosis. Synthesized images are passed through a fixed multi-frequency, multi-orientation Gabor filter bank, and the loss penalizes deviations between the per-frequency, per-orientation kurtosis and the corresponding measured targets. When applied in isolation, this constraint generates solutions that matches measured Gabor responses numerically while producing visually implausible artifacts such as streaks and worm-like structures.

To suppress such erroneous solutions, a set of auxiliary losses imposes complementary constraints that explicitly compete with the primary kurtosis objective. The full training objective is given by

$$\mathcal{L}_{\text{total}} = \lambda_{\text{kurt}}\,\mathcal{L}_{\text{kurt}} + \lambda_{\text{iso}}\,\mathcal{L}_{\text{iso}} + \lambda_{\text{coh}}\,\mathcal{L}_{\text{coh}} + \lambda_{\text{res}}\,\mathcal{L}_{\text{res}} + \lambda_{\text{dc}}\,\mathcal{L}_{\text{dc}} + \lambda_{\text{var}}\,\mathcal{L}_{\text{var}} + \lambda_{\text{lowf}}\,\mathcal{L}_{\text{lowf}} + \lambda_{\text{tv}}\,\mathcal{L}_{\text{tv}}, \tag{5}$$

where each term is further defined below. The loss weights were empirically set as follows: $[\lambda_{\text{res}}, \lambda_{\text{dc}}, \lambda_{\text{var}}, \lambda_{\text{lowf}}, \lambda_{\text{tv}}] = [2.0, 0.25, 0.015, 10^{-4}, 0.10, 0.15, 0.10]]$

1) Data terms: Let x_s denote the synthesized noise image and let $R_{f,o}(x_s)$ denote the response of the Gabor filter at frequency f and orientation o. The primary kurtosis loss is

$$\mathcal{L}_{\text{kurt}} = E_{f,o}\left[\left(\kappa(R_{f,o}(x_s)) - \kappa_f^*\right)^2\right], \tag{6}$$

where $\kappa(\cdot)$ denotes sample kurtosis and κ_f^* is the measured target kurtosis at frequency f, replicated across orientations.

The isotropy penalty suppresses orientation-dependent variation in the Gabor-domain kurtosis,

$$\mathcal{L}_{\text{iso}} = E_f\left[\text{Var}_o\big(\kappa(R_{f,o}(x_s))\big)\right]. \tag{7}$$

2) Stabilizing terms: To discourage spatially organized or edge-like residual structures, a coherence penalty is applied to the residual noise using a structure-tensor–based anisotropy measure,

$$\mathcal{L}_{\text{coh}} = E\left[\frac{\sqrt{(J_{xx} - J_{yy})^2 + 4J_{xy}^2}}{J_{xx} + J_{yy} + \epsilon}\right], \tag{8}$$

where J_{xx}, J_{yy}, and J_{xy} are components of the local structure tensor computed from spatial gradients of the residual image.

Additional stabilizing terms include a residual-energy penalty

$$\mathcal{L}_{\text{res}} = E\left[\|r\|_2^2\right], \tag{9}$$

a DC suppression term

$$\mathcal{L}_{\text{dc}} = E\left[\mu(r)^2\right], \tag{10}$$

and a variance-matching penalty

$$\mathcal{L}_{\text{var}} = E\Big[\big(\text{Var}(x_s) - \text{Var}(x_b)\big)^2\Big]. \tag{11}$$

To prevent excessive low-frequency residual structure and overly sharp modulation fields, a low-frequency residual penalty and a total-variation penalty on the learned noise-amplitude map m are applied,

$$\mathcal{L}_{\text{lowf}} = E\left[\sum_{k\in\mathcal{K}_{\text{low}}} \text{PSD}_{\text{rad}}^{(r)}(k)\right], \qquad \mathcal{L}_{\text{tv}} = E[\text{TV}(m)]. \tag{12}$$

As training progresses, the auxiliary losses increasingly oppose solutions that exploit the non-uniqueness of the higher-order kurtosis constraint. Convergence comes from a dynamic equilibrium of competing objectives, yielding synthesized noise that is statistically faithful to measured frequency-dependent kurtosis while remaining visually plausible and free of structured artifacts.

For evaluation and observer-study preparation, batches of 32 images were synthesized per condition. The initial target was set to the measured frequency dependent kurtosis, with additional simulation targets set by scaling the discrete frequency points of the measured response by $\pm 5\%$, enabling systematic manipulation of higher-order noise statistics while preserving lower-order spectral characteristics.

III. Results

A. Gabor response

Fig. 2 show the Gabor response for the target CT noise, synthesized noise, and Gaussian reference noise. When trained to match the measured target, the synthesized kurtosis follows the measured frequency-dependent trend with residual deviations around 3.5% across all spatial frequency bands. When the target kurtosis is increased by 5%, the synthesized curves consistently track this increase, yielding kurtosis values approximately 4–6% higher than the measured reference while preserving the same frequency dependence. The relative deviation between the measured and the target curve is never more than 4% and in most cases below 2%.

B. Synthesized images

Compared to measured CT noise, the synthesized images show closely matching texture characteristics, including similar granularity and spatial homogeneity, without visible bias toward smoothness or artificial structure. NPS-matched Gaussian noise appears visually more uniform and lacks the black and white splotches present in both measured and synthesized images, consistent with its near-Gaussian kurtosis.

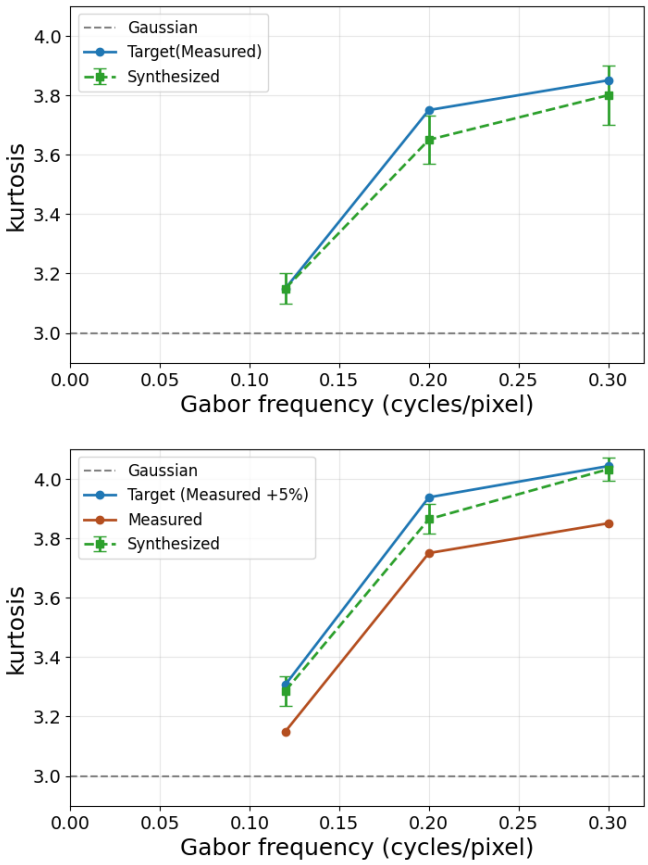
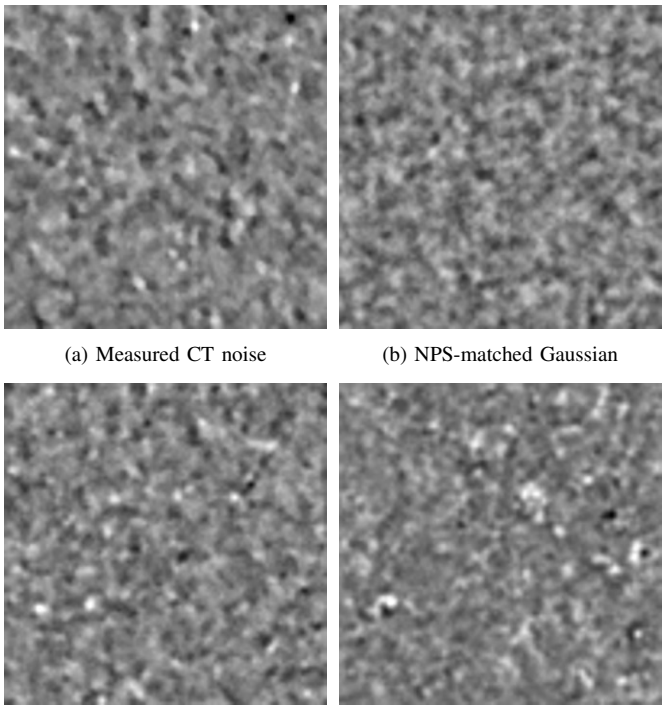


Fig. 2: Top: plot of Gabor responses of measured and synethesized images post-training. Bottom: plot of Gabor responses of measured images entirely moved up by 5% and synthesized images trained to match it.



(a) Measured CT noise (b) NPS-matched Gaussian

(c) Synthesized: Gabor target (d) Synthesized: Gabor target +5%

Fig. 3: Examples: (a) measured CT noise, (b) NPS-matched Gaussian noise, (c) synthesized noise targeting measured Gabor response, (d) synthesized noise targetting measured Gabor response increased by 5%.

As expected, increases in kurtosis (fig.3) correspond to a higher prevalence of localized deviations, which appear visually as a greater density of small, high-contrast black and white spots relative to Gaussian noise. These changes occur without introducing spatial coherence, banding, or directional artifacts, due to the fact that the network does not generate structure directly but instead modulates the amplitude of stochastic noise in a controlled manner.

IV. Discussion

The results demonstrate the GrainNet was able to effectively inject frequency-dependent higher-order structure while visually preserving second-order structure and visual plausibility.

The observed behaviors of the synthesized images are a direct result of the competitive regularization framework used during training. Higher-order statistical objectives promote high-amplitude fluctuations that manifest visually as localized black and white blotches, while auxiliary constraints—such as isotropy, coherence suppression, and smoothness of the noise magnitude field—prevent these fluctuations from causing structured or visually implausible patterns such as worm-like artifacts and streaks. Additionally, the shallow network design enhances interpretability, as each predicted output directly corresponds to a local modulation of noise variance rather than an implicit or learned texture representation. By enabling controlled, bounded modulation of non-Gaussian noise characteristics without introducing visual artifacts, the proposed approach establishes a practical foundation for systematic observer studies of noise texture and diagnostic performance.

In the current design, the multiple stabilizing terms in the loss function were added sequentially during design as problems were found. It is likely that some terms are mutually redundant and we are looking into simplifying this aspect of the network. Future work will leverage GrainNet for controlled observer studies designed to isolate the perceptual and diagnostic impact of frequency-dependent higher-order noise statistics.

V. Conclusion

GrainNet injects frequency-dependent higher-order noise statistics into CT noise fields while preserving the NPS. Rather than a complete model of nonlinear reconstructed

noise, it serves as a practical synthesis framework for observer-study stimuli with prescribed non-Gaussian texture properties.

Here, injection refers to controlled noise-texture synthesis, allowing higher-order statistics to be varied under constrained and visually plausible conditions. This supports future observer studies on the perceptual relevance of frequency-dependent kurtosis for noise texture, perceived sharpness, and diagnostic task performance.

REFERENCES

[1] C. K. Abbey, J. N. Sohl-Dickstein, B. A. Olshausen, M. P. Eckstein, and J. M. Boone, "Higher-order scene statistics of breast images," Proc. SPIE 7263, Medical Imaging: Image Perception, Observer Performance, and Technology Assessment, 726317 (2009).

[2] K. L. Boedeker et al., "Impact of non-Gaussian noise properties on CT noise texture," Med. Phys. 49(6), 4054–4066 (2022).

[3] D. R. Brillinger, *Time Series: Data Analysis and Theory*, 2nd ed., Holden-Day, San Francisco (1981).

[4] J. T. Bushberg et al., "Noise and noise texture," in The Essential Physics of Medical Imaging. 4th ed., S. Zinner and E. McDermott, Eds., pp. 89–97, Wolters Kluwer (2020).

[5] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," J. Opt. Soc. Am. A 2(7), 1160–1169 (1985).

[6] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," J. Opt. Soc. Am. A 4(12), 2379–2394 (1987).

[7] B. Julesz, "Textons, the elements of texture perception, and their interactions," Nature 290, 91–97 (1981).

[8] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory," Proc. IEEE 79(3), 278–305 (1991).

[9] C. L. Nikias and J. M. Mendel, "Signal processing with higher-order spectra," IEEE Signal Process. Mag. 10(3), 10–37 (1993).

[10] L. J. Oostveen, K. Boedeker, D. Shin, C. K. Abbey, and I. Sechopoulos, "Perceptual thresholds for differences in CT noise texture," Journal of Medical Imaging, vol. 11, no. 3, p. 035501, May/Jun. 2024, doi: 10.1117/1.JMI.11.3.035501.

[11] G. Saiprasad et al., "Evaluation of low-contrast detectability of iterative reconstruction across multiple institutions, CT scanner manufacturers, and radiation exposure levels," Radiology 277(1), 124–133 (2015).

[12] E. Samei and J. Solomon, "The noise power spectrum in CT," Med. Phys. 41(9), 091908 (2014).

[13] W. Stiller, "Basics of iterative reconstruction methods in computed tomography: a vendor-independent overview," Eur. J. Radiol. 109(Aug.), 147–154 (2018).

[14] G. Vegas-Sánchez-Ferrero et al., "Statistical characterization of noise in CT images," Med. Image Anal. 40, 44–59 (2017).

[15] M. J. Willemink et al., "Iterative reconstruction techniques for computed tomography. Part 1: Technical principles," Eur. Radiol. 23(6), 1623–1631 (2013).

CT Reconstruction from Dual-View Depth Images and Topograms for Prospective Risk Minimization

Aniol Serra Juhé*, Daniel Mosig*, Chang Liu, Philipp Oelze, Michael Lell, Marc Kachelrieß, and Andreas Maier

***Abstract*—Prospective radiation risk minimization in computed tomography (CT) requires a coarse patient-specific 3D reconstruction prior to full acquisition. Existing deep learning-based approaches typically rely on two orthogonal topograms, which may not always be available in clinical workflows. At the same time, modern CT systems are often equipped with depth cameras mounted above the patient table. In this work, we investigate whether a depth image can replace one of the two standard topograms for CT volume reconstruction. We propose a generative adversarial network (GAN) that reconstructs a 3D CT volume from a lateral X-ray projection and an anterior–posterior (AP) depth image. Due to the lack of paired clinical data, topograms and depth images were synthetically generated from CT volumes of the LIDC-IDRI dataset using a realistic fan-beam geometry. The proposed multimodal input configuration is evaluated against biplanar X-ray projections, a single X-ray projection and depth-only variants using RMSE, PSNR, and SSIM. While biplanar X-ray input yields the highest reconstruction accuracy, the combination of a lateral X-ray and an AP depth image achieves comparable quantitative performance. Qualitative results show anatomically plausible reconstructions but limited soft-tissue contrast and blurred organ boundaries. Overall, the results suggest that depth images can provide useful geometric information, although they do not fully compensate for missing radiographic views.**

I. Introduction

COMPUTED Tomography (CT) is a non-invasive imaging modality used to visualize internal structures of the human body. Due to its short acquisition time and high image quality, CT is routinely employed in clinical practice. However, CT involves exposure of the patient to ionizing radiation during projection data acquisition for three-dimensional volume reconstruction. Consequently, radiation dose reduction is an important area of ongoing research. Several approaches are used to reduce patient dose including pre-filters, iterative reconstruction algorithms, shielding techniques, and tube current modulation (TCM) [1], [2]. In TCM, the X-ray tube current is modulated depending on the angular and z-position of the tube with respect to the patient to reduce the mAs-product while maintaining image quality. Not all the organs have the same sensitivity to radiation. The effective dose is defined as the sum of the dose absorbed by the organ-at-risks (OAR) during the exposure, weighted with the organ-specific tissue factor [3], [4]. RiskTCM [5] is a method that accounts for organ-specific radiation sensitivities to mitigate overall patient risk. It assumes the availability of a prospective coarse CT reconstruction, organ segmentation and dose distribution estimation. Therefore, a coarse CT reconstruction is needed prior to the full scan. Several deep learning methods have been studied to achieve this using only a limited number of projections.

Generative adversarial network (GAN) [6] methods have successfully proven their capability for CT reconstruction using only few views. X2CT-GAN [7] produces a 3D volume from two orthogonal X-ray images using a generator that increases the data dimension and a feature fusion approach to combine information of both projections. Building on X2CT-GAN, CVAE-GAN [8] introduced a GAN architecture augmented with a conditional variational autoencoder to improve generator regularization and to reconstruct a 3D volume from a single X-ray projection. Furthermore, Liu et al. [9] expanded X2CT-GAN by introducing a 3D perceptual loss and a pre-trained organ segmentation network during the training that improve anatomical information. OX2CT-GAN [10] incorporates Feature Enhancement Blocks to better capture critical anatomical information. Moreover, PerX2CT [11] accounts for the projection perspective by generating coordinate-specific feature combinations. In addition to GANs, diffusion processes have also been explored. DiffuX2CT [12] models CT reconstruction from orthogonal biplanar X-rays employing a 3D globally coherent denoising model with an implicit conditioning strategy.

In routine clinical workflows, one or two (anterior-posterior AP and lateral) CT localizer radiographs - also known as topograms or scout views - are typically acquired prior to the CT scan. Existing reconstruction methods are generally designed for scenarios involving two X-ray projections. Generally, using two X-ray projections leads to better reconstruction results than using a single view, as mentioned by the authors of X2CT [7]. However, in some cases, only a single topogram may be available. Meanwhile, the use of depth cameras in CT reconstruction has been investigated, with motion and overexposure artifact mitigation being the primary focus for C-arms systems [13]–[15]. Today, modern CT systems are also increasingly equipped with depth cameras mounted above the patient table. Motivated by this, we propose replacing one of the topograms, the AP view, with a depth image to enable accurate 3D volume reconstruction. Due to the lack of publicly

A. Serra Juhé, D. Mosig and A. Maier are with the Pattern Recognition Lab, Department Computer Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany.

C. Liu is with the Erlangen National High Performance Computing Center (NHR@FAU), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany.

P. Oelze and M. Kachelrieß are with the Division of X-Ray Imaging and Computed Tomography, German Cancer Research Center (DKFZ), Heidelberg, Germany.

P. Oelze and M. Kachelrieß are with the Ruprecht-Karls-University, Heidelberg, Germany.

M. Lell is with the Department of Radiology and Nuclear Medicine, Klinikum Nürnberg, Paracelsus Medical University, Nuremberg, Germany.

Corresponding authors: A. Serra Juhé and D. Mosig (e-mail: aniol.serra@fau.de and daniel.mosig@fau.de)

*These authors contributed equally to this work.

available datasets containing paired depth images, topograms, and CT volumes, the model was trained and evaluated using a public CT dataset. This is a common practice when paired real data are unavailable or when system validation is required [16]. The corresponding topograms and depth images were generated by simulating the acquisition process from the CT volumes.

II. Methods and Materials

The proposed model is a GAN that takes an X-ray projection and a depth image as input to reconstruct a 3D volume. Specifically, the X-ray input corresponds to a lateral topogram, while the depth image is acquired from above the patient. This provides complementary information from two orthogonal planes. Model training follows the standard adversarial learning strategy used in GANs with the following minmax objective

$$\min_G \max_D \quad \mathcal{L}_{\text{GAN}} = \mathbb{E}_y[\log D(y)] + \mathbb{E}_x[\log(1 - D(G(x)))] \tag{1}$$

x denotes the input X-ray/depth images, y denotes the corresponding ground-truth CT volume. The generator and discriminator are indicated as $G(x;\theta_g)$ and $D(y;\theta_d)$. $\mathbb{E}$ denotes the mean function. The loss functions used to optimize the discriminator and the generator are defined as follows:

$$\mathcal{L}_{\text{D}}(x,y) = \frac{1}{2}\left(\mathbb{E}_y\Big[\|D(y)-1\|_2^2\Big] + \mathbb{E}_x\Big[\|D(G(x))-1\|_2^2\Big]\right) \tag{2}$$

$$\mathcal{L}_{\text{G}}(x) = \mathbb{E}_x\Big[\|D(G(x))-1\|_2^2\Big] \tag{3}$$

Building on the work of Liu et al. [9], the discriminator adopts the architecture proposed by Isola et al. [17], while the generator is derived from the X2CT-GAN framework introduced by Ying et al. [7]. The generator processes the two input inputs, a lateral X-ray and an anterior-posterior (AP) depth image, through two separate U-Net–based encoding branches. The resulting 3D feature representations from the two branches are combined by addition at each upsampling stage, and a subsequent 3D decoding network integrates the fused features to generate the final reconstructed CT volume [9]. A visualization of the generator pipeline is provided in Fig. 1.

A. Dataset

The dataset used for training and evaluation of the proposed model is the Lung Image Database Consortium and Image Database Resource Initiative (LIDC-IDRI) dataset [18]. The LIDC-IDRI dataset consists of 1010 anonymized clinical thoracic CT scans acquired from multiple institutions under standardized imaging protocols. Their spatial resolution ranges from 0.46 to 0.98 mm in the axial plane, and from 0.6 to 5.0 mm in the z-direction. The volumes were resampled to a voxel size of 2.5 mm in all axes and then cropped to a

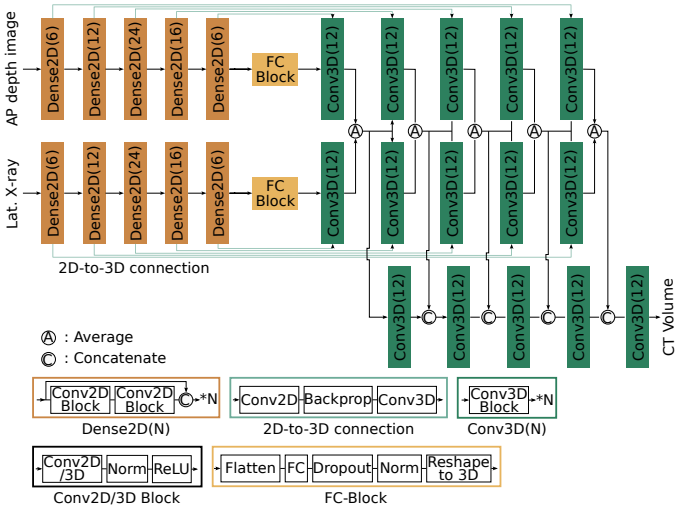


Fig. 1: The generator used in our experiments as proposed in [7]. Two 2D input images, a lateral X-ray and an AP depth image, are reconstructed to a 3D CT volume. Figure adapted from Liu et al. [9], CC BY 4.0.

volume size of 128. Following the work by Liu et al., the X-ray projections are synthetically generated to approximate the fan-beam CT forward projection. The projection geometry was configured to match the specifications of a Siemens SOMATOM Force scanner, with a source-to-detector distance of 1085.6 mm, a source-to-isocenter distance of 595 mm, and 920 rays within the fan. Similarly, depth images were generated using the same fan-beam geometry as the CT. For each detector element, the surface location was estimated by identifying the first voxel along the projection ray whose attenuation exceeded a predefined high-density criterion. This ensures perfect alignment of the X-ray projections and depth images. An example of these synthetically generated images is visualized in Fig. 2. We split the resulting dataset into training, validation, and test subsets using a 70/20/10 ratio.

B. Experimental setup

We establish a baseline experiment using the lat. X-ray/AP X-ray input combination and benchmark different input combinations against it. The GAN is trained for 100 epochs for each experiment. Generator and discriminator are each optimized using an Adam optimizer with a learning rate of $2 * 10^{-4}$. We utilize a single NVIDIA V100 GPU and employ a batch size of 1. No cross-validation was used due to computational demand and the large training set size. We evaluate the results using root mean squared error (RMSE) in Hounsfield units (HU), peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM).

III. Results & Discussion

Four different input configurations were evaluated for the proposed model. These configurations included: two biplanar X-ray projections; a lateral X-ray projection combined with an AP depth image; a lateral X-ray projection replicated across

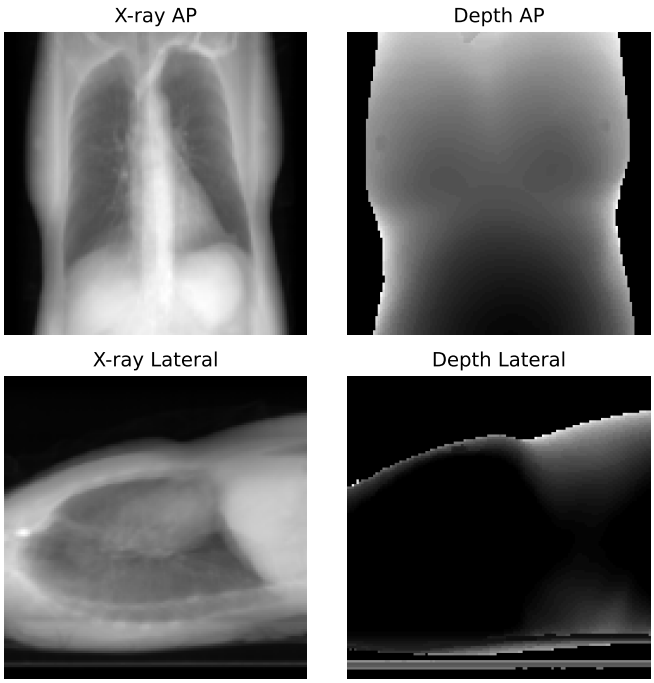


Fig. 2: The input data used in our experiments is generated synthetically using the corresponding CT volumes of the LIDC-IDRI dataset [18]. Depth images, AP and lateral X-Ray projections are simulated.

both input channels; and paired biplanar depth images. Quantitative results for each configuration are shown in Table I.

The evaluation reveals a consistent dependence of reconstruction quality on the diversity and complementarity of the input views. The configuration using paired biplanar X-ray projections achieves the strongest overall performance across all evaluation metrics, confirming that orthogonal radiographic views provide rich and complementary geometric information that is critical for accurate 3D reconstruction. Multimodal input combining a lateral X-ray projection with an AP depth image produces competitive results, suggesting that surface information can partially mitigate the absence of a second X-Ray projection. Duplicating the lateral X-ray across both input channels achieves similar performance to using multimodal input.

However, a t-test conducted at a significance level of 0.05 indicates that the differences in PSNR values between the following input configurations - lat. X-ray/AP X-ray, lat. X-ray/AP depth, and dual lat. X-ray - are not statistically significant. Furthermore, no statistically significant difference is observed between the RMSE values of the lat. X-ray/AP X-ray and the dual lat. X-ray input configuration.

Qualitative results are shown in Fig. 3. All input configurations demonstrate the ability to reconstruct anatomically correct CT volumes. However, the reconstructions lack clearly defined boundaries between organs, muscle and fat tissue, with exception of the lung. Especially in the lower abdomen, organ texture appears washed and muddy. Compared to the lat. X-ray/AP X-ray combination, lat. X-ray/AP depth is able to more accurately reconstruct the patient surface, leading to clearer reconstructions in the mediastinal area. As the lat. depth/AP depth combination does not provide information about the patients internals, the resulting reconstructions expectedly lack patient specific details such as correct organ sizes and appear generic.

TABLE I: Comparison of different input combinations and resulting evaluation metrics.

Inputs		**Metrics**		
Input 1	Input 2	RMSE ↓	PSNR ↑	SSIM ↑
Lat. X-ray	AP X-ray	237.17 ± 34.66	24.83 ± 1.18	0.55 ± 0.08
Lat. X-ray	AP Depth	247.32 ± 32.90	24.45 ± 1.10	0.54 ± 0.07
Lat. X-ray	Lat. X-ray	246.92 ± 37.19	24.49 ± 1.26	0.55 ± 0.07
Lat. Depth	AP Depth	279.86 ± 34.45	23.37 ± 1.03	0.48 ± 0.08

IV. Conclusion

This study assessed the influence of various input configurations on 3D reconstruction performance. Paired biplanar X-ray projections provided the highest accuracy, while a combination of a lateral X-ray and a depth image - leveraging the accessibility of depth cameras in modern CT systems - produced promising results. Qualitative results showed that both input configurations can reconstruct anatomically accurate CT volumes. Future work should explore improved integration of depth data and investigate the impact of noise and low-resolution depth images. Additionally, the integration of an organ segmentation loss could result in cleaner organ boundaries in the reconstructed volumes.

Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) under grant KA 1678/37-1, LE 2763/5 and MA 4898/26. The authors gratefully acknowledge the HPC resources provided by NHR@FAU using hardware funded by the German Research Foundation (DFG).

References

[1] Kalender, W., Wolf, H., Suess, C., Gies, M., Greess, H., and Bautz, W., "Dose reduction in CT by on-line tube current control: Principles and validation on phantoms and cadavers," *European radiology* **9**, 323–28 (1999).

[2] Gies, M., Kalender, W., Wolf, H., and Suess, C., "Dose reduction in CT by anatomically adapted tube current modulation. I. Simulation studies," *Medical physics* **26**, 2235–47 (1999).

[3] ICRP, "Managing patient dose in multi-detector computed tomography (MDCT). ICRP Publication 102.," *Annals of the ICRP* **37**, 73–78 (2007).

[4] ICRP, "The 2007 recommendations of the International Commission on Radiological Protection. ICRP Publication 103.," *Annals of the ICRP* **37**, 261 (2007).

[5] Klein, L., Liu, C., Steidel, J., Enzmann, L., Knaup, M., Sawall, S., Maier, A., Lell, M., Maier, J., and Kachelrieß, M., "Patient-specific radiation risk-based tube current modulation for diagnostic CT," *Medical Physics* **49** (2022).

[6] Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y., "Generative adversarial nets," in [*In Proc. NeurIPS*], 2672–2680, MIT Press (2014).

[7] Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., and Zheng, Y., "X2CT-GAN: Reconstructing CT from biplanar X-rays with generative adversarial networks," 10611–10620 (2019).

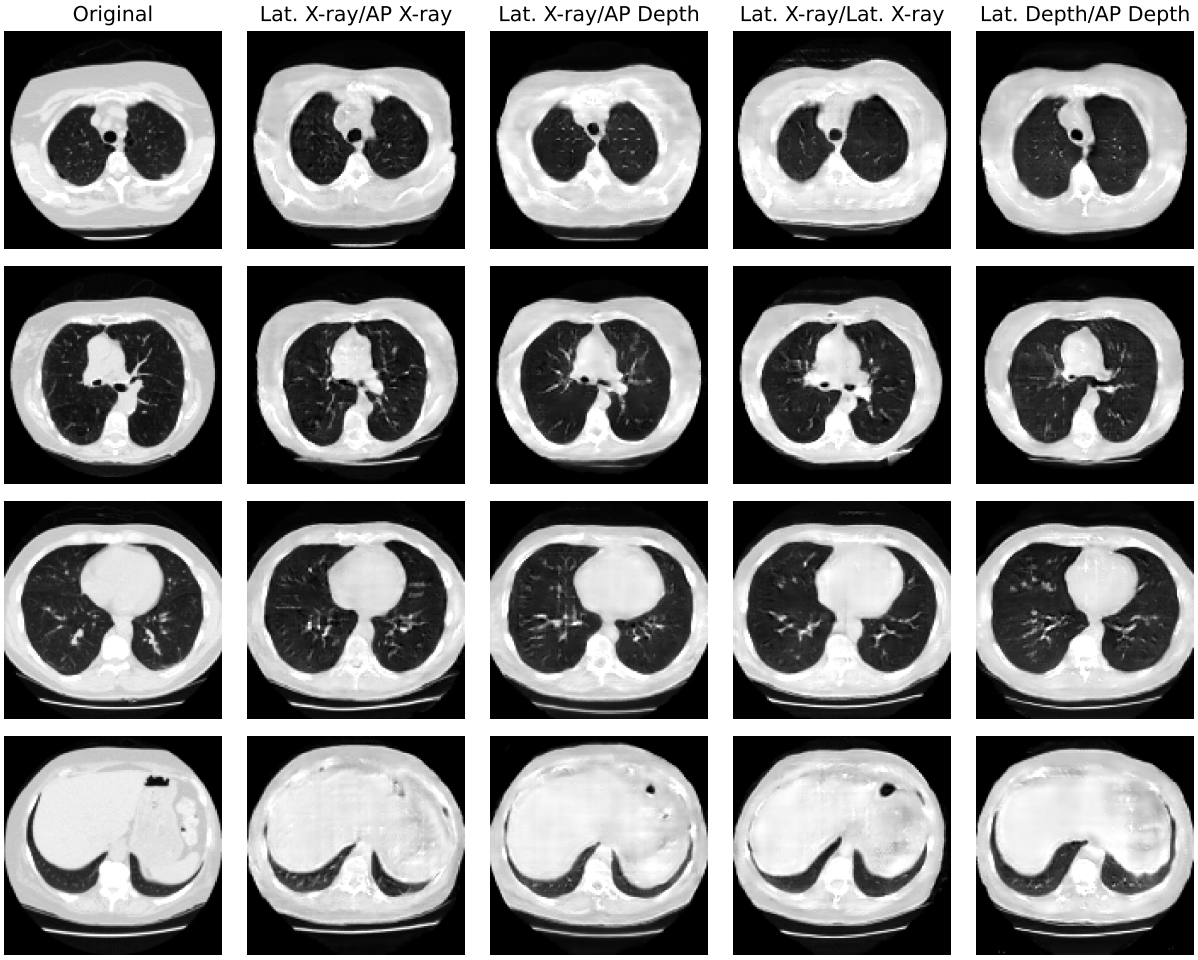


Fig. 3: Representative slices from reconstructed CT volumes using different inputs. Each row shows a slice from a different test sample, windowed to C = -600 HU, W = 1500 HU.

[8] Jiang, L., Zhang, M., Wei, R., bo, L., Bai, X., and Zhou, F., "Reconstruction of 3d CT from a single X-ray projection view using CVAE-GAN," 1–6 (2021).

[9] Liu, C., Klein, L., Huang, Y., Baader, E., Lell, M., Kachelrieß, M., and Maier, A., "Two-view topogram-based anatomy-guided CT reconstruction for prospective risk minimization," *Scientific Reports* **14** (2024).

[10] Tang, T., Cai, W., Yang, B., Huang, J., Zhou, Y., Liu, Z., Wang, Q., and Li, G., [*Reconstructing Chest CT from Orthogonal Biplanar X-rays via Feature Enhancement Blocks and Perceptual Consistency Loss*], 138–152 (2025).

[11] Kyung, D., Jo, K., Choo, J., Lee, J., and Choi, E., "Perspective projection-based 3d CT reconstruction from biplanar X-rays," in [*ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*], 1–5, IEEE (2023).

[12] Liu, X., Qiao, Z., Liu, R., Li, H., Zhang, J., Zhen, X., Qian, Z., and Zhang, B., "Diffux2ct: Diffusion learning to reconstruct CT images from biplanar X-rays," in [*European conference on computer vision*], 458–476, Springer (2024).

[13] Bier, B., Unberath, M., Geimer, T., Maier, J., Gold, G., Levenston, M., Fahrig, R., and Maier, A., "Motion compensation using range imaging in c-arm cone-beam ct," in [*Annual Conference on Medical Image Understanding and Analysis*], 561–570, Springer (2017).

[14] Bier, B., Ravikumar, N., Unberath, M., Levenston, M., Gold, G., Fahrig, R., and Maier, A., "Range imaging for motion compensation in c-arm cone-beam ct of knees under weight-bearing conditions," *Journal of Imaging* **4**(1), 13 (2018).

[15] Rausch, J., Maier, A., Fahrig, R., Choi, J.-H., Hinshaw, W., Schebesch, F., Haase, S., Wasza, J., Hornegger, J., and Riess, C., "Kinect-based correction of overexposure artifacts in knee imaging with c-arm ct systems," *International Journal of Biomedical Imaging* **2016**(1), 2502486 (2016).

[16] Rybakov, O., Bier, B., Maier, J., Unberath, M., and Maier, A., "Simultaneous generation of x-ray and range images using xcat under motion," in [*2017 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC)*], (2017).

[17] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., "Image-to-image translation with conditional adversarial networks," in [*Proceedings of the IEEE conference on computer vision and pattern recognition*], 1125–1134 (2017).

[18] Armato III, S. G., McLennan, G., Bidaut, L., McNitt-Gray, M. F., Meyer, C. R., Reeves, A. P., Zhao, B., Aberle, D. R., Henschke, C. I., Hoffman, E. A., et al., "The lung image database consortium (LIDC) and image database resource initiative (IDRI): a completed reference database of lung nodules on CT scans," *Medical physics* **38**(2), 915–931 (2011).

In-house nano-xCT with an x-ray spectrometer and a hybrid photon counting detector for integrated circuit metrology

Jordan Fonseca, Joseph W. Fowler, Felix H. Kim, Zachary H. Levine, John Henry Scott, Daniel S. Swetz, Pual Szypryt, Joel N. Ullom, Andras E. Vladar, and Nathan Nakamura

***Abstract*—Nanoscale x-ray computed tomography (nano-xCT) is a promising tool for nondestructive 3D imaging of integrated circuits (ICs) and advanced packaging. Already, 3D x-ray imaging of ICs has achieved sub-10 nm spatial resolutions when performed at synchrotrons, and there are numerous commercial tools for in-house imaging that achieve resolutions of 300 nm to 1000 nm. Nevertheless, there remains a great need amongst failure analysis divisions of semiconductor manufacturing companies for a fast, laboratory-based tool that can reconstruct portions of ICs at spatial resolutions in the 10s of nm in less than a day. It is also useful to be able to distinguish constituent metals in ICs. One promising approach to doing this with x-ray metrology is to use the spectrally distinct mass attenuation coefficient of different metals. The constraints of high resolution, high imaging speed, and spectral sensitivity offer a challenging trifecta of engineering obstacles. Here, we explore tradeoffs between performing tomography with an x-ray spectrometer versus a commercial hybrid photon counting detector (HPCD) with a scanning-electron-microscope-based nano-xCT tool. While the spectrometer offers 100x better energy resolving power, the HPCD collects data between 200x to 800x faster, depending on which spectral regions are used. We explore the challenges and nuance of how a photon counting detector with an energy resolution of ~1 keV and a broadband 1 keV – 25 keV x-ray source can be used to distinguish between Cu and W in an IC fabricated at the 130-nm node with spatial resolutions of 75 nm to 80 nm. At these spatial resolutions and imaging speeds for planar IC samples, this instrument offers a new and important capability for in-house imaging tools.**

***Index Terms*— Nano-xCT, hybrid photon counting detector (HPCD), element-sensitive, spectral binning**

The authors thank all team members who contributed to the development of and data collection using the proof-of-concept instrument, from both NIST and Sandia National Laboratories. Thanks to Amber Dagel, who led the fabrication of the original sample that was reused in this work. This work was performed with funding from the CHIPS Metrology Program, part of CHIPS for America, National Institute of Standards and Technology, U.S. Department of Commerce. Mention of commercial products does not imply endorsement by the authors or their institutions. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

J. F. is with the Quantum Sensors Division of the National Institute of Standards and Technology, 325 Broadway, Boulder, CO, 80305, USA (email: jordan.fonseca@nist.gov). J. W. F. is with the Quantum Sensors Division of the National Institute of Standards and Technology, 325 Broadway, Boulder, CO, 80305, USA (email: joe.fowler@nist.gov). F. H. K. is with the Intelligent Systems Division of the National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD, 20899, USA (email: felix.kim@nist.gov). Z. H. L. is with the Quantum Measurement Division of the National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD, 20899, USA (email: zachary.levine@nist.gov). J. H. S. is with the Materials Measurement Science Division of the National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD, 20899, USA (email: johnhenry.scott@nist.gov). D. S. S. is with the Quantum Sensors Division of the National Institute of Standards and Technology, 325 Broadway, Boulder, CO, 80305, USA (email: daniel.swetz@nist.gov). P.S. is with the Department of Physics of University of Colorado, Boulder, 2000 Colorado Ave., Boulder, CO, 80309, USA (email: paul.szypryt@nist.gov). A. V. is with the Microsystems and Nanotechnology Division of the National Institute of Standards and Technology, 100 Bureau Dr., Gaithersburg, MD, 20899, USA (email: andras.vladar@nist.gov). N. N. is with the Quantum Sensors Division of the National Institute of Standards and Technology, 325 Broadway, Boulder, CO, 80305, USA (email: nathan.nakamura@nist.gov)

I. Introduction

INTEGRATED circuits (ICs) are some of the most sophisticated human-made structures of all time. They are fabricated at a massive scale, and any given chip may contain billions of individual transistors. Because of the tremendous scale and cost of this hardware, there is a constant need for fast, high-resolution, in-house inspection tools that can perform various aspects of failure analysis. Historically, electron-based imaging techniques or synchrotrons have provided the highest-resolution 3D imaging of ICs [1], [2]; however, the penetration depth of electrons is low enough to require significant milling away and consumption of the sample being imaged, while synchrotrons are often unsuited for failure analysis labs that require day-of inspection or that must navigate export controls for proprietary chips.

Nanoscale x-ray computed tomography (nano-xCT) [3] is a promising imaging technique in this application space. Even relatively soft x-rays have sub-nanometer wavelength, putting their fundamental resolution limit well below the needs of current IC fabrication. Furthermore, x-rays can penetrate far deeper into an IC, making the sample preparation process less destructive. Finally, there are several experimental paradigms that make it possible to perform nano-xCT with laboratory-based tools that achieve sub-micron spatial resolution with sub-day measurement turnaround times.

While there are a wide variety of industrial and academic options for nano-xCT in the 300 nm – 600 nm range [4], [5], there is a burgeoning need for even higher resolutions—into the tens of nanometers—to meet the needs of modern semiconductor manufacturing and failure analysis. Since the number of photons required to preserve a given signal-to-

noise ratio in a planar sample scales as the fourth power of resolution, enhancing the capabilities of current nano-xCT tools is a nontrivial task that is butting up against fundamental hardware constraints with many of the current experimental paradigms.

In this work, we describe an approach to nano-xCT that uses a scanning electron microscope (SEM) and an IC sample with an x-ray target deposited directly onto its surface to generate a tunable, nanoscale x-ray spot size. We use both a NIST-developed x-ray spectrometer and a commercial hybrid photon counting detector (HPCD) [6] to perform element-specific tomography on an IC sample fabricated at the 130-nm node, with wiring as narrow as 160 nm clearly resolved. We explore the tradeoffs between using a slower spectrometer for element-resolving tomography and using a large photon counting detector for high-speed and high-resolution imaging. Finally, we discuss the technical barriers to using the energy thresholding capabilities of an HPCD for spectral tomography, as well as the avenues that can be explored to overcome these challenges.

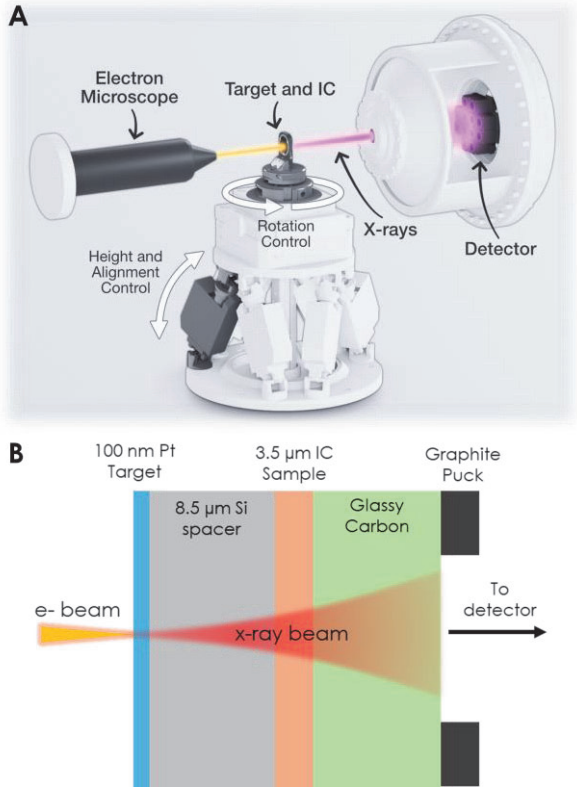


Figure 1: Experimental schematic. (A) Cartoon schematic of the tomography tool. An electron column of an SEM generates 25 keV electrons that impinge on a sample surface. The sample can be accurately positioned and rotated within the SEM vacuum chamber. X-rays that are generated in the sample pass through a Be window in the SEM chamber to an x-ray detector placed outside the SEM vacuum space. (B) Cartoon schematic of the side-view of the sample showing a key innovation of our approach to tomography: the Si backing of the IC is thinned down to below 10 μm, and a 100 nm thick layer of Pt is deposited directly onto this surface (left). The IC wiring layers can also be thinned down to include only the layers of interest, and the entire circuit is then attached to a glassy carbon substrate for structural integrity and to a graphite puck for mounting to the sample holder in the SEM chamber. Electrons are tightly focused to a <100 nm spot size on the Pt surface, which generate a spectrum of bremsstrahlung and characteristic x-ray emission from the Pt target. These x-rays pass through the IC sample ROI, through the glassy carbon, and out through a Be window at the back of the SEM vacuum chamber to the x-ray detector on the other side. This experimental apparatus makes it possible to have a nanoscale x-ray spot size without the use of x-ray optics, and the tiny distance between x-ray spot and sample compared to the much larger distance between the spot size and the x-ray detector enables massive geometric magnification.

II. Methods

Our nano-xCT tool consists of an SEM chamber, IC sample mounted to a rotation stage, and x-ray detector, either a spectrometer or HPCD (Fig. 1A). By using an SEM, we leverage electron optics to tightly focus an electron beam onto the sample surface, generating x-rays in a Pt target a mere 10 μm away from the IC wiring that we are interested in imaging (Fig. 1B). This approach guarantees the ability to create a tunable, nanoscale x-ray spot size without the use of x-ray optics. Furthermore, it allows us to achieve unparalleled geometric magnification ($M_g \approx 25\,000$) of the sample while keeping the x-ray detector close to the sample to collect as many photons as possible. In this way, we achieve experimental conditions that enable both high spatial resolution as well as high imaging speeds.

The details of our experimental apparatus have been explained elsewhere. For a thorough description of the instrumentation with an x-ray spectrometer, see ref. [7]. More information about TomoScatt, the NIST-developed reconstruction algorithm used here, can be found in refs. [8] and [9]. A more detailed description of the aspects of the technical considerations associated with using a large HPCD for nano-xCT can be found in ref. [10].

III. Results

The first iteration of our experiment used a NIST-developed x-ray spectrometer to collect signal photons. This detector made it possible to measure the energy-dependent attention of photons that interact with the sample, thereby distinguishing voxels by the specific material present in each one. Fig. 2A shows the average x-ray spectrum generated in the Pt target that has passed through the IC sample to the spectrometer, overlayed by the x-ray attenuation lengths for two common metals used in IC fabrication: Cu and W. By using the four energy bins noted in the plot and the corresponding different attenuation lengths, it is possible to generate reconstructions for either Cu (Figs. 2B, E) or W (Figs. 2C, F) and to compare them with the GDS file for those layers (Figs. 2D, G). In this way, the geometry *and the composition* of each layer of the sample can be independently verified by nano-xCT.

While the information value in collecting a high-resolution spectrum for each ray that interacts with the sample is hard to overstate, a core motivation of our work is to produce a tool that is capable of industry-relevant spatial resolutions at

industry-relevant speeds, and collecting a full dataset took over 210 hours for a sample ROI of ~ 700 μm^3. To overcome the photon-starved nature of our measurement, we replace the spectrometer with a commercial HPCD, which has over four million pixels spanning an area over 800 times larger than the spectrometer. This change enabled significantly improved measurement speeds and photon collection efficiency but did so at the cost of spectral resolution (Table 1).

Table 1: Spectrometer vs. HPCD

Property [units]	Spectrometer	HPCD
Solid angle [sr]	0.015	0.35
Projected area [μm^2]	1.5	36
Projected pixel [nm]	50	3
Operation Mode	**Spectrometer &** Photon counter	Photon counter
Time for 10^9 photons [min]	**2,300**	**14**

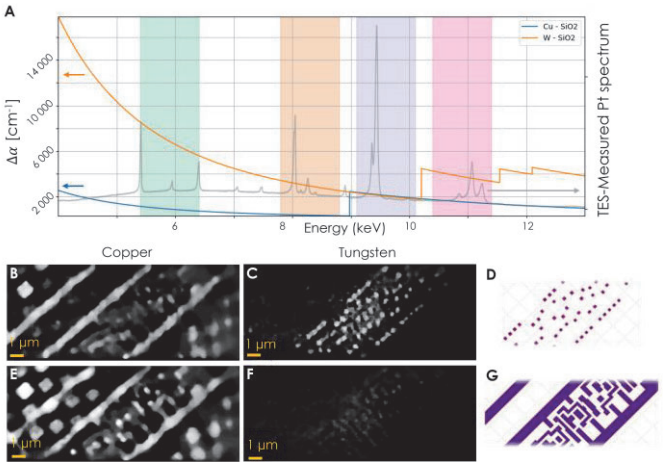


Figure 2: Element-specific tomography enabled by a spectrometer. (A) Plot showing the Pt source x-ray spectrum (gray curve, right axis) and the attenuation lengths of Cu and W (orange and blue curves, left axis). Note that the Pt target was selected intentionally because of the bright Lα emission line above the Cu absorption edge. This guarantees that there is a bright source of x-rays that are heavily attenuated by the Cu wiring in the circuit, thereby enhancing contrast. The green, orange, lilac, and pink boxes show the four spectral regions that were input into the reconstruction algorithm, which then produced two distinct reconstructions from a single dataset, one representing Cu and the other representing W. Comparison of the Cu (B) and W (D) reconstructions for a slice corresponding to small vias that connect the Si to the circuitry (D). Comparison of the Cu (E) and W (F) reconstructions for a different slice corresponding to metal wiring (G). It is worth noting that the vias show up much more clearly in the W reconstruction while the metal wiring is more apparent in the Cu reconstruction. This distinction, which was only possible because of the unique attenuation information provided by the availability of multiple spectral regions in the data, highlights the utility of being able to distinguish between materials in an IC with minimal prior knowledge of the circuit.

To benchmark the performance of our HPCD, we collect a tomography dataset of nearly the same sample region studied with the spectrometer. Fig. 3A shows the GDS file for the ROI, with Fig. 3B (Fig. 3C) showing stitched x-ray radiographs for the ROI collected at a sample rotation angle of 0° (+15°). These data can subsequently be input into a reconstruction algorithm to generate the full 3D reconstruction of the circuit.

While using an HPCD simplifies data collection and offers dramatic improvements to photon collection rate, one open question is whether the basic energy thresholding capabilities can be leveraged to preserve some semblance of spectrally resolved tomography that would enable elemental identification. To address this, we begin by exploring how the "ideal" spectrum that an HPCD is designed to measure (Fig. 4A) compares with the broadband spectrum generated by the SEM beam impinging upon the Pt target (Fig. 4B). The energy thresholds on the HPCD primarily serve to guarantee that signal photons are counted correctly, particularly when an incident photon distributes its energy across multiple pixels. This effect, commonly known as charge-sharing, imposes significant design, readout, and calibration needs to work correctly for monochromatic sources, let alone for broadband spectra. While the details of how charge-sharing effects result in erroneous count rates or misassigned photon energies is beyond the scope of this proceedings, a core focus of our ongoing work is quantifying these errors and determining how they can be overcome with clever experimental choices to achieve spectral/elemental sensitivity with the approximately 1 keV energy resolution we can expect from our HPCD.

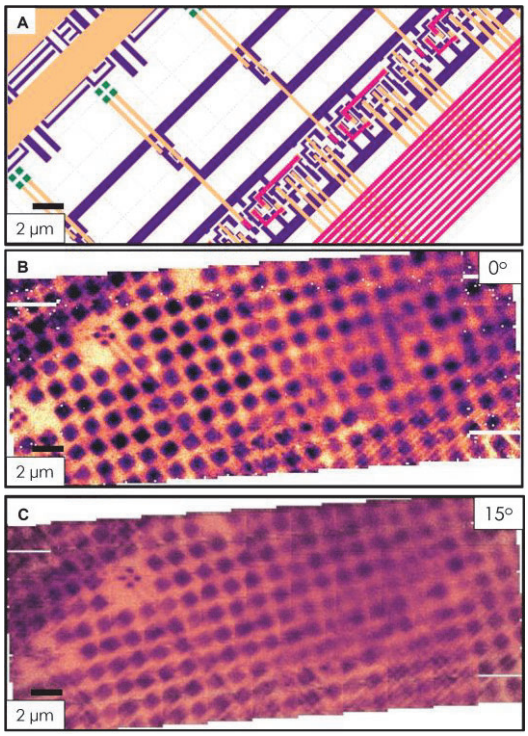


Figure 3: Imaging a portion of an IC with an HPCD. (A) The GDS file for the ROI of the IC being imaged, showing several different wiring features and vias at different layer depths, as indicated by the different colors. (B) x-ray radiograph collected at normal incidence with the HPCD. The

dark cubes across the image are chemical mechanical polish (CMP) fill, which is added by the foundry and does not appear in the GDS file. Even prior to reconstruction, wiring features and the clover-shaped collection of four vias are clearly visible. (C) x-ray radiograph collected with the sample rotated to 15°, where blurring and elongation of the features begins to become apparent.

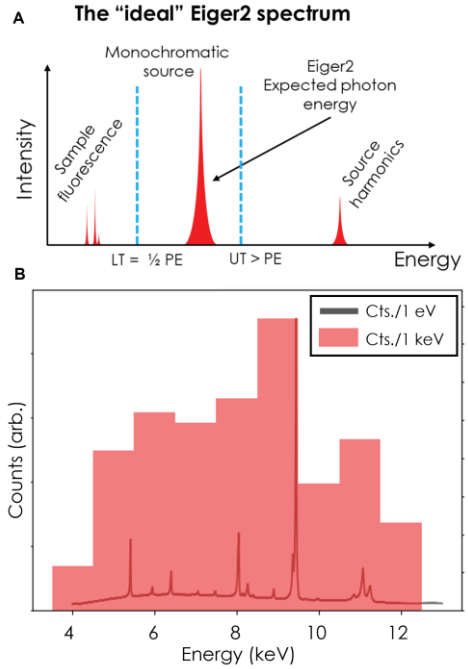


Figure 4: HPCD spectral considerations. (A) Cartoon schematic of the type of x-ray spectrum that an HPCD is optimized to detect. The target photons that will be detected come from a monochromatic source, such as a synchrotron beamline. A lower threshold can be set on the HPCD that will still detect signal photons that deposit energy across multiple pixels while rejecting low-energy fluorescence from the sample. An optional upper energy threshold rejects high energy background light, such as from cosmic rays, as well as high harmonics that may be generated by a high intensity, pulsed, x-ray source. (B) Plot showing the x-ray source generated by the electron beam on the Pt target inside our SEM, as measured by the spectrometer (gray). This high-resolution spectrum can be downsampled to 1 keV bins (red) to better reflect the energy resolution that might be expected from an HPCD.

IV. Conclusion

This work details the development of a laboratory-based tabletop, SEM-based, nano-xCT tool capable of achieving sub-200 nm spatial resolution of planar IC samples and of doing so either at industry-relevant speeds or with high spectral resolution that enables sensitivity to distinct constituent metals that make up the circuit, and possibly with both. This and prior preliminary work establish the viability of this technique for semiconductor failure analysis labs. The next generation of this tool, which includes a newly designed SEM chamber, a brighter, tighter-focusing electron gun, and an improved sample positioning stage, is already under development. New samples at the 45 nm node are being prepared, and significant progress has been made toward understanding the HPCD energy thresholding well enough to distinguish between metals in the IC without the use of a high-resolution spectrometer, all while achieving even greater spatial resolution and data collection/analysis speeds. Finally, the implementation of standard image quality metrics will make it possible to predict and compare performance across a variety of commercial and academic nano-xCT tools.

References

[1] M. Holler *et al.*, "X-ray ptychographic computed tomography at 16 nm isotropic 3D resolution," *Sci. Rep.*, vol. 4, no. 1, p. 3857, Jan. 2014, doi: 10.1038/srep03857.

[2] T. Aidukas *et al.*, "High-performance 4-nm-resolution X-ray tomography using burst ptychography," *Nature*, vol. 632, no. 8023, pp. 81–88, Aug. 2024, doi: 10.1038/s41586-024-07615-6.

[3] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. Description of system," *Br. J. Radiol.*, vol. 46, no. 552, pp. 1016–1022, Dec. 1973, doi: 10.1259/0007-1285-46-552-1016.

[4] F. Nachtrab *et al.*, "NanoXCT: development of a laboratory nano-CT system," in *Developments in X-Ray Tomography IX*, SPIE, Sep. 2014, pp. 154–161. doi: 10.1117/12.2061752.

[5] T. Dreier, D. Nilsson, and J. Hållstedt, "Fast and high-resolution X-ray nano tomography for failure analysis in advanced packaging," *Microelectron. Reliab.*, vol. 168, p. 115694, May 2025, doi: 10.1016/j.microrel.2025.115694.

[6] S. Manolopoulos *et al.*, "X-ray powder diffraction with hybrid semiconductor pixel detectors," *J. Synchrotron Radiat.*, vol. 6, no. 2, Art. no. 2, Mar. 1999, doi: 10.1107/S0909049599001107.

[7] N. Nakamura *et al.*, "Nanoscale Three-Dimensional Imaging of Integrated Circuits using a Scanning Electron Microscope and Transition-Edge Sensor Spectrometer," Mar. 04, 2024, *arXiv*: arXiv:2212.10591. Accessed: Sep. 26, 2024. [Online]. Available: http://arxiv.org/abs/2212.10591

[8] Z. H. Levine *et al.*, "A tabletop X-ray tomography instrument for nanometer-scale imaging: reconstructions," *Microsyst. Nanoeng.*, vol. 9, no. 1, pp. 1–11, Apr. 2023, doi: 10.1038/s41378-023-00510-6.

[9] Z. H. Levine, T. J. Blattner, A. P. Peskin, and A. L. Pintar, "Scatter Corrections in X-Ray Computed Tomography: A Physics-Based Analysis," *J. Res. Natl. Inst. Stand. Technol.*, vol. 124, p. 124013, May 2019, doi: 10.6028/jres.124.013.

[10] J. Fonseca *et al.*, "Toward In-House 3D Nanoscale X-Ray Tomography of Integrated Circuits for Quality Control and Hardware Security at Speed of Need," in *2025 IEEE Physical Assurance and Inspection of Electronics (PAINE)*, Oct. 2025, pp. 1–7. doi: 10.1109/PAINE66113.2025.11320158.

DBP-based geometry-robust artifact removal for rotational CL

Wuliang Shi, Changyu Chen, Yuxiang Xing*

***Abstract*—Computed laminography (CL) is widely used for three-dimensional imaging of plate-shaped objects. However, due to the inherent incompleteness of projection data, CL reconstructions often suffer from severe interlayer aliasing artifacts. In addition, projection truncation often occurs in practical CL imaging scenarios, further degrading reconstruction quality. To address these challenges, we propose a deep learning-based framework for CL artifact removal integrated with a differentiated backprojection (DBP) reconstruction. The proposed method explicitly incorporates the geometric characteristics of CL aliasing artifacts into a convolutional neural network (CNN), enabling effective learning of artifact-related features and improved its adaptability to various imaging geometries. Meanwhile, the DBP reconstruction robustly handles truncated projections. Extensive experiments on both simulated and real data demonstrate that the proposed method significantly suppresses aliasing artifacts and achieves robust performance across different CL imaging geometries.**

***Index Terms*—Computed Laminography, Differentiated Backprojection, Deep Learning, Deformed Convolution, Non-destructive Testing.**

I. Introduction

COMPUTED laminography (CL) is an X-ray imaging technique for non-destructive inspection of plate-shaped objects. By tilting the rotation axis with respect to the incident X-ray beam, CL enables effective imaging of samples such as printed circuit boards and other industrial components.

Currently, numerous studies have focused on CL reconstruction, which can be categorized into three types: analytical reconstruction, iterative reconstruction, and deep learning-based reconstruction. Analytical reconstruction in CL drew inspiration from traditional CT algorithms, mainly filtered backprojection (FBP) [1]. We have proposed a DBP reconstruction algorithm for CL (CL-DBP) [2]. which significantly outperforms FBP-type methods in addressing truncated projections. However, the inability of CL to acquire complete projection data results in severe aliasing artifacts in the analytical reconstructed images. Iterative reconstruction methods can reduce aliasing artifacts with manually designed prior knowledge [3]. Lu introduced adaptive weights based on the image gradient in different directions, effectively suppressing noise and artifacts while preserving edges and details as much as possible [4]. However, iterative reconstruction algorithms and iterative deep learning networks typically incur high computational costs and highly dependent on the choice of algorithm parameters. In recent years, advances in deep learning have provided new perspectives for CL reconstruction. Most existing approaches first obtain preliminary 3D images using analytical or iterative reconstruction algorithms, and then employ neural networks for image-domain artifact removal. In our previous work [5], a U-Net-based networks was developed to suppress artifacts and noise in reconstructed images from CL-FDK, leading to improved image quality of Ball Grid Array (BGA) solder joints. However, Image-domain post-processing deep learning approaches often rely on fixed system configurations. Therefore, robust reconstruction across different geometric configurations and spatial scales remains an open challenge in CL. Given the need for multi-scale analysis in materials and device studies [6] and the frequent adjustment of CL system geometries in practice, geometrically robust reconstruction across different geometric configurations and spatial scales are essential. In addition, projection truncation, commonly encountered in ROI CL scanning, is an issue that requires careful consideration.

This work proposes a geometry-aware CNN that models CL aliasing artifacts by deformed convolutions to achieve robust artifact suppression across different imaging geometries without retraining or finetuning. The integration of CL-DBP further ensures robustness to projection truncation in ROI scanning. The paper then describes the characteristics of artifacts in CL and proposed method, followed by experimental validation and discussion.

II. Methods and Materials

A. Basics and Characteristics of CL Imaging

Without losing generality, we formulate CL imaging under the most widely used rotating CL configuration as shown in Fig. 1 (a). In this system, both the X-ray source and detector move along circular trajectories on opposite sides of a scanned object, with the orientation of the detector array remaining fixed during rotation.

Let the image to be reconstructed denoted by $f = f(\mathbf{x}) = f(x, y, z)$ with $\mathbf{x} = (x, y, z)$ being the spatial coordinates. Its projection can be expressed by the following

Submitted date: Jan 24, 2026. This work was supported in part by National Key R&D Program of China (No.2022YFF0607802) and China Postdoctoral Science Foundation (No. 2025M780861).

Wuliang Shi, Changyu Chen, and Yuxiang Xing are with the Department of Engineering Physics, Tsinghua University, Beijing. 100084, China, and also with the Key Laboratory of Particle and Radiation Imaging (Tsinghua University), Ministry of Education, Beijing 100084, China. (Author correspondence: xingyx@mail.tsinghua.edu.cn; ccy@mail.tsinghua.edu.cn; swl22@mails.tsinghua.edu.cn).

equation: $p(\xi,u,v)=\int_0^{\infty} f(\boldsymbol{A}(\xi)+t\boldsymbol{\lambda}(\xi,u,v))dt$,where $\boldsymbol{A}(\xi)$ denotes the spatial coordinates of the X-ray source, and $\boldsymbol{\lambda}(\xi,u,v)$ is the unit direction vector of the X-ray. We have derived a DBP-type reconstruction algorithm for rotating CL imaging system (CL-DBP) in [2] that enables fast reconstruction and effectively addresses data truncation problem:

$$f^z(x,y)=f^z(t\boldsymbol{\beta}^{\perp}+s\boldsymbol{\beta})=\frac{-1}{(s-L_t)(U_t-s)}\int_{L_t}^{U_t}\frac{(s'-L_t)(U_t-s')\,b_{\theta}^z(t\boldsymbol{\beta}^{\perp}+s'\boldsymbol{\beta})}{-2\pi^2(s-s')}ds'+C_t^z \tag{1}$$

where b_{θ}^z denotes the differential backprojection result with PI line oriented along the θ direction. More details of the derivation for CL-DBP can be found in [2].

CL project data does not satisfy the Tuy-Smith condition, which makes exact reconstruction through analytic methods unattainable. As shown in Fig. 1 (b), projections acquired by CL systems results in severe absence of frequency domain information along the z-axis, which leads to severe aliasing artifacts in CL reconstructions. In Fig. 1 (c), we show an exemplar reconstruction obtained by CL-DBP that is contaminated by aliasing artifacts manifesting as structural distortions in the z-axis direction.

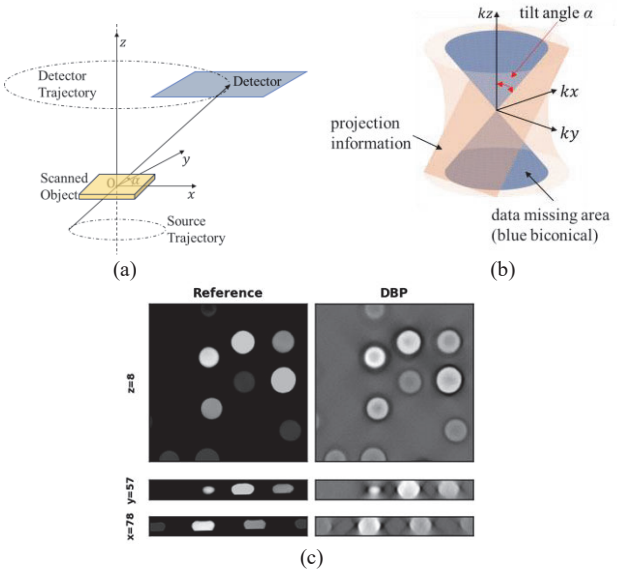


Fig. 1. Characteristics of CL imaging: (a) illustration for imaging geometry of rotating CL, (b)schematic of the data missing area in frequency-domain, (c) the influence of aliasing artifacts on DBP reconstructed images.

In analytic reconstructions, the geometrical characteristics of aliasing artifacts are closely tied to backprojection. During scanning, projection data are integrations of voxel intensity along each ray path. Backprojection redistributes the integrated information back to the voxels along the same path in an "aggregation" manner. Due to incomplete projections in CL, the information of a single voxel is distributed over a three-dimensional biconical surface, which can be interpreted as a geometry-dependent point spread function. This artifact structure varies with both spatial position and scanning geometry. Near the rotation center, the biconical surface is symmetric, whereas it becomes tilted at off-center locations, leading to pronounced changes in artifact morphology.

In horizontal slices, aliasing artifacts from different structural layers appear as overlapping ring-like patterns that are difficult to distinguish. In contrast, vertical slices exhibit intersecting line patterns with well-defined geometric characteristics, making them more suitable for artifact modeling and suppression. Motivated by these observations, this work leverages the geometry-dependent biconical artifact distribution to achieve robust CL artifact reduction.

B. Reconstruction Framework

The proposed CL reconstruction method integrates the DBP analytical reconstruction operator with deep learning networks. The overall workflow of the proposed method is illustrated in Fig. 2.

Let $\mathbf{P}=p(\xi,u,v)$ denote the measured projection data. The proposed framework uses a dual-branch pipeline. The projection data are first processed by a DBP-based analytical module to get two preliminary reconstructions with PI-line along x- and y-axes, denoted as $\mathbf{f}_x^{\mathrm{DBP}}=\mathcal{D}_x(\mathbf{P})$ and $\mathbf{f}_y^{\mathrm{DBP}}=\mathcal{D}_y(\mathbf{P})$, respectively. We design a Geometry-Robust Artifact Suppression Network (GRASN) module to suppress artifacts in $\mathbf{f}_x^{\mathrm{DBP}}$ and $\mathbf{f}_y^{\mathrm{DBP}}$ slice-by-slice. For each target slice, a stack of neighboring slices centered on the target is extracted to provide sufficient contextual information for network processing. Specifically, the network module takes this multi-slice stack as input and produces a single artifact-suppressed slice as output. For $\mathbf{f}_x^{\mathrm{DBP}}$, we process xz-slices. For $\mathbf{f}_y^{\mathrm{DBP}}$, we process yz-slices. GRASN explicitly models the geometry-dependent characteristics of aliasing artifacts to efficiently suppress aliasing artifacts. The outputs of the GRASN are then fused using a bowtie filter to produce the final 3D reconstructed volume $\hat{\mathbf{f}}$.

The overall reconstruction process can be summarized as $\hat{\mathbf{f}}=\mathcal{B}_{\mathrm{SF}}\left(\mathcal{G}_{\boldsymbol{\Theta}}\left(\mathcal{T}_x(\mathcal{D}_x(\mathbf{P})),\mathcal{T}_y(\mathcal{D}_y(\mathbf{P}))\right)\right)$, with $\mathcal{T}(\cdot)$ denoting the image stack extraction, $\mathcal{G}_{\boldsymbol{\Theta}}(\cdot)$ representing the GRASN with learnable parameters $\boldsymbol{\Theta}$ and $\mathcal{B}_{\mathrm{SF}}(\cdot)$ the bowtie-based spectral fusion module.

We present the details of each module in the following:

1. DBP Module: The first step of the proposed method is to use a DBP module to perform analytical reconstruction with PI line oriented along the x-axis and y-axis, respectively. The computation of C_t^z in (1) requires estimating $p^z(\theta,t)$, which cannot be directly obtained in CL and inevitably introduces errors in estimation [2]. Since C_t^z primarily affects only the reconstruction magnitude, we omit this term and rely on the network for compensation. Therefore, the DBP module is:

$$f^z(x,y) = f^z(t\boldsymbol{\beta}^{\perp} + s\boldsymbol{\beta}) = \frac{-1}{(s-L_t)(U_t-s)} \int_{L_t}^{U_t} \frac{(s'-L_t)(U_t-s')\, b_\theta^z(t\boldsymbol{\beta}^{\perp} + s'\boldsymbol{\beta})}{-2\pi^2(s-s')}\, ds' \tag{3}$$

By exploiting this DBP module, we can reduce truncation artifacts in reconstructions, so that the network can focus on removing aliasing artifacts.

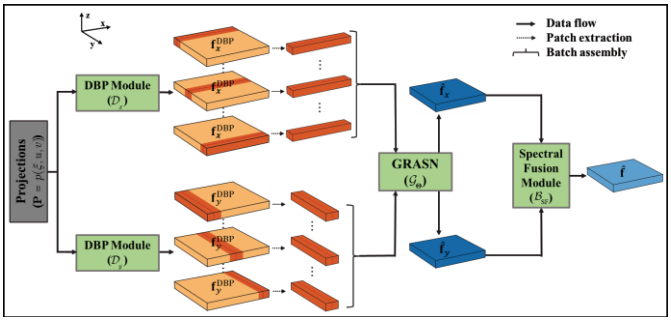


Fig. 2. Framework of the proposed CL reconstruction method.

2. GRASN: Our network for artifact reduction is based on a standard U-Net backbone. A 2D CNN is adopted where the third dimension of a multi-slice stack is encoded through multiple input channels, achieving a balance between computational cost and network capability. Compared with the classical U-Net, GRASN introduces block-wise residual connections between the input and output feature maps to facilitate stable gradient propagation and improved feature preservation. In addition, pooling and upsampling operations are restricted to the in-plane directions to better preserve the structural features of plate-like objects with limited extent along the depth direction, thereby improving artifact suppression.

To further model the geometry-dependent characteristics of CL artifacts, deformed convolutions are employed in the first three encoder blocks of GRASN. Unlike standard convolutions that assume spatial invariance, CL artifacts exhibit strong spatial and geometry dependence. Deformed convolutions allow the receptive field to adapt to spatially variant structures, thereby improving the network's ability to capture location-dependent artifact morphology.

In implementation, standard convolution layers in the encoder are replaced with deformable convolution modules provided by PyTorch [7]. Moreover, the sampling offsets are explicitly derived from the known geometric configuration of intersecting artifact lines. Specifically, the convolutional sampling points are shifted to positions on the intersecting artifact lines, whose orientations are characterized by angles γ_1 and γ_2 as illustrated in Fig. 3, to embed geometric priors into the network. For a 3×3 kernel, only diagonal sampling points are adjusted, with offsets defined by horizontal and vertical displacements Δw and Δh in Fig. 3.

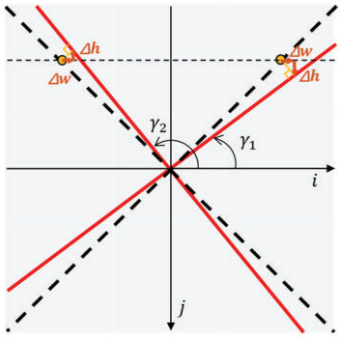


Fig. 3. Illustration of the explicit calculation of offsets for deformed convolution.

3. Spectral Fusion Module: After artifact suppression by GRASN, the outputs from the two branches are fused to form the final reconstruction. To compensate for directional frequency loss in DBP, z-direction slices are transformed into the frequency domain, where a bowtie-shaped spectral weighting is applied to selectively combine reliable frequency components from each branch, yielding the final result [2].

III. Experimental Study Results

A. Data preparation and Network Training

The dataset includes spherical, BGA, line-pair, and PCB phantoms, with a training-to-test split of 9:1. Numerical phantoms of sphere and BGA are generated using the method in [5]. Numerical line-pair phantoms are formed by generating rectangular line pairs with varying lengths, widths, spacing, and intensities distributed across the full field of view. PCB phantoms are derived from PCB design drawings augmented by cutting, rotation, scaling, stitching, and stacking. All phantom intensities were normalized to [0,1.6] to match CT reconstruction values of real PCBs, and all phantom were resized to 256×256×25. CL projections are simulated using the geometry in Fig. 1 (a), with a hybrid noise model applied to mimic realistic acquisition conditions. The projection data contains 64 views, each 300×300.

For each stack, 51 consecutive slices are input as 51 channels to GRASN for a reasonable coverage of 3D artifact extent, and the network outputs the artifact-corrected central slice. Model training was performed using two Quadro RTX 8000 GPUs with PyTorch Distributed Data Parallel (DDP) for acceleration, while inference was conducted on a single GPU.

B. Experimental Results

In the simulation study, the FDK+U-Net method [5] and the AwDaRTV method [4] are implemented for comparison. The network architectures, implementation details, and hyperparameter settings of all methods strictly follow their respective original studies. Moreover, we replace the standard convolutions in the first three stages of the U-Net with deformed convolutions (DC) proposed in this work to obtain an FDK+U-Net(DC) method for comparison too. The performance on a BGA phantom from the test set is shown as a representative example.

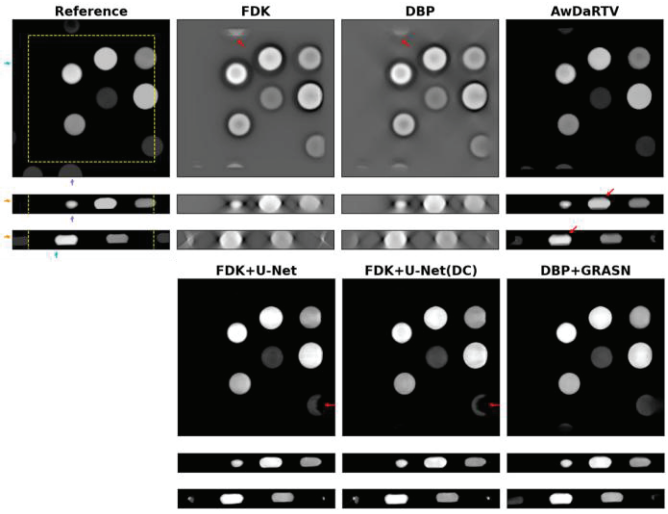


Fig. 4. Reconstructed images using different methods. The display window is [-0.5, 0.8] for FDK and DBP reconstructions, and [0.1, 1.6] for other images. The yellow dashed box indicates the ROI. In the reference, colored arrows indicate the relative positions of slices along different directions within the other slices. The red arrows highlight regions showing pronounced differences among the reconstructions.

TABLE I
QUANTITATIVE COMPARISON FOR RECONSTRUCTIONS ACROSS VARIOUS METHODS

Method	PSNR	RMSE	SSIM
FDK	8.5713	0.3728	0.0624
DBP	8.1926	0.3894	0.0593
AwDaRTV	31.3486	0.0271	0.8865
FDK+U-Net	30.1718	0.0310	0.9169
FDK+U-Net(DC)	30.1045	0.0312	0.9200
DBP+GRASN	**31.5671**	**0.0264**	**0.9236**

Reconstructed slices in three directions (z-, y-, and x-directions, from top to bottom) for each reconstruction method are shown in Fig. 4. Compared with FDK, DBP shows its advantage in handling truncated projections. AwDaRTV is less effective in mitigating aliasing artifacts than the proposed method and the learning-based baselines. The remaining comparative methods suffer from evident structural degradation, particularly in the lower-left region of the z-direction slice. In contrast, the proposed method preserves structural integrity within the ROI more effectively and substantially suppresses truncation artifacts. As shown in Table I, the proposed method outperforms the compared approaches across all metrics.

TABLE II
QUANTITATIVE COMPARISON FOR RECONSTRUCTION FROM DIFFERENT ABLATION CONFIGURATIONS

Method	PSNR	RMSE	SSIM
Without Deformed Convolution	32.8661	0.0227	0.9039
With Pooling In Z-direction	33.9422	0.0201	0.9062
Proposed	**33.9837**	**0.0200**	**0.9064**

Ablation studies were conducted to evaluate the effects of deformed convolution and the z-direction pooling strategy. Three configurations were compared using identical training data and settings: the proposed GRASN, GRASN without deformed convolutions, and GRASN with z-direction pooling.

As summarized in Table II, incorporating deformed convolution leads to clearer improvements in reconstruction accuracy, achieving higher PSNR and lower RMSE compared with the version without deformed convolution. We will present more results demonstrating adaptation to various scanning geometries on conference.

IV. CONCLUSION

In this work, we propose a geometry-aware CL reconstruction framework that effectively addresses aliasing and truncation artifacts commonly encountered in industrial inspection. By embedding the geometric characteristics of CL artifacts into deformed convolutions, the proposed method achieves superior artifact suppression, improved reconstruction accuracy, and strong robustness across different imaging geometries and spatial scales. Mainly, leveraging the intrinsic ROI reconstruction capability of DBP enables reliable performance under truncated projection conditions; the explicit modeling of artifact geometry using deformed convolution provides a flexible mechanism for capturing geometry-dependent artifacts. Although demonstrated on circular cone-beam CL, the framework is readily extendable to other CL trajectories and related geometry-limited imaging problems. Our future work will focus on integrating projection-domain learning to further enhance reconstruction quality.

REFERENCES

[1] M. Yang, G. Wang, and Y. Liu, "New reconstruction method for x-ray testing of multilayer printed circuit board," *Opt. Eng.*, vol 49, no. 5, pp. 056501, May 2010.

[2] W. Shi, X. Zou, C. Chen, and Y. Xing, "Differentiated backprojection reconstruction method for square cross-section FOV rotational CL," *CT Theory and Applications*, vol. 34, no. 4, pp. 621–631, Jul. 2025.

[3] Z. Gui, Y. Liu, Y. Liu, W. Liu, R. Yan, and L. Wang, "3-D computed laminography based on prior images and total variation," *IEEE Trans. Nucl. Sci.*, vol. 70, no. 3, pp. 189–199, Mar. 2023.

[4] J. Lu, Y. Liu, Y. Chen, H. Shu, Z. Li, J. Kang, and Z. Gui, "Cone beam computed laminography based on adaptive-weighted dynamic-adjusted relative total variation," *Nucl. Instrum. Methods Phys. Res. A, Accel. Spectrom. Detect. Assoc. Equip.*, vol. 1051, p. 168200, Oct. 2023.

[5] X. Zou, W. Shi, M. Du, and Y. Xing, "Artifact reduction in rotational computed laminography using a deep learning method," *Opt. Lasers Eng.*, vol. 187, p. 108881, Nov. 2025.

[6] V. Wood, "X-ray tomography for battery research and development," *Nature Rev. Mater.*, vol. 3, no. 9, pp. 293–295, Sep. 2018.

[7] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei. (2017, Oct.). Deformable Convolutional Networks. presented at IEEE Int. Conf. Comput. Vis. (ICCV).

Simulation of tire phantom based on X-ray computed tomography

Jiaying Zhang, Zhencheng Lin, Zhicong Yu

***Abstract*—Computed tomography has been widely used for the inspection of high-value tires. However, the lack of dedicated tire phantoms limits systematic simulation and validation. In this work, we propose an analytical 3D tire phantom that mimics a typical aviation tire (H44.5×16.5R21). The phantom was constructed using torus, cylinder, and cone primitives with clipping planes to represent a realistic tire body and internal features. An analytical line-integral formulation was derived for accurate forward projection, and reconstruction experiments were conducted to validate the proposed framework. Although the present work focuses on an aviation tire containing three types of objects, the proposed approach can be naturally extended to other types of tires by adjusting dimensions and incorporating additional classes of quadratic objects, such as ellipsoids and hyperbolas.**

***Index Terms*—computed tomography, phantom, tire**

I. Introduction

TIRE inspection plays a critical role in ensuring product safety, improving manufacturing quality, and enabling efficient retreading. In recent years, X-ray-based nondestructive testing (NDT) has become an important tool for tire inspection, particularly for high-value and heavy-duty applications [1], [2] such as aircraft tires [3]. Compared with alternative techniques such as laser shearography [4] and ultrasound methods [5], X-ray imaging provides superior capability for probing internal structures, enabling sophisticated analysis for material characterization and structural consistency evaluation. However, tires possess complex toroidal geometries with multi-layered compositions and fine-scale textures [6], which introduce challenges for the assessment of CT imaging in both global structural distributions and cross-sectional images acquired at different orientations. Several studies have proposed phantoms for X-ray tire imaging evaluation, but these works primarily focused on 2D examples and did not support detailed analytical simulation [7], [8]. There remains a lack of dedicated tire phantoms that support fully analytical 3D simulation and systematic verification during the development of advanced CT tire imaging systems.

In parallel, there has been increasing interest in automatic X-ray tire inspection, particularly with the adoption of deep learning techniques. Recent studies focus on improving detection accuracy, achieving finer defect localization, and handling complex curved defect patterns [9], [10], [11]. Nevertheless, these approaches predominantly rely on real radiographs, which inherently lack precise ground truth. In addition, the limited diversity and volume of available datasets often constrain the performance and generalizability of such methods. A comprehensive and controllable tire phantom generation framework would therefore be beneficial for neural network training and quantitative task assessment, such as crack detection, material discrimination, and structural defect analysis.

In this work, continuing our previous work [7], we propose a comprehensive framework for 3D tire phantom design and imaging simulation using a widely used aviation tire as an example. The phantom was constructed using combinations of torus, cylinder, and cone primitives with clipping planes to represent complex geometric features. Additional embedded structures with varying shapes and material properties were incorporated to simulate realistic internal textures and heterogeneity. We further provide detailed geometric formulations and analytical line-integral calculations to enable accurate forward projection simulation. Reconstruction experiments were performed to demonstrate the feasibility of the proposed framework. The proposed method is scalable in both dimension and object types, enabling extension to different tire types and imaging scenarios.

II. Methods

A. Analytical 3D Phantom Construction

The widely used aviation tire H44.5×16.5R21 was selected as the reference for this study. The 3D tire phantom consists of a main body and multiple internal structures, represented as object 0 and objects 1-10, respectively, as shown in Fig.1. The main body represents the complex geometry of a 3D tire, while the internal structures enable the assessment of various tire inspection tasks. All objects were analytically modeled using combinations of torus, cylinder, and cone primitives.

The main tire body was composed of rubber and comprised the sidewall, tread, and bead regions (Fig.2A). Owing to the ring-shaped geometry of the tire, the construction followed a "shell" strategy. Each shell component was constructed by combining an outer primitive with a corresponding inner primitive assigned a negative linear attenuation coefficient (LAC). During forward projection, the line integral contributed by the inner primitive was subtracted from that of

J. Zhang and Z. Yu are with University of California, Santa Cruz. Z. Lin is with University of California, San Diego.

the outer primitive, effectively removing the interior region and forming a shell-like structure. For example, the curved sidewalls were modeled using a pair of torus shells, where the outer torus represented the external tire surface and an inner torus with negative LAC represented the hollow interior. The superposition of these two torus generated a torus shell with controllable thickness. The torus shell was further clipped along the axial direction such that only the side portions were retained, thereby forming the sidewall region of the tire. The central region between the sidewalls was replaced by a cylinder shell, representing the flattened tread region. The bead regions were constructed symmetrically in the inner central portion of the tire. Each bead region was formed by connecting a clipped cone shell (purple region in Fig.2A) with a clipped torus (yellow region in Fig.2A), ensuring a smooth geometric transition between the sidewall and the bead region. The cone shell was clipped along the axial directions to retain only the curved transition region, while the torus was clipped to preserve the rounded tail structure.

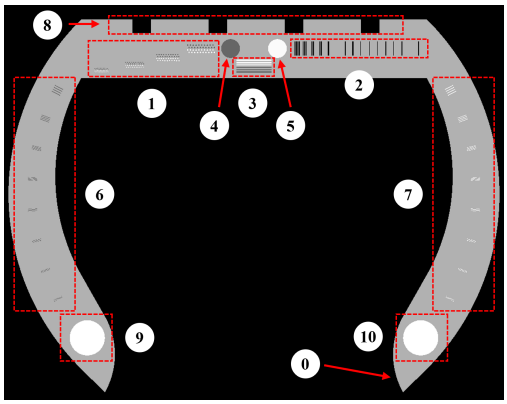


Fig. 1. Cross-sectional view of the internal structures, where red dashed rectangles highlight individual objects and white circles denote their corresponding indices.

For the internal structures (Fig.1), object 1 consisted of multiple torus groups with varying sizes, made of Nylon 66 and aramid, allowing the assessment of feature detectability at different scales. Object 2 contained a series of air resolution bar patterns, which were modeled by cylinder shells. Object 3 comprised multiple cylinder shells forming layered Nylon 66 and aramid structures to mimic the belt layers in realistic tires. Objects 4 and 5 were designed for contrast-to-noise evaluation and were modeled as torus structures composed of Nylon 66 and aramid, respectively. Objects 6 and 7 further extended spatial-resolution assessment into the sidewall regions and were modeled by cone shells, containing resolution bar patterns made of Nylon 66 and aramid, respectively. Object 8 was designed to mimic grooves embedded in the tread and was modeled by cylinder shells. In addition, optional bead-wire structures were provided as objects 9 and 10, modeled using torus primitives, but those objects were not included in the reconstruction experiments.

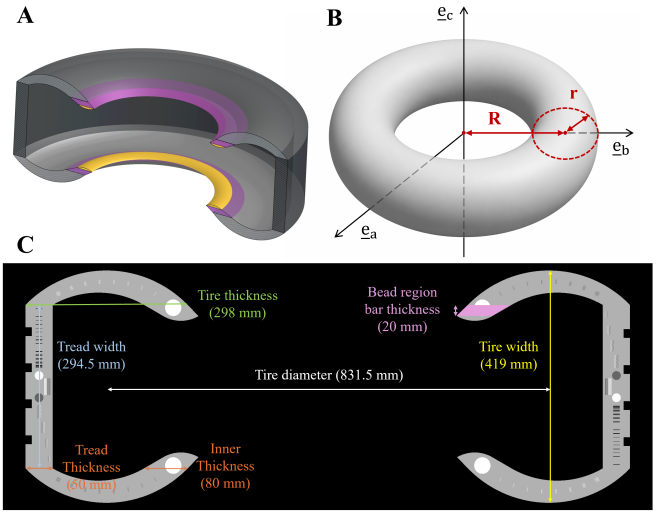


Fig. 2. Illustration of the 3D tire phantom. (A) Three-dimensional view of the main tire body, with a cross-sectional view showing the spatial distribution of the side wall (light gray), tread (dark gray), and bead regions (purple and yellow). (B) Geometric parameters and local coordinate system of the torus primitive. (C) Key geometric parameters in the tire phantom.

B. Mathematical Representation and Dimension Parameters

Torus is one of the most important primitives in the 3D tire phantom and is rarely incorporated in X-ray phantom simulation. A general torus centered at $\underline{x}_0$ is defined as

$$\left((\underline{x}-\underline{x}_0)^T(\underline{x}-\underline{x}_0)+R^2-r^2\right)^2 = 4R^2(\underline{x}-\underline{x}_0)^T M(\underline{x}-\underline{x}_0), \tag{1}$$

where R is the major radius, denoting the distance from the torus center to the centerline of the tube, and r is the minor radius, representing the radius of the tube cross-section (Fig.2B). The matrix M determines the orientation of the torus in 3D space:

$$M = Q^T D Q,\ Q = [\underline{e}_a, \underline{e}_b, \underline{e}_c]^T,\ D = \mathrm{diag}[1,1,0].$$

Here, $\underline{e}_a$ and $\underline{e}_b$ span the circular centerline plane of the torus, referred to as the torus plane in this study, while $\underline{e}_c$ is normal to this plane.

The cylinder and cone components are widely used geometries in CT phantom design and can be described by a unified quadratic formulation

$$(\underline{x}-\underline{x}_0)^T M (\underline{x}-\underline{x}_0) = c, \tag{2}$$

where $M = Q^T S Q$ defines the orientation and scaling of the primitive. For a cylinder,

$$c^{\mathrm{cy}} = 1, \qquad S^{\mathrm{cy}} = \mathrm{diag}\left[\left(\frac{1}{r_a^{\mathrm{cy}}}\right)^2, \left(\frac{1}{r_b^{\mathrm{cy}}}\right)^2, 0\right].$$

where r_a^{cy} and r_b^{cy} determine the cross-sectional dimensions perpendicular to the cylinder axis. For the cone,

$$c^{\mathrm{co}} = 0, \qquad S^{\mathrm{co}} = \mathrm{diag}\left[\left(\frac{1}{r_a^{\mathrm{co}}}\right)^2, \left(\frac{1}{r_b^{\mathrm{co}}}\right)^2, -\left(\frac{1}{r_c^{\mathrm{co}}}\right)^2\right].$$

where the parameters r_a^{co} and r_b^{co} control the spread of the surface in the transverse direction, while r_c^{co} governs the rate of variation along the axial direction. More implementation details for 2D examples are provided in [12].

The geometric parameters of the tire phantom were defined based on the reference tire (H44.5×16.5R21), while remaining fully adjustable to accommodate different tire types. Our previous work [7] provided a detailed definition of 2D tire phantom, which is extended in this study to 3D version. As shown in Fig.2C, the overall tire width was set to 419 mm, while the tire thickness was 298 mm. The nominal tire diameter was 831.5 mm, which determined the major radius of the outer torus. The tread region was characterized by a tread width of 294.5 mm and a tread thickness of 50 mm. The inner thickness was set to 80 mm, defining the internal hollow region and overall wall thickness distribution. Additionally, the bead region included a bead bar with a thickness of 20 mm. Based on these parameters, together with geometric continuity and smooth transitions between adjacent primitives, the main body of the tire phantom was fully defined. For internal structures, multiple torus groups in object 1 had tube radii ranging from 0.3 to 0.6 mm. The bar width of object 2 was inversely proportional to spatial frequency (8-38 lp/cm). Six cylindrical layers in object 3 had a uniform thickness of 1 mm, and inter-layer spacing of 2 mm. Objects 4 and 5 were torus disks with a fixed tube radius of 8 mm. Bar patterns in the sidewalls (objects 6 and 7) spanned spatial frequencies from 12 to 40 lp/cm with increments of 4 lp/cm. Groove structures (Object 8) had a fixed width of 16 mm, depth of 12 mm, and spacing of 65 mm between one another.

The LACs of the phantom components were assigned to represent three typical tire materials, including Nylon 66, aramid, tread rubber and metal. In addition, several regions were assigned air-equivalent LAC to simulate internal voids. Using the corresponding molecular compositions and mass densities, the LACs of these materials were calculated based on data from NIST [13], which could be further adjusted according to the selected X-ray tube and spectrum. In this study, a fixed effective energy of 74 keV was used.

C. Forward Projection and Reconstruction

Further experiments included generating forward projection data and validating the corresponding reconstructed results. In this study, a monochromatic x-ray beam was assumed for computing analytical line integrals. Accurate determination of the intersection points between X-rays and

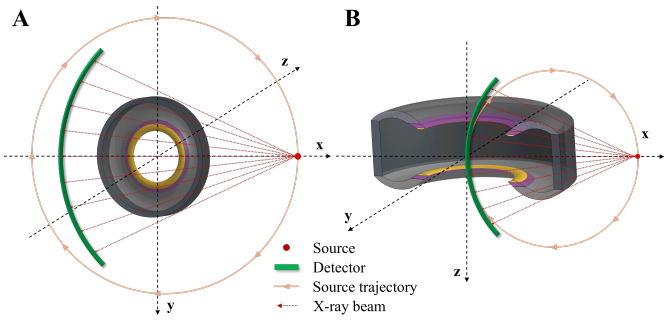


Fig. 3. Illustration of two forward projection geometries for the 3D tire phantom. In both cases, the X-ray source (red) follows a circular trajectory (orange), while a curved detector (green) collects the projected signals. (A) Scan geometry for full-tire imaging, in which x-rays cover the entire tire. (B) Geometry for detailed scanning of a single tire tube.

geometric surfaces is therefore essential for the forward projection. Each x-ray can be modeled parametrically as

$$\underline{x} = \underline{a} + t\underline{\beta}, \tag{3}$$

where $t \in [0, +\infty)$, $\underline{a}$ denotes the source position and $\underline{\beta}$ is a unit vector representing the ray direction. Intersections between rays and cylinder or cone primitives can be reduced to quadratic equations for t, which have at most two solutions. The corresponding solutions are described in [12]. In contrast, ray-torus intersection can yield up to four real solutions depending on the ray trajectory, corresponding to cases of no intersection, tangential contact, or multiple entry-exit crossings of the torus surface. Substituting Eq.3 into Eq.1 leads to a quartic equation

$$t^4 + At^3 + Bt^2 + Ct + D = 0. \tag{4}$$

Only real solutions were considered, and physically valid intersections must satisfy $t > 0$. By introducing the substitution $t = x - \frac{A}{4}$, the cubic term could be eliminated, resulting in the depressed quartic

$$x^4 + Ex^2 + Fx + G = 0. \tag{5}$$

The expressions of coefficients A through G are provided in Appendix. The solution of this equation follows the standard approach of reducing it to the product of two quadratic polynomials. When $F = 0$, the equation reduces directly to a biquadratic form and can be solved straightforwardly. For $F \neq 0$, an auxiliary variable α was introduced, and the equation could be rewritten as

$$(x^2 + \alpha)^2 - [(2\alpha - E)x^2 - Fx + (\alpha^2 - G)] = 0. \tag{6}$$

If the second term is a perfect square, the equation could be factorized into two quadratic polynomials,

$$\begin{aligned}&\left(x^2 - \sqrt{2\alpha - E}\,x + \alpha + \frac{F}{2\sqrt{2\alpha - E}}\right)\\ &\times\left(x^2 + \sqrt{2\alpha - E}\,x + \alpha - \frac{F}{2\sqrt{2\alpha - E}}\right) = 0,\end{aligned} \tag{7}$$

from which the roots can be obtained. Determining a suitable α reduces to solving an auxiliary cubic equation, as detailed in Appendix. A valid solution must satisfy $2\alpha - E > 0$, ensuring that the square-root term is well defined. The valid α provides the roots of the quartic equation, which are then mapped back to the original ray parameter using $t = x - \frac{A}{4}$.

Notably, all computations were performed in a discrete numerical setting. A strict separation of cases based on $F = 0$ and $F \neq 0$ may introduce discontinuities due to floating-point errors, leading to inaccuracies in intersection calculations. To improve numerical stability, branch conditions were evaluated using scaled tolerances rather than exact comparisons. In addition, the terms F and $2\alpha - E$ may approach zero simultaneously, and numerical instability can arise when $2\alpha - E$ converges to zero faster than F. To address this issue, an equivalent expression derived by equating Eq.5 and Eq.6, $\text{sign}(F\sqrt{\alpha^2 - G})$, was used to replace the unstable term, $F/(2\sqrt{2\alpha - E})$, to avoid large artificial values caused by division by a very small number.

After selecting valid roots, an interior test was performed by evaluating the midpoint between consecutive intersections to determine whether each ray segment lay inside the torus. By collecting all valid ray segments across objects of different shapes, the line integral for each projection was computed as the sum of segmented lengths weighted by the corresponding LAC along the x-ray path.

We performed the simulation using two protocols corresponding to the two scan geometries shown in Fig.3. The curved detector was used in this study, and one-dimensional projection data were collected at each view. To image the cross-section in the torus plane, the geometry illustrated in Fig.3A was employed, with a source-to-isocenter and source-to-detector distance of 1000 mm and 2000 mm, respectively. For this protocol, a detector pixel size of 1.05 mm, detector channel number of 4065, and 3600 views per rotation were used. Another scan geometry was designed to scan a single cross section of the tire tube (Fig.3B). In this configuration, the source-to-detector distance was set equal to the tire diameter, 831.5 mm, and the source-to-isocenter distance was 415.75 mm. In this protocol, the detector pixel size and pixel number were set to 0.44 mm and 3065, respectively, and 3600 views per rotation were acquired. To further improve numerical accuracy and reduce aliasing, each projection was computed by first up-sampling the detector grid with a factor of 3 and then averaging back the high-resolution projections to the original detector. Image reconstruction was then performed using filter back-projection (FBP) algorithm [14]. The image size was 1024×1024 and field-of-view (FOV) was 1280×1280 mm^2 and 427×427 mm^2 for scan A and B, respectively. The reconstructed LACs were converted to Hounsfield Units (HU) using water as the reference material, with a water LAC of 0.0184 mm^{-1}.

III. Results

The sinograms of the tire phantom generated using the proposed forward projection under the two scan geometries are shown in the top row of Fig.4. The corresponding reconstructed cross-sectional images are presented in the second row. The last row displays the difference images obtained by subtracting the phantom from the reconstruction results. The reconstruction error profiles extracted from different locations for the two scan geometries (top: scan A and bottom: scan B) are shown in Fig.5.

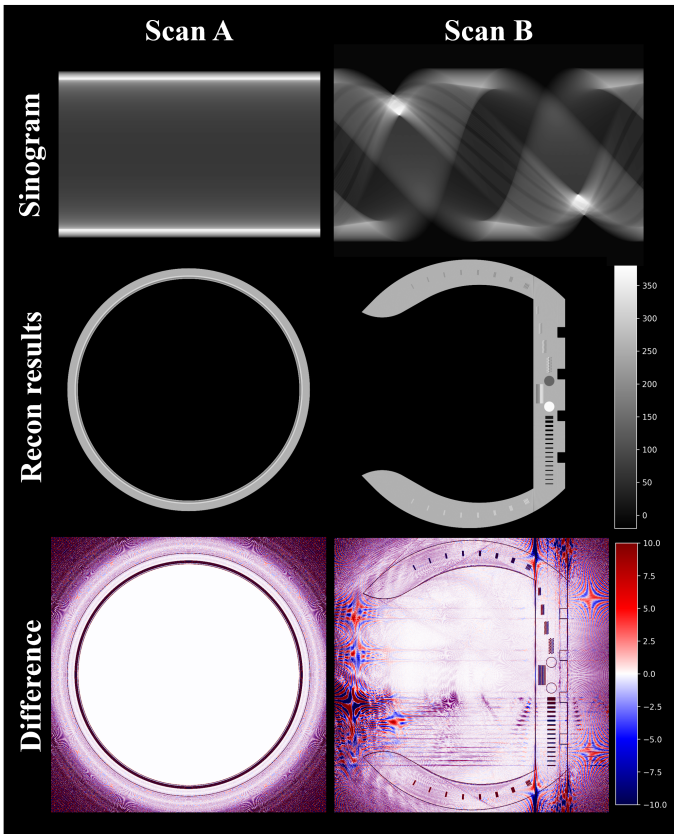


Fig. 4. Forward projection and reconstruction results of the tire phantom under two scan geometries. Scan A corresponds to imaging in the torus plane, while scan B corresponds to imaging the torus tube cross-section. The rows show the simulated sinogram, reconstructed image, and the difference between phantom and reconstruction results, respectively. Display: L/W = 180/400 HU.

Scan A images the cross-section in the torus plane, where the reconstructed image mainly captures the global tire body together with the layered structure of object 3. The reconstructed image demonstrates close agreement with the simulated phantom, and the corresponding difference image shows that reconstruction errors are primarily concentrated near object boundaries. These difference are primarily attributed to finite detector sampling spacing and limited sampling rate. Most errors remain within approximately ±0.5 HU (LAC differences on the order of 10^{-6}), indicating high reconstruction accuracy in homogeneous regions.

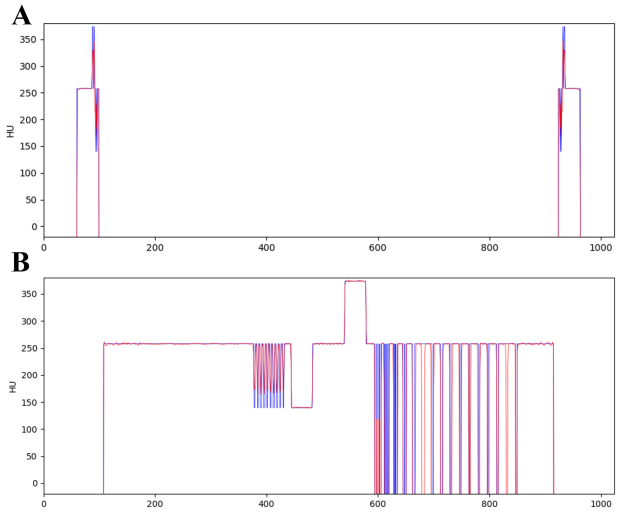


Fig. 5. HU profiles extracted from phantom slices and reconstructed results for two scan protocols. The blue and red curves represent the ground truth and reconstructed results, respectively. (A) Profile corresponding to scan A, extracted from the middle row of the reconstructed image. (B) Profiles corresponding to scan B, extracted from the column passing through the middle of the tread region.

In contrast, scan B images the torus tube cross-section, covering all embedded objects within the tire phantom. The reconstructed image successfully resolves the major geometric features and the internal objects. For some extremely high-resolution structures, finer detector sampling or more projection reviews are required to fully recover the fine details. Compared with scan A, the error profile exhibits stronger local fluctuations due to the increased structural complexity and sharp transitions between materials. Nevertheless, the majority of smooth regions still maintain errors near ± 0.5 HU, while larger deviations are mainly localized around fine-scale structures and object boundaries.

IV. Discussion

The tire phantom was designed to mimic the aviation tire H44.5×16.5R21 for the development of tire-related CT systems in terms of geometric structure and material composition. Globally, the phantom consisted of multiple combinations of torus, cylinder and cone geometries with clipping planes. Several embedded objects with different shapes and materials were incorporated to evaluate various imaging characteristics relevant to tire inspection. For the torus, which is not a commonly used geometry in conventional CT phantom design, detailed geometric formulations and corresponding line-integral calculations were provided. To address different inspection requirements, forward projections of the phantom were evaluated under two scanning protocols for global tire imaging and internal structure imaging, respectively. Reconstruction experiments were then performed to assess the effectiveness of the proposed simulation pipeline.

The reconstruction results demonstrated that the proposed analytical framework can accurately simulate the forward projection of complex tire structures. Under both scan geometries, the reconstructed images showed strong agreement with the simulated phantom, successfully capturing both global tire geometry and fine internal structures, while reconstruction errors in homogeneous regions remained within approximately ± 0.5 HU. Although this work considers only three geometric types, the framework can be readily extended to incorporate other quadratic geometries, such as ellipsoids and hyperboloids, and is inherently scalable with respect to dimensionality, shape complexity, and material composition.

Several aspects will be investigated in future work. First, the current implementation assumes a monochromatic x-ray beam. Further studies will explore forward projection simulation under realistic polychromatic x-ray spectra. Second, due to the computational complexity of ray-torus intersection calculations and the need for dense sampling, the current forward projection process is computationally intensive. To improve computational efficiency, hardware acceleration using GPU-based implementations will be explored.

V. Conclusion

In this work, we developed an analytical 3D tire phantom and derived the corresponding forward projection, both validated through reconstruction experiments. The results demonstrate the feasibility of the proposed framework, which can be readily extended to other tire types and imaging configurations.

Acknowledgments

The authors would like to thank Frederic Noo for helpful discussions.

Appendix
The ray-torus intersection calculation

To compute the intersection between a ray and a torus, Eq.3 was substituted into Eq.1, which reduces to

$$t^4 + At^3 + Bt^2 + Ct + D = 0,$$

$$\begin{cases} A = 4(\underline{a} - \underline{x}_0)^T \underline{\beta} \\ B = 4\left((\underline{a} - \underline{x}_0)^T \underline{\beta}\right)^2 + 2\left(\|\underline{a} - \underline{x}_0\|^2 + R^2 - r^2\right) \\ \quad - 4R^2 \underline{\beta}^T M \underline{\beta} \\ C = 4\left(\|\underline{a} - \underline{x}_0\|^2 + R^2 - r^2\right)(\underline{a} - \underline{x}_0)^T \underline{\beta} \\ \quad - 8R^2 (\underline{a} - \underline{x}_0)^T M \underline{\beta} \\ D = \left(\|\underline{a} - \underline{x}_0\|^2 + R^2 - r^2\right)^2 - 4R^2 (\underline{a} - \underline{x}_0)^T M (\underline{a} - \underline{x}_0). \end{cases} \tag{8}$$

A variable substitution, $t = x - \frac{A}{4}$, was first applied to eliminate the cubic term, which converts the original quartic into a depressed quartic equation,

$$x^4 + Ex^2 + Fx + G = 0,$$

$$\begin{cases} E = B - \frac{3}{8}A^2 \\ F = C - \frac{1}{2}AB + \frac{1}{8}A^3 \\ G = D - \frac{1}{4}AC + \frac{1}{16}A^2B - \frac{3}{256}A^4. \end{cases} \tag{9}$$

This depressed quartic equation was solved in two cases: $F = 0$ and $F \neq 0$. When $F = 0$, the equation reduces to a biquadratic form and can be solved as a quadratic equation in x^2 as

$$(x^2)^2 + E(x^2) + G = 0. \tag{10}$$

When $F \neq 0$, an auxiliary variable α was introduced to transform the quartic equation into a structure that can be factored as shown below,

$$(x^2 + \alpha)^2 - [(2\alpha - E)x^2 - Fx + (\alpha^2 - G)] = 0. \tag{11}$$

If the second term $(2\alpha - E)x^2 - Fx + (\alpha^2 - G)$ is a perfect square, the quartic equation can be factorized into the product of two quadratic polynomials and solved futher.

To make the second term of Eq.11 a perfect square, the variable α must satisfy a cubic equation as

$$\alpha^3 - \frac{E}{2}\alpha^2 - G\alpha + \frac{4EG - F^2}{8} = 0. \tag{12}$$

Another variable substitution $\alpha = \upsilon + \frac{E}{6}$ was applied to yield a depressed cubic equation as

$$\upsilon^3 + H\upsilon + L = 0,$$

$$\begin{cases} H = -\frac{E^2}{12} - G \\ L = -\frac{E^3}{108} - \frac{F^2}{8} + \frac{EG}{3}. \end{cases} \tag{13}$$

The number of real solutions of this cubic is determined by its discriminant:

$$\Delta_{cubic} = \left(\frac{L}{2}\right)^2 + \left(\frac{H}{3}\right)^3. \tag{14}$$

If $\Delta_{cubic} > 0$, the cubic equation only has one solution, which is

$$\upsilon^* = \sqrt[3]{-\frac{L}{2} + \sqrt{\Delta_{cubic}}} + \sqrt[3]{-\frac{L}{2} - \sqrt{\Delta_{cubic}}} \tag{15}$$

If $\Delta_{cubic} = 0$ and $H = 0$, the cubic equation also has one solution, which is zero. If $\Delta_{cubic} = 0$ but $H \neq 0$, it has two solutions:

when $L - 2\left(\sqrt{-\frac{H}{3}}\right)^3 = 0$,

$$\upsilon_1^* = \sqrt{-\frac{H}{3}} \quad \text{and} \quad \upsilon_2^* = -2\sqrt{-\frac{H}{3}}; \tag{16}$$

when $L + 2\left(\sqrt{-\frac{H}{3}}\right)^3 = 0$,

$$\upsilon_1^* = -\sqrt{-\frac{H}{3}} \quad \text{and} \quad \upsilon_2^* = 2\sqrt{-\frac{H}{3}} \tag{17}$$

If $\Delta_{cubic} < 0$, the cubic equation has three solutions as follows,

$$\begin{aligned} \upsilon_i^* &= 2\sqrt{-\frac{H}{3}}\cos\left(\frac{\theta + 2\pi i}{3}\right),\ i = 0, 1, 2 \\ \theta &= \arccos\left(\frac{3L}{2H}\sqrt{-\frac{3}{H}}\right). \end{aligned} \tag{18}$$

After obtaining the candidate values for υ, the corresponding α values were recovered. A valid α must satisfy the condition $2\alpha - E > 0$, ensuring that the quadratic factorization is well-defined. Substituting the selected α into Eq.11, the depressed quartic can be factorized into two quadratic polynomials (Eq.7). The roots of these quadratic equations provide the solutions of the quartic, which are then transformed back to the original ray parameter using $t = x - \frac{A}{4}$. Finally, only roots satisfying $t > 0$ are retained as valid ray-torus intersection points.

REFERENCES

[1] Continental Tyres, "X-ray and CT tyre analysis," 2026.

[2] Hamamatsu Photonics, "Industrial x-ray inspection for tire quality control," 2026.

[3] The Goodyear Tire and Rubber Company, *Aircraft Tire Care and Maintenance*, 2024. Accessed: July 2025,https://www.goodyearaviation.com/resources/tirecare.html.

[4] C. Cai and L. He, "Improved mach–zehnder interferometer-based shearography," *Optics and Lasers in Engineering*, vol. 50, no. 12, pp. 1699–1705, 2012.

[5] N. Trivisonno, "Nondestructive evaluation of tires," *Rubber chemistry and technology*, vol. 58, no. 3, pp. 469–526, 1985.

[6] B. Rodgers, *Tire engineering: an introduction*. CRC press, 2020.

[7] Z. Yu, R. Zahn, A. Xu, A. Jain, and J. He, "Can a medical imaging x-ray tube be used for radial aviation tire inspection?," in *15th Conference on Industrial Computer Tomography, Linz, Austria*, 2026.

[8] L. Sun, L. He, C. Hai, X. Han, Z. Gui, and M. Yang, "Design of imaging system and tomography detection method for radial tires structure under x-ray short-scan mode," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[9] G. Zhao and S. Qin, "High-precision detection of defects of tire texture through x-ray imaging based on local inverse difference moment features," *Sensors*, vol. 18, no. 8, p. 2524, 2018.

[10] Z. Zheng, S. Zhang, B. Yu, Q. Li, and Y. Zhang, "Defect inspection in tire radiographic image using concise semantic segmentation," *Ieee Access*, vol. 8, pp. 112674–112687, 2020.

[11] G. Xu, A. Li, X. Wang, C. Xu, J. Chen, and F. Zheng, "Research on x-ray nondestructive defect detection method of tire based on dynamic snake convolution yolo model," *Scientific reports*, vol. 14, no. 1, p. 29587, 2024.

[12] Z. Yu, F. Noo, F. Dennerlein, A. Wunderlich, G. Lauritsch, and J. Hornegger, "Simulation tools for two-dimensional experiments in x-ray computed tomography using the FORBILD head phantom," *Physics in Medicine & Biology*, vol. 57, no. 13, p. N237, 2012.

[13] https://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html. Accessed: 2025-12-16.

[14] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*. SIAM, 2001.

An iterative filtered-backprojection algorithm to exactly and stably solve the long object problem in cone-beam CT with source positions evenly distributed on a cylinder

Murdock G. Grewar, Frédéric Noo

***Abstract*—A filtered backprojection (FBP) method is presented for reconstruction in transmission cone-beam Computed Tomography (CBCT). This method applies when the X-ray source points are evenly distributed on a cylinder, with an otherwise conventional imaging geometry similar to helical scanning. The cone-beam projections are assumed to be axially truncated. The required filtration of the cone-beam projections is a spatially-varying 2D convolution, which is computationally challenging to complete in an efficient manner. However, accurate results can be achieved by choosing a random 1D filtration for each cone-beam projection, resulting in an FBP algorithm that requires a simple 1D convolution as the filtration operation. This algorithm is iterated to ameliorate the errors introduced by this approximation and by the finite source point sampling. Each iteration requires one full backprojection of the data, one full projection of the reconstruction volume, and a second projection which may be computed at lower resolution. Iteration time is thus similar to conventional methods such as OS-SIRT, and the computation may be similarly parallelised on modern GPU hardware in a familiar manner. Herein, a comparison is made between axial region-of-interest reconstructions of the FORBILD head phantom using this iterated FBP method, and using OS-SIRT with appropriately tuned hyperparameters.**

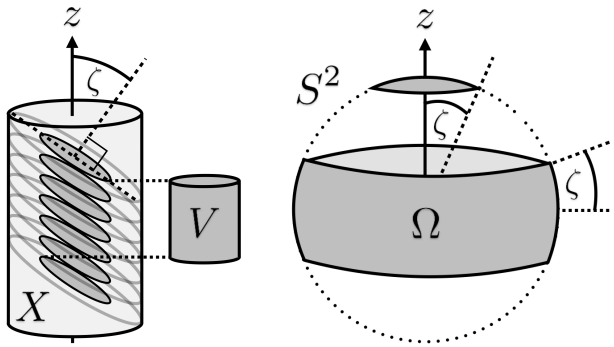
I. Introduction

Cone-beam (CB) CT imaging typically concerns datasets acquired with source points sampled along a curve. This abstract concerns CBCT imaging with source points sparsely sampled on the surface of a cylinder. These types of CT scans have been proposed to reduce directional metal streaking artefacts and increase resolution isotropy [1]; they are well-suited for nondestructive materials evaluation, but they may play an important role in medical imaging in the future.

The first author has previously posted on arXiv [2], [3] results pertaining to an exact method of reconstruction for CBCT datasets acquired with source points sampled on a cylinder and other surfaces. That method is called 'global backprojection-convolution' (GBC). GBC suffers from the 'long object problem' in the sense that the entire object needs to be scanned for an exact reconstruction to be achieved on any region of interest.

Here, we present a filtered-backprojection (FBP) method for reconstruction with source points sparsely distributed on the surface of a cylinder. This method allows theoretically exact and stable reconstruction of axial regions of interest due to its FBP structure. The underlying formula includes a computationally-demanding spatially-varying convolution. To circumvent this issue, we use an iterative implementation that replaces the spatially-varying convolution with a randomly-chosen spatially-invariant 1D convolution on each cone-beam projection.

M. Grewar is with the Department of Materials Physics, Australian National University, Australia. F. Noo is with the Department of Radiology and Imaging Sciences, University of Utah, USA.



(a) (b)

Fig. 1: The source point cylinder, X, is sliced obliquely into a stack of ellipses. Each ellipse admits a 2D FBP formula for its interior. **(a):** X, as a stack of ellipses. Our region of reconstruction, V, is depicted. The radius of V depends on the width of the detector. **(b):** depiction of the Orlov sphere. The maximum polar angle ζ is determined by the range of measurement lines captured in the experiment. V is defined as that region of the interior of X which receives measurements from all angles in Ω. See also fig. 3 for an example.

II. Theoretical basis for the algorithm

There is insufficient room in this abstract to work through the derivation of the underlying formula, but the basic ideas can be outlined. The formula is given in (1). It can be rigorously derived in a similar manner to the approach in [4]. We will now describe the basic ideas underlying the derivation.

We begin by conceptualising the cylinder as a union of ellipses, of varying orientations (most sections of a cylinder are ellipses). Each one of these ellipses is amenable to an exact 2D FBP formula for reconstruction for the interior of

that ellipse. See fig. 1a. To keep the formula simple for this abstract, we restrict the region of reconstruction—call it V—to a smaller set than is strictly necessary. We choose V as the set of points that receive measurements through all angles in Ω, illustrated in fig. 1b.

Let $v \in V$. The attenuation coefficient at v may be recovered by considering any of the source ellipses that is coplanar with v (i.e. contains v in its interior), and applying a 2D FBP formula within the plane. Due to the way we have defined V, we are guaranteed the existence of a coplanar source ellipse, so long as the normal vector of the ellipse has a polar angle less than ζ (see fig. 1). Therefore, we have an FBP formula for reconstructing the whole volume V in the following manner: fix a normal vector to be used for all ellipses, and recover the attenuation coefficient at v by applying the 2D FBP formula for the ellipse coplanar with v and the given normal vector.

However, two practical problems emerge when attempting to formulate the 3D reconstruction in this manner. The first problem is that if we use ellipses with only a single orientation (e.g. a fixed normal vector), then the reconstruction formula for V uses only a single line from each cone-beam projection. (See fig. 2.) This is very data-inefficient, and obviously not practical. This is resolved by integrating the 3D FBP formula over every allowable ellipse orientation.

It is here that a second practical barrier emerges: although the resulting 3D formula is now data-efficient, using most lines on the detector[1], the formula requires a 2D spatially-varying convolution on each cone-beam projection. Spatially-varying convolutions are computationally expensive. It is here that we introduce a practical compromise, which involves iterating our method—iterating helps with other sources of error anyway, such as the finite source point sampling. This compromise allows us to replace the 2D spatially-varying convolution with a simple 1D convolution.

Mathematically, this reformulation of the method is achieved by reparameterising the integration over ellipse orientations into a 1D integration over detector rotation angles, expressed by the rotation coordinate $\theta_{\text{rot}} \in [-\zeta, \zeta]$. The 2D convolution is thus expressed as an integral of 1D convolutions, as visible in (1). Then, instead of integrating over the 1D convolutions, we simply pick a randomly uniformly sampled rotation angle $\theta_{\text{rot}} \in [-\zeta, \zeta]$, independently for each cone-beam projection and each iteration of the method. (See fig. 2.) This significantly reduces the computational burden, while maintaining good accuracy. In fact, in the limit of a large number of source points, the method becomes exact within a single iteration, despite the random sampling of rotation angles.

That is the basic foundation of the iterated FBP algorithm. Pseudocode is provided in alg. 1. Illustrations of the steps are provided in figs. 4 and 5. Following the convention in [5], the detector coordinates $(Y, X$ in fig. 2) are measured in the midplane, i.e. measured as if the detector were located inside the object, intersecting the z axis.

[1]It uses the lines which are contained entirely within the Colsher window.

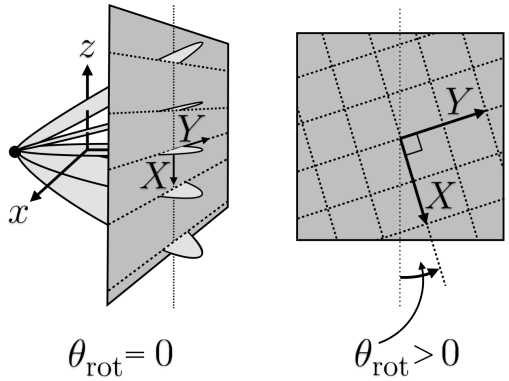


Fig. 2: Illustration of how the rotated lines of the detector contribute to different reconstruction ellipses. Each detector line, at each rotation angle $\theta_{\text{rot}} \in [-\zeta, \zeta]$, contributes to the ellipse of the reconstruction support which is coplanar with both that detector line and the respective source point. By combining the solutions obtained from a range of ellipse orientations, we are able to efficiently use the detector data.

III. Evaluation with simulated data

We demonstrate the performance of our method with a reconstruction of an axial region-of-interest of the FORBILD head phantom [6], designed to mimic a human head. The phantom extends above and below the source cylinder, as illustrated in fig. 3. To keep the FBP formula simple for this abstract, we have selected the reconstruction support as a cylinder through which projections in all directions Ω are captured.[2] We generated $5,000$ cone-beam projections with Poisson noise based on a monochromatic X-ray source of energy $\approx 80\text{keV}$ and isotropic intensity, with a clearfield photon number of approximately $250,000$ photons per pixel at the centre of the detector. (Each point in the reconstruction support is seen by about $2,000$ projections, matching the predicted number required for data sufficiency in this experiment.) The detector resolution was 667×100 pixels. The detector pixel size at field-of-view centre was $0.50\text{mm} \times 0.50\text{mm}$. The X-ray source points were distributed around the object in a low-discrepancy sequence [2], [7], as depicted in fig. 6.

We compared our iterated FBP reconstruction method with a cone-beam version of the OS-SIRT [8] algorithm. The OS-SIRT algorithm is a generalisation of SART [9]. SART is a Kaczmarz method which proceeds through each cone-beam projection, each time orthogonally projecting the reconstructed volume onto the subspace of solutions compatible with the measurements in that cone-beam projection. The OS-SIRT algorithm is similar, except that it averages the correction within subsets of projections. This provides stability in the presence

[2]The formula can be extended to reconstruct on the full support, which is a cylinder with a height equal to that of the source cylinder; this is a subject for future study.

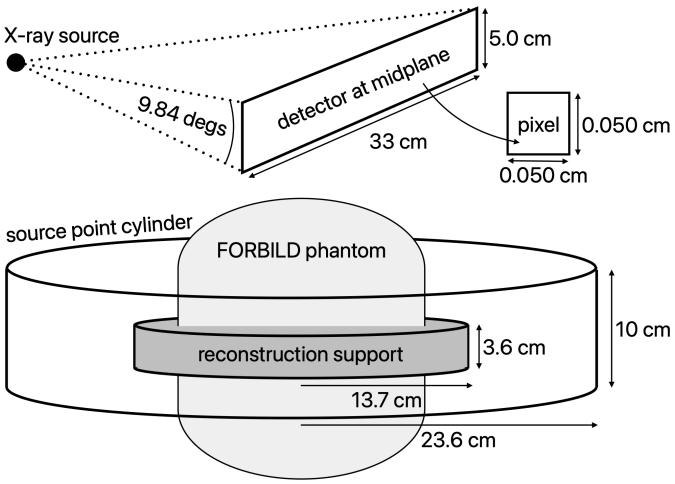


Fig. 3: Schematic of the simulated CT scan. The maximum polar angle of the ellipses used for reconstruction is $\zeta \approx 9.84^\circ/2 \approx 0.086\,\text{rad}$.

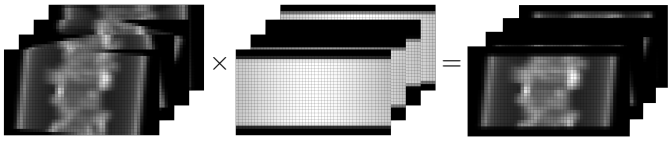


Fig. 4: The rotated projections **(a)** are multiplied by the prefiltration factor **(b)**. This factor depends on the rotation angle θ_i. The product of the multiplication is shown in **(c)**.

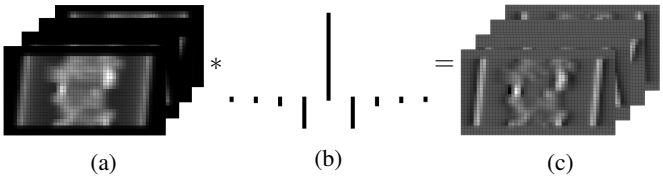


Fig. 5: The 'premultiplied' rotated projection data **(a)** (same as fig. 3c) are linearly discretely convolved with a band-limited, discretised ramp filter kernel **(b)** along the horizontal axis. The product of the convolution is depicted in **(c)**. The projections are ready to be backprojected, with an appropriate weighting, after rotating them back to their original orientation.

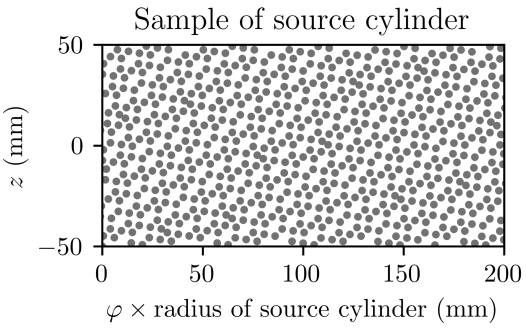


Fig. 6: A small window of the source points used in generating the simulated cone-beam projection data. The cylinder is much wider than it is tall, so it has been cropped to show only part of the angular range.

Algorithm 1 *Iterated filtered backprojection*

Require: $N \in \{1, 2, 3, \dots\}$ ▷ number of source points
Require: $\mathbf{m} \in$ projections ▷ linearised measurement data
Require: $\zeta \in [0, \pi/2]$
initialization
  $\mathbf{x} \leftarrow \underline{0}$ ▷ initialise voxelised volume with zeroes
end initialization
procedure ITERATE( )
  $\mathbf{y} \leftarrow \underline{0}$ ▷ initialise a voxelised volume with zeroes
  for $i = 1$ to N **do**
    $r_i \leftarrow m_i - \texttt{proj}(\mathbf{x})_i$ ▷ projection residual
  `/* Copy the 2D projection residual, and filter it*/`
    $q_i \leftarrow r_i$
    $\theta_i \leftarrow \texttt{random_uniform}(-\zeta, \zeta)$
    $q_i \leftarrow \texttt{rotate}(q_i, \theta_i)$
    $q_i \leftarrow \texttt{multmask2D}(q_i, \theta_i)$ ▷ see fig. 4
    $q_i \leftarrow \texttt{rampFilterHorz}(q_i)$ ▷ see fig. 5
    $q_i \leftarrow \texttt{rotate}(q_i, -\theta_i)$
  `/* Backproject the filtered 2D projection residual */`
    $\mathbf{y} \leftarrow \mathbf{y} + \texttt{weightedBackproj}(q_i, i)$
  end for
  `/* Calculate correction coefficient to minimise the 1^2 projection residual of the next volume. Note: alpha is forced to 1 for the first iteration.*/`
  $\alpha_n \leftarrow 0$
  $\alpha_d \leftarrow 0$
  for $i = 1$ to N **do**
    $q_i \leftarrow \texttt{proj}(\mathbf{y})_i$
    $\alpha_n \leftarrow \alpha_n + r_i \cdot q_i$
    $\alpha_d \leftarrow \alpha_d + q_i \cdot q_i$
  end for
  $\alpha \leftarrow \texttt{max}(0, \texttt{min}(1, \alpha_n/\alpha_d))$ ▷ converged: $\alpha = 0$
  `/* Update the volume */`
  $\mathbf{x} \leftarrow \mathbf{x} + \alpha\mathbf{y}$
end procedure

of noise, but requires more iterations to achieve full sharpness. We averaged the correction from subsets of 25 projections each. In each iteration of OS-SIRT, we proceeded through all the projections once, in the order of the low-discrepancy sequence, thereby increasing the rate of convergence. We allowed 10 iterations of OS-SIRT, in order to visually match the sharpness achieved by the iterated FBP method, which converged in 4 iterations.

Reconstructions were both performed with voxels of volume $0.25\text{mm} \times 0.25\text{mm} \times 0.25\text{mm}$; subsequently, the reconstructions were downbinned in a $2 \times 2 \times 2$ fashion, producing a final volume with voxels of volume $0.50\text{mm} \times 0.50\text{mm} \times 0.50\text{mm}$. Slices of the final volumes are depicted in fig. 7. It is seen that the iterated FBP method produced similar results to OS-SIRT in the $z = 0$ slice, in half the number of iterations. The FBP

Let $\mathbf{r} \in V$ be given in cylindrical polar coordinates as (r_z, r_ϕ, r_ρ). Let $P_{(z,\varphi)}(Y,X)$ be the line attenuation from source point (z,φ), along the line intersecting the (Y,X) point of the detector midplane (see fig. 2). The attenuation coefficient at $\mathbf{r}$ is

$$f(\mathbf{r}) = \frac{\csc^2(\zeta/2)}{32\pi^3 R^2} \int_0^{2\pi} R \mathrm{d}\varphi \frac{R^3}{Z_{\mathbf{r}}^3} \int_{z_{\min}}^{z_{\max}} \mathrm{d}z \, FP_{(z,\varphi)}(Y_{\mathbf{r}}, X_{\mathbf{r}}) \,, \tag{1}$$

where

$$FP_{(z,\varphi)}(Y,X) = \int_{-\zeta}^{\zeta} \mathrm{d}\theta \, \mathrm{Rot}_{-\theta} \left\{ \mathrm{Ramp}_{1\mathrm{D}} \left\{ \mathrm{Mult}_{M_\theta} \left\{ \mathrm{Rot}_\theta \left\{ P_{(z,\varphi)} \right\} \right\} \right\} \right\} (Y,X),$$

$$z_{\min} = r_z - (R + r_\rho)\tan\zeta \,, \quad z_{\max} = r_z + (R + r_\rho)\tan\zeta \,, \quad \mathrm{Mult_a}\left\{\mathrm{b}\right\}(Y,X) = \mathrm{a}(Y,X) \times \mathrm{b}(Y,X)$$

$$Y_{\mathbf{r}} = -\frac{R}{Z_{\mathbf{r}}} \sin(r_\phi - \varphi) \,, \quad X_{\mathbf{r}} = -\frac{R}{Z_{\mathbf{r}}}(r_z - z) \,, \quad \mathrm{Rot}_\theta \left\{\mathrm{func}\right\}(Y,X) = \mathrm{func}(Y\cos\theta + X\sin\theta, X\cos\theta - Y\sin\theta)$$

$$Z_{\mathbf{r}} = R - r_\rho \cos(r_\phi - \varphi) \,, \quad \mathrm{Ramp}_{1\mathrm{D}} \left\{\mathrm{func}\right\}(Y,X) = \int_{-\infty}^{\infty} \mathrm{d}Y' \int_{-\infty}^{\infty} \mathrm{d}\omega |\omega| \exp\left\{i\omega(Y - Y')\right\} \mathrm{func}(Y', X) \,,$$

$$M_\theta(Y,X) = \Theta\left(\left(\cos^2\theta \sec^2\zeta - 1\right) - X^2/R^2\right)\left(1 + X^2/R^2\right)^{-3/2} \left(\sqrt{1 + Y^2/R^2 + X^2/R^2}\right)^{-1} .$$

The Heaviside step function is defined as $\Theta(x) = 1$ if $x > 0$, and $\Theta(x) = 0$ otherwise.

reconstruction has more significant streak artefacts emanating from the ear cavities. The FBP reconstruction appears to resolve the eyes and ventricle with greater conspicuity. The OS-SIRT method suffers from severe artefacts on the axial extremities of the region of interest, seen clearly in the $x = 0$, $y = 0$ and $z = 8\,\mathrm{mm}$ slices. As expected, the FBP reconstruction is capable of reconstructing the entire axial region of interest, though with some loss of quality toward the axial extremities (seen in the $z = 8\,\mathrm{mm}$ slice). The FBP method is capable of axial region-of-interest reconstruction because it expresses the attenuation coefficient as a function of measurement data (1), and depends only on (rotated) detector lines that include measurements taken through V.

We used zeroth-order interpolation for all projections and backprojections. Both of these methods require the equivalent of one forward projection operation and one backprojection operation, per iteration. The iterated FBP method also requires a second projection to determine the optimum value of the correction factor α (see alg. 1), but this may be computed at a lower resolution and with a stochastic estimation element if desired, since the precise value of α has little effect. The remaining operations (including the 1D convolution of the FBP method) are relatively inexpensive. Therefore, one iteration of either of these algorithms carries a comparable computational load. Both algorithms were GPU-accelerated.

IV. Conclusion

We have demonstrated a practical implementation for an iterated FBP method of reconstruction in CBCT which applies to scans with source points uniformly distributed on the surface of a cylinder. The method is theoretically exact and stable. It allows rapid and reliable axial region-of-interest reconstructions from axially-truncated cone-beam projections.

References

[1] A. M. Kingston, G. R. Myers, S. J. Latham, B. Recur, H. Li, and A. P. Sheppard, "Space-filling x-ray source trajectories for efficient scanning in large-angle cone-beam computed tomography," *IEEE Transactions on Computational Imaging*, vol. 4, no. 3, pp. 447–458, 2018.

[2] M. Grewar, G. Myers, and A. Kingston, "Exact inversion from space-filling trajectories in cone-beam transmission tomography," 2024.

[3] M. G. Grewar, "Fourier-based inversion of partial x-ray transforms in n dimensions," 2025.

[4] J.-M. Wagner and F. Noo, "Three-dimensional image reconstruction from exponential parallel-beam projections," *IEEE Transactions on Nuclear Science*, vol. 48, no. 3, pp. 743–749, 2001.

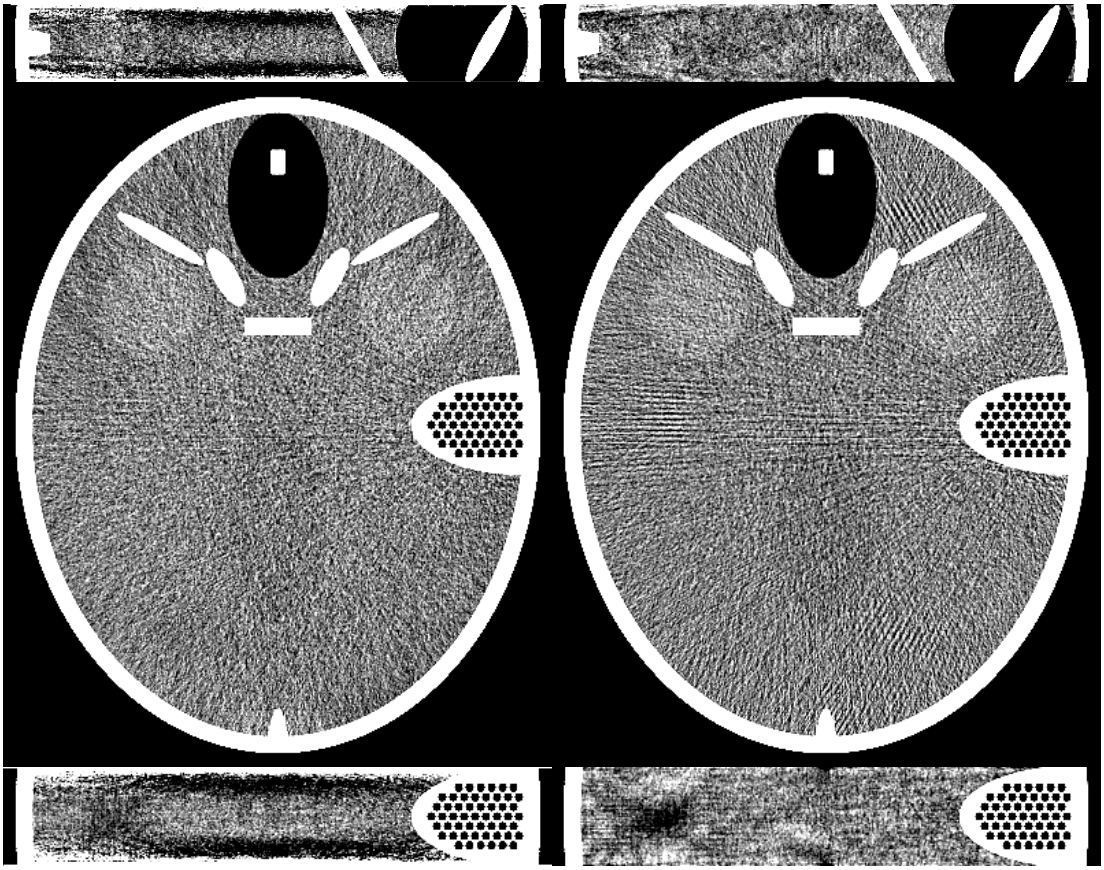
[5] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Josa a*, vol. 1, no. 6, pp. 612–619, 1984.

[6] G. Lauritsch and H. Bruder, "Head phantom," *University of Erlangen*, June 2007. [Online]. Available: http://www.imp.uni-erlangen.de/forbild/english/results/doc_phan.htm

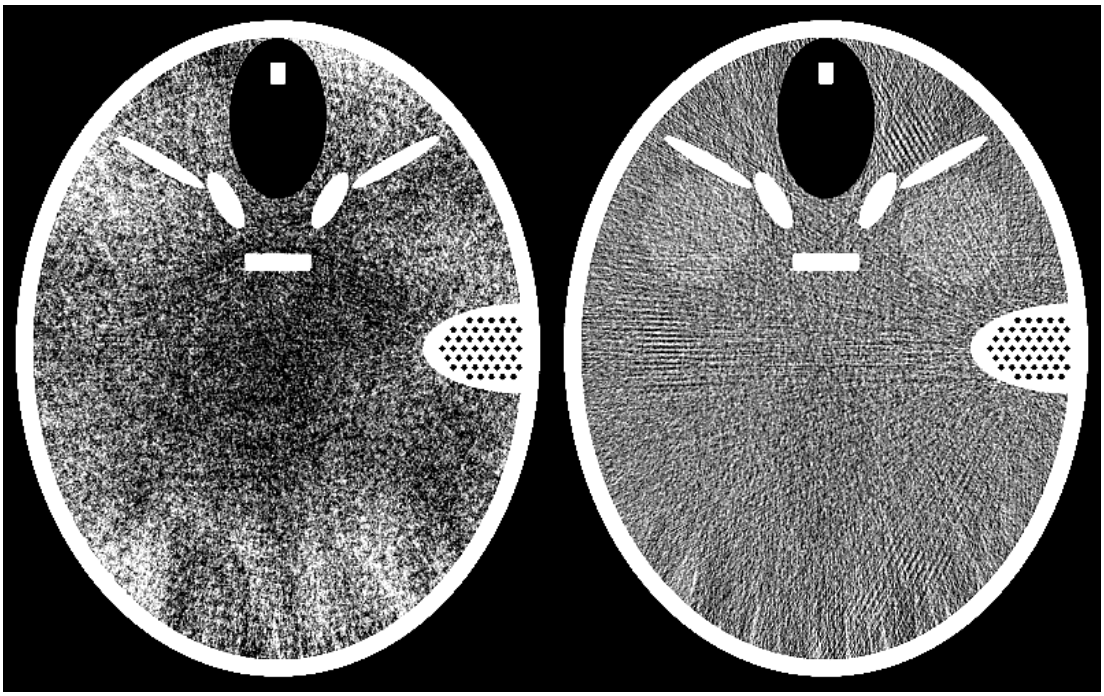
[7] 2024, private conversations with Andrew Kingston and Adrian Sheppard.

[8] J. Gregor and T. Benson, "Computational analysis and improvement of sirt," *IEEE transactions on medical imaging*, vol. 27, no. 7, pp. 918–924, 2008.

[9] A. H. Andersen and A. C. Kak, "Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm," *Ultrasonic imaging*, vol. 6, no. 1, pp. 81–94, 1984.



(a) slices from top to bottom: $y = 0, z = 0, x = 0$.



(b) slice $z = 8\,\mathrm{mm}$.

Fig. 7: **Left:** slices from the 10th iteration of the OS-SIRT method. **Right:** slices from the 4th iteration of the FBP method. It has converged. **Both:** plotting range: [0, 100] HU.

Practical Dental CBCT Reconstruction under Truncated Projections

Kiwan Jeon and Hyoung Suk Park

***Abstract*—Cone-beam computed tomography (CBCT) is widely used in dental diagnosis and treatment planning due to its cost-effective and compact system design compared with multi-detector CT (MDCT). However, the limited detector size often leads to field-of-view (FOV) truncation and incomplete projection data, which can significantly degrade reconstruction quality and introduce artifacts. Although the FDK algorithm remains a standard choice for CBCT reconstruction because of its high computational efficiency, truncation-related degradations are not fully resolved by conventional approaches, and simple extrapolation methods may still suffer from contrast mismatch. In this study, we propose a practical reconstruction framework that combines the fast reconstruction capability of FDK with the representational capacity of implicit neural representations.**

***Index Terms*—Dental Cone-Beam Computed Tomography, Truncated Projections, Implicit Neural Representation, Projection Extension.**

I. Introduction

Dental cone-beam computed tomography (CBCT) has become an essential imaging modality in modern dentistry because it provides volumetric anatomical information that directly supports treatment planning and decision-making [1]–[4]. Compared with conventional 2D radiography, CBCT reduces the ambiguity caused by structural superposition and offers a clear representation of complex maxillofacial anatomy in three dimensions. This capability is particularly valuable in a wide range of clinical applications, including implant planning, evaluation of impacted teeth, orthodontic assessment, and the detection of periapical lesions or bone defects. Compared with MDCT, dental CBCT is a cost-effective imaging modality owing to its use of relatively small flat-panel detectors, which enables a compact system design and reduces installation space requirements. However, the limited detector size can restrict the imaging coverage and cause field of view (FOV) truncation, leading to incomplete projection data and subsequent degradation of reconstructed image quality, often accompanied by amplified artifacts.

FDK (Feldkamp–Davis–Kress) algorithm is one of the most widely used in CBCT and is often regarded as a de facto standard in the class of analytic reconstruction methods [5]. By approximately extending the concept of filtered backprojection (FBP) to cone-beam geometry, it enables computationally efficient reconstruction with a relatively simple processing pipeline. Moreover, since the filtering and backprojection steps are highly amenable to parallelization, the adoption of parallel computing accelerators such as GPUs allows high-resolution three-dimensional volumetric images to be reconstructed within a few seconds.

In particular, the FDK algorithm applies reconstruction filter kernels using Fourier transform–based filtering, typically implemented via the FFT. However, when projection data are missing due to the use of small detectors, the projection signal becomes inherently non-periodic and exhibits boundary discontinuities, which can compromise the accuracy of Fourier-based filtering that implicitly assumes periodicity. Moreover, abrupt discontinuities at the ends of the projection data can induce excessive high-frequency components, and when combined with singular filter kernels such as ramp-type filters, they may lead to severe reconstruction errors, including pronounced streak artifacts.

Constant extension can be effectively used to mitigate artifacts caused by projection truncation. This approach introduces little to no additional computational cost while preventing boundary-induced high-frequency artifacts from propagating into the imaging region. However, such a simple extension may still fail to preserve accurate image contrast, as inaccurate projection extrapolation can introduce distortions in the reconstructed volume. Beyond constant extension, numerous projection completion and extrapolation methods have been developed to compensate for truncated projections, providing more reliable estimates of missing data and alleviating truncation-related artifacts [6], [7].

Recently, implicit neural representation (INR) based methods have emerged as a promising solution for mitigating reconstruction degradations caused by incomplete projection data, and extensive research has been conducted in this direction [8]–[11]. However, INR-based approaches often involve substantial training time for 3D CBCT reconstruction, which continues to be a major constraint for practical and time-sensitive applications.

In this study, we aim to develop a practical 3D CBCT reconstruction method by leveraging the rapid computation of FDK and the representational capacity of INRs. We use an INR to reconstruct a prior image with a coarse voxel size over an extended imaging domain. The forward projection of this prior image fully covers the patient's head, enabling the estimation of missing measurements outside the truncated projection. The resulting extended projections are then used to reconstruct a high-resolution 3D volume using conventional FDK.

II. Methods

Consider a dental CBCT system equipped with a small flat-panel detector, which leads to a truncated field of view. Let Ω

K. Jeon and H. S. Park are with National Institute for Mathematical Sciences, Daejeon, 34047, Republic of Korea

Corresponding: Kiwan Jeon (jeonkiwan@nims.re.kr)

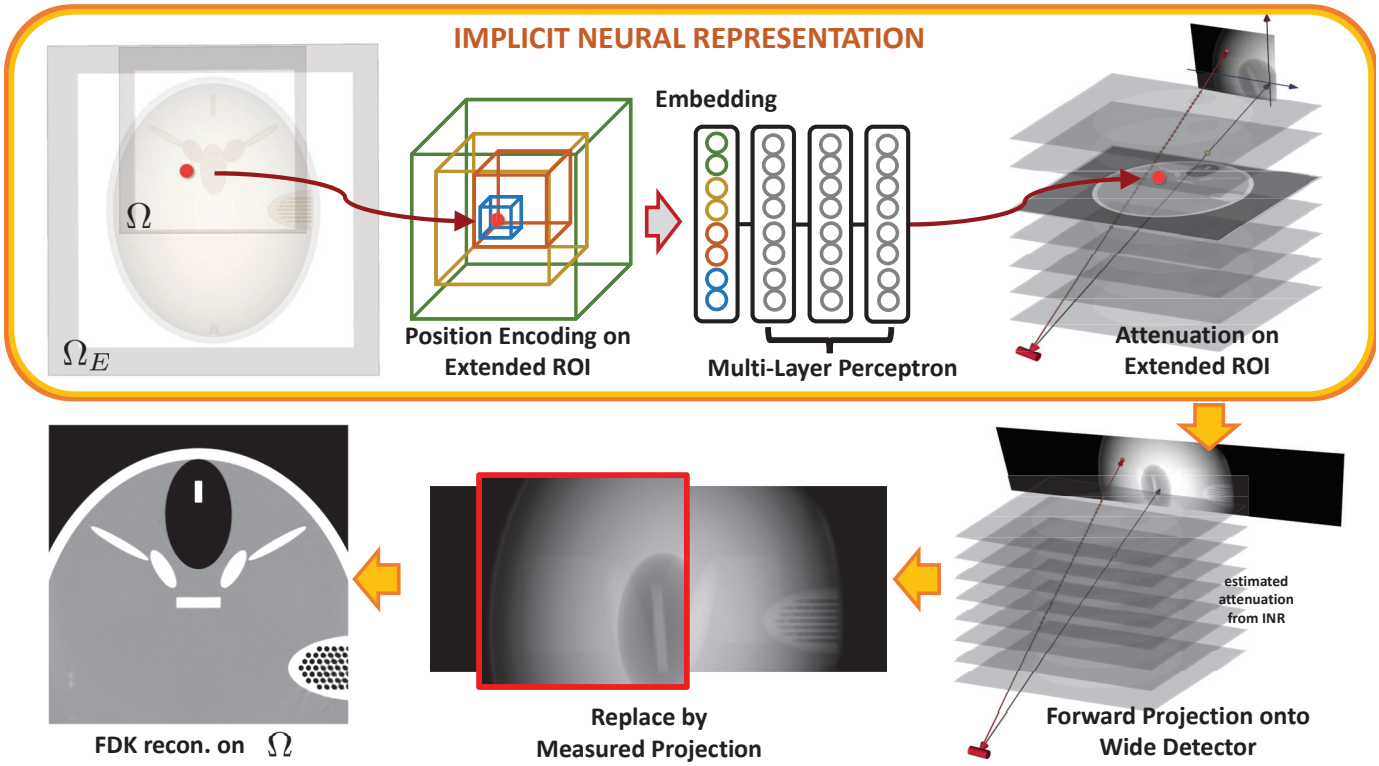


Fig. 1. Overview of the proposed 3-D dental CBCT reconstruction framework that leverages an implicit neural representation (INR) prior under truncated projections.

be the imaging domain arising from this configuration, and let Ω_E denote an extended domain that fully covers the patient's head. We begin by employing an INR to reconstruct a coarse-resolution prior volume over Ω_E. Given a 3D coordinate $\mathbf{x} \in \Omega_E$, we encode its spatial location using a hash-grid positional encoding [8], [12], denoted by $\mathrm{ENC}(\mathbf{x})$. The encoded features are then fed into a multi-layer perceptron $\mathrm{MLP}_\theta(\cdot)$, which predicts the corresponding attenuation value at $\mathbf{x}$. Following the Lambert–Beer law, the network parameters θ are learned by minimizing the coarse-stage objective $\mathcal{L}_c$, defined as:

$$\mathcal{L}_c = \sum_{l_{\varphi,u,v}} \left| \mathrm{P}(\varphi,u,v) - \sum_{\mathbf{x}_k \in l_{\varphi,u,v}} \mathrm{MLP}_\theta(\mathrm{ENC}(\mathbf{x}_k))\, d_{\mathbf{x}_k} \right|, \tag{1}$$

where, $\mathrm{P}(\varphi,u,v)$ denotes the measured projection data at detector position (u,v) and projection angle φ, $l_{\varphi,u,v}$ denotes the cone-beam ray connecting the X-ray source and the detector position (u,v) at angle φ, $\mathbf{x}_k$ denotes sampled positions along the ray $l_{\varphi,u,v}$, and $d_{\mathbf{x}_k} = \|\mathbf{x}_{k+1} - \mathbf{x}_k\|$ denotes the distance between adjacent sample positions. A graphical overview of the INR implementation is shown in the top row of Fig. 1. The trained $\mathrm{MLP}_\theta(\mathrm{ENC}(\mathbf{x}))$ is used to synthesize a coarse-resolution prior volume $\mathbf{u}_0$ by querying it at coarse-grid positions $\mathbf{x} \in \Omega_E$. Applying the forward projection to $\mathbf{u}_0$ over a detector wide enough to fully cover the imaged subject, we obtain an non-truncated projection P_0. The extended projection $\hat{\mathrm{P}}$ for applying FDK algorithm is defined as:

$$\hat{\mathrm{P}} = \mathrm{M} \odot \mathrm{P} + (1 - \mathrm{M}) \odot \mathrm{P}_0, \tag{2}$$

where M is a binary mask that indicates the region of the measured projection.

A. Parameters

The CBCT acquisition geometry for generating truncated projection data is provided in Table I. The imaging domain settings for the coarse grid used in the INR stage and the fine grid used in FDK reconstruction are provided in Table II. For the positional encoder (ENC), we used parameters similar to those in [12]: a minimum resolution $\mathrm{N}_{min} = 16$, a maximum resolution $\mathrm{N}_{max} = 420$, $L = 9$, feature vectors $T = 2^{17}$, and feature dimension $F = 2$. The fully connected network (MLP) is given by 256 nodes with 3 hidden layers. For INR optimization, we used a learning rate of 0.001 and sampled 256 rays per batch, with 448 points randomly sampled along each ray. To ensure practical efficiency, the INR prior was trained for less than 30 minutes, corresponding to 60,000 iterations in our implementation.

III. Results

To provide both qualitative and quantitative validation, we used the FORBILD head phantom [13]. The INR model was implemented in PyTorch [14], while the forward/backprojection operators and the FDK reconstruction were implemented using the ASTRA Toolbox [15]. All experiments were performed on a workstation with an Intel Core i9-7900X processor, DDR4-2400 128GB RAM, and an NVIDIA GeForce RTX 3090 Ti GPU (24 GB VRAM). To evaluate the performance of the proposed algorithm, we compared

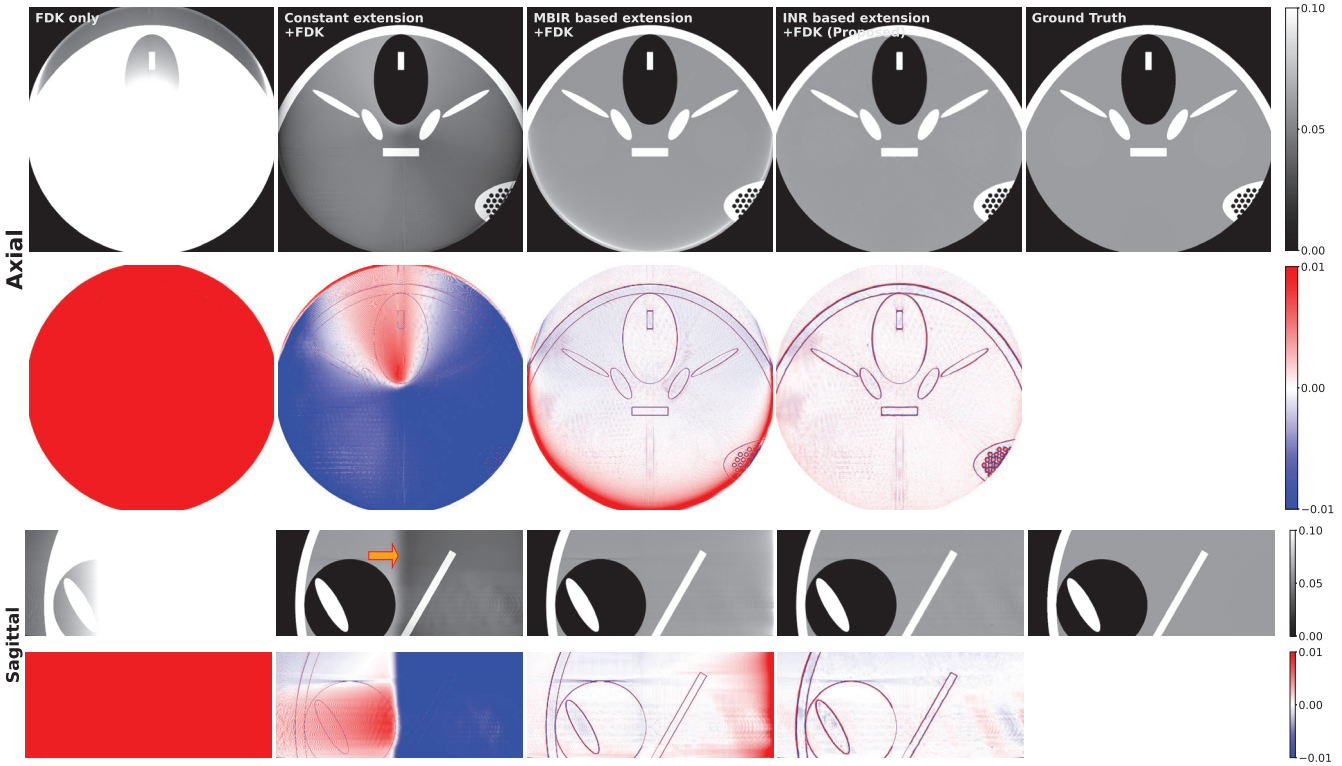


Fig. 2. Comparison of different reconstruction results for the FORBILD head phantom in 3D dental CBCT imaging. From left to right, the columns show FDK reconstructions using: (1) no preprocessing, (2) constant projection extension, (3) coarse-grid MBIR–based projection extension, and (4) the proposed coarse-grid INR-prior–based projection extension. The first two rows correspond to axial views, and the last two rows correspond to sagittal views. In both axial and sagittal views, the first row presents the reconstructed images, while the second row presents the corresponding difference (error) maps.

TABLE I
CT GEOMETRY PARAMETERS FOR TRUNCATED PROJECTION DATA GENERATION. THE LENGTH UNIT IS mm.

Property	Value
Distance (Source to Detector)	600.0
Distance (Source and Iso Center)	400.0
Number of Detector Pixels	640×640
Detector Pixel Resolution	0.2×0.2
Detector Offset	(57.0, 29.0)
Number of Projections	1200
Range of Projection Angle	$0-2\pi$
Number of Extended Detector Pixels	3072×640

TABLE II
IMAGING DOMAIN (BOUNDING BOX) DESCRIPTION FOR THE COARSE (INR) AND FINE (FDK) GRID. THE UNIT IS mm.

	Property	Values
Coarse Grid (INR)	Min. Position	$[-140.0, -80.0, -40.0]$
	Max. Position	$[140.0, 200.0, 105.0]$
	Voxel Resolution	$[0.5, 0.5, 0.5]$
Fine Grid (FDK)	Min. Position	$[-80.0, -80.0, -32.0]$
	Max. Position	$[80.0, 80.0, 88.0]$
	Voxel Resolution	$[0.2, 0.2, 0.2]$

it with several baseline reconstruction approaches using the dental CBCT geometry specified in Table I. Fig. 2 compares the reconstruction results obtained with different approaches. When FDK is directly applied to truncated projections without any preprocessing, the reconstructed image quality is severely degraded, making it difficult to identify meaningful anatomical structures under the given window level. In contrast, applying constant extension enables a visually interpretable reconstruction by mitigating truncation-related artifacts, allowing anatomical structures to be reasonably observed. However, because constant extension relies on a simplistic extrapolation, the reconstruction still suffers from contrast mismatch, which remains a notable limitation. When model-based iterative reconstruction (MBIR) is used as a baseline method, it provides improved anatomical fidelity and better contrast consistency in the reconstructed images. However, boundary-related artifacts introduced during the iterative reconstruction process may persist, resulting in noticeable distortions in the final reconstruction. In contrast, the proposed method with an INR-based prior substantially mitigates truncation artifacts and contrast mismatch, yielding visually consistent reconstructions with improved anatomical fidelity.

IV. DISCUSSION

INR-based reconstruction has been recognized as an effective approach for mitigating image degradation caused by truncated projections. However, obtaining a 3D reconstruction with reasonable quality typically requires substantial computational time, and prior studies have reported that image quality improves progressively as the optimization time increases [11]. In this study, to alleviate such heavy computational requirements, we propose to use INR only for generating a prior image and to limit its optimization to a relatively short duration (less than 30 minutes) on a coarse grid. As

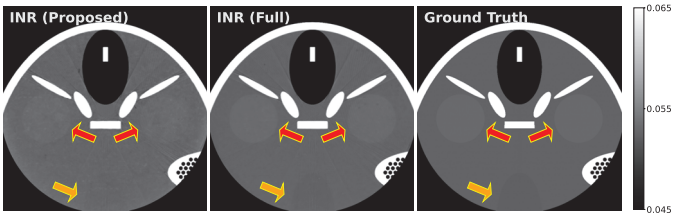


Fig. 3. Comparison of reconstructed images obtained with different INR prior optimization times.

demonstrated in the results, even with this constrained training time, the resulting INR prior is sufficient to effectively reduce the reconstruction errors and truncation-induced degradations observed in conventional FDK reconstruction. In this way, the proposed framework combines the efficiency of FDK with the global modeling capability of INR, enabling a practical trade-off between reconstruction quality and computational cost. However, our experiments also indicate that reconstructing low-contrast subjects still requires considerable computational time to achieve high-quality results. Fig. 3 compares the final reconstructions obtained using the proposed short INR-prior optimization (less than 30 minutes, 60,000 iterations) and a full-time optimization setting (at least 14 hours, 1,920,000 iterations). While the proposed approach remains capable of recovering low-contrast structures, the reconstruction exhibits reduced sharpness and increased image noise compared with the full-time result. These findings suggest that further improving the computational efficiency of INR-based priors is crucial for practical dental CBCT reconstruction.

V. Conclusion

In this study, we proposed a practical reconstruction framework for dental CBCT under truncated acquisitions by combining an INR-based coarse prior with a conventional FDK algorithm. The synthesized extended projections generated from the INR prior effectively reduce truncation-induced artifacts and improve contrast consistency compared with standard FDK and simple projection extension baselines. While the proposed method achieves meaningful improvements with a limited INR training time, the results also indicate that longer optimization remains beneficial for improving the visibility of low-contrast regions. Future work will focus on developing more computationally efficient INR strategies to further enhance reconstruction quality under short time constraints. In addition, validation on real clinical CBCT datasets and evaluation under realistic noise conditions will be important to assess the robustness and practical applicability of the proposed framework.

Acknowledgment

The authors were supported by the National Institute for Mathematical Sciences (NIMS) grant funded by the Korean government (No. NIMS-B26A10000) and partially supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2024-00338419).

References

[1] W. C. Scarfe and A. G. Farman, "What is cone-beam ct and how does it work?" *Dental Clinics of North America*, vol. 52, no. 4, pp. 707–730, 2008, contemporary Dental and Maxillofacial Imaging.

[2] J. B. Ludlow and M. Ivanovic, "Comparative dosimetry of dental cbct devices and 64-slice ct for oral and maxillofacial radiology," *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, vol. 106, no. 1, pp. 106–114, 2008.

[3] T. Kaasalainen, M. Ekholm, T. Siiskonen, and M. Kortesniemi, "Dental cone beam ct: An updated review," *Physica Medica*, vol. 88, pp. 193–217, 2021.

[4] C. M. Hyun, K. Jeon, S. H. Kang, S.-H. Lee, J. K. Seo, and H. S. Park, *Clinical Applications of Low-Dose Dental Cone-Beam Computed Tomography*. Springer Nature Switzerland, 2024, pp. 147–163.

[5] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, Jun 1984.

[6] B. Ohnesorge, T. Flohr, K. Schwarz, J. P. Heiken, and K. T. Bae, "Efficient correction for ct image artifacts caused by objects extending outside the scan field of view," *Medical Physics*, vol. 27, no. 1, pp. 39–46, 2000.

[7] J. Hsieh, E. Chao, J. Thibault, B. Grekowicz, A. Horst, S. McOlash, and T. J. Myers, "A novel reconstruction algorithm to extend the ct scan field-of-view," *Medical Physics*, vol. 31, no. 9, pp. 2385–2391, 2004.

[8] R. Zha, Y. Zhang, and H. Li, "NAF: Neural attenuation fields for sparse-view cbct reconstruction," in *Medical Image Computing and Computer Assisted Intervention (MICCAI 2022)*. Springer Nature Switzerland, 2022, pp. 442–452.

[9] H. S. Park, J. K. Seo, and K. Jeon, "Implicit neural representation-based method for metal-induced beam hardening artifact reduction in x-ray ct imaging," *Medical Physics*, 2025.

[10] H. S. Park and K. Jeon, "An iterative reconstruction method for dental cone-beam computed tomography with a truncated field of view," in *Proceedings of the 18th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, Shanghai, China*, 2025.

[11] H. S. Park and K. Jeon, "Adaptive multi-resolution hash-encoding framework for inr-based dental cbct reconstruction with truncated fov," *Medical Physics*, 2026.

[12] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Trans. Graph.*, vol. 41, no. 4, 2022.

[13] "Processing of forbild phantom files for rtk," https://github.com/RTKConsortium/Forbild, gitHub Repository.

[14] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017.

[15] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the astra toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, Oct 2016.

Two-Stage Cascaded Caries Detection on Panoramic X-Ray via Slicing-Aided Hyper Inference

Subin Park, Yoonha Eo, and Jongduk Baek*

**** ***Abstract*—While deep learning methods for dental caries detection have shown promising results, detecting small lesions remains challenging due to CNN-induced spatial information loss, particularly when lesions occupy only a small portion of the panoramic mage. This often suppresses diagnostic features of caries, which are characterized by subtle contrast gradients, leading to a high false-positive rate and clinical fatigue. To address this issue, we propose a two-stage cascaded framework that leverages Slicing Aided Hyper Inference (SAHI) to maintain original resolution while mitigating detail loss inherent in CNN-based methods. In the first stage, SAHI partitions high-resolution radiographs into overlapping slices, which are processed by the detection module (YOLOv11n) to produce potential candidates. In the second stage, the classification module (YOLOv11n-cls) performs localized analysis on candidate-centered patches to classify whether each candidate corresponds to caries. Our results demonstrate that a multi-stage, resolution-preserving approach is critical for detecting small lesion areas that are otherwise lost in single-stage detection pipelines.**

***Index Terms*—Slicing Aided Hyper Inference, Caries, Detection, Panoramic Radiograph.**

I. Introduction

DENTAL caries is one of the most common oral diseases in dentistry, necessitating early and accurate detection to prevent irreversible complications such as pulp infection and peri-apical abscesses [1]. A primary challenge in caries detection task on panoramic radiographs is the strong dependence on image resolution. Unlike general object detection tasks, where shape and color-relevant features are dominant, dental caries are primarily identified through subtle contrast variations. While standard convolutional neural networks (CNNs), such as YOLO [2], resize images to reduce computational cost, it downsamples and compresses small lesions into only a few pixels [3]. This is further complicated by small object dominance and spatial imbalance in dental radiographs [4]; For example, a single carious lesion in our dataset occupies, on average, 0.2% of the total panoramic area, while the majority are non-pathological structures.

Subin Park and Yoonha Eo are with the department of Artificial Intelligence, Yonsei University, Seoul, South Korea (e-mail: subinp0507@gmail.com, yoonhaeo@yonsei.ac.kr)

*Corresponding author: Jongduk Baek (e-mail: jongdukbaek@yonsei.ac.kr)

To address these limitations, we propose a two-stage cascaded inference framework that leverages an open-source framework called Slicing Aided Hyper Inference (SAHI) methodology [5], which partitions a high-resolution panorama (that typically averages $2,440 \times 1,280$ pixels) into smaller, overlapping slices of 640×640 pixels and applies an object detection forward pass independently to each slice. By bypassing downsampling of the entire image to a fixed input size, the YOLOv11n-based detection module processes patches while preserving the original resolution, capturing nearly all potential lesions during initial screening.

In the second phase, the proposed detections are passed onto the classification module for verification. A high-resolution 320×320 pixels region is extracted around each detection by centering the bounding box on the crop. This enables the YOLOv11n-cls backbone-based classifier to perform fine-grained morphological feature analysis, effectively distinguishing carious lesions from non-carious and non-pathological artifacts to boost precision. By decoupling the high-recall spatial localization from the high-precision verification phase, this framework achieves improved precision and F1-score compared to single-stage detectors, providing a more reliable tool for clinical diagnosis.

II. Methods

The training strategy and the inference framework are illustrated in Figure 1(a) and (b), respectively. For the inference, we utilize the YOLOv11n architecture [2] for the initial detection stage and the YOLOv11n-cls module [2] for the second stage of classification-based refinement.

A. Detection with SAHI

To handle spatial information loss inherent in small object detection in large-scale panoramic radiographs, we employ SAHI approach in the inference stage [5]. Specifically, the panoramic image I with dimensions $W \times H$ is segmented into l overlapping patches, defined as:

$$\mathcal{P} = \{P_1^I, P_2^I, \ldots, P_l^I\} \tag{1}$$

Panoramic radiographs from the AAAI dataset [6] we used vary in size, and the majority were sized at 2,440 × 1,280

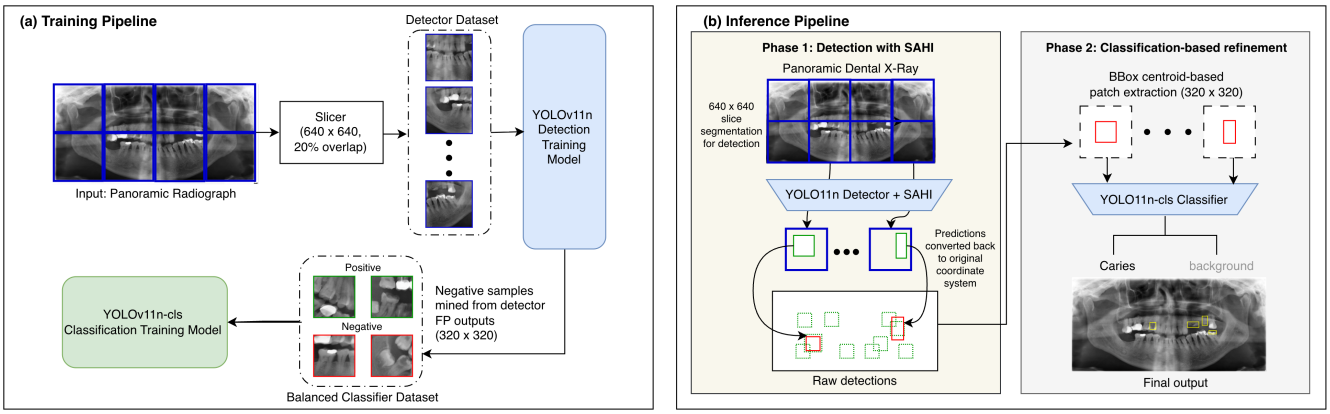


Figure 1. Diagram of proposed (a) training pipeline and (b) two-stage cascaded inference pipeline.

pixels. Consequently, we defined $M = N = 640$ pixels, as these dimensions sufficiently enclose the local context, such as neighboring teeth and interproximal spaces. During inference, patches were extracted with a stride S, determined by an overlap ratio σ:

$$S = M \times (1 - \sigma) \tag{2}$$

Given that the average dimensions of the lesion in our dataset are 65×79 pixels, an overlap ratio of $\sigma = 0.2$ (which yields 128 pixels overlap zone) ensures lesions are situated near slice boundaries. This redundancy prevents truncated detections since all potential lesions are captured within at least one adjacent patch.

These overlapping predictions merge and translate back to original panoramic system coordinates after Non-Maximum Suppression (NMS) [7], a post-processing algorithm that merges all detections that belong to the same object. To maximize sensitivity, the detection module utilizes a permissive confidence threshold ($\tau_{det} = 0.05$). This low threshold strategically prioritized high recall to capture all potential carious candidates, which includes a significant number of false-positives from non-pathological artifacts. Since dental caries are minute in size, an optimal NMS matching threshold of 0.5 was selected for precise bounding box localization. Since minor pixel shifts can result in a bounding box that encloses irrelevant structure, this threshold effectively collapsed redundant candidate boxes generated by the low-threshold detector.

B. *Classification-based Refinement*

To address the high false-positive rate inherent in the detection stage while preserving high recall, a second-stage classifier module acts as a precision-enhancing filter, suppressing non-carious artifacts. We extracted a 320×320 patch which centered on each detection proposal, and each patch was processed by YOLOv11n-cls classification architecture.

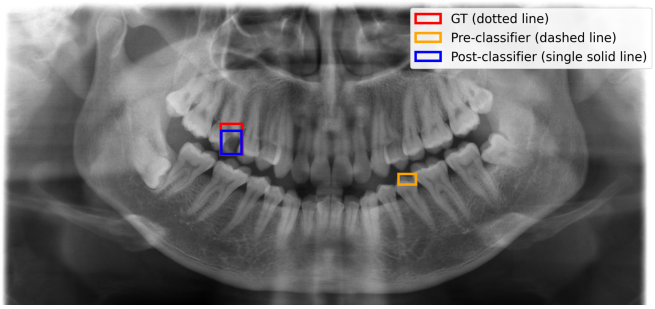


Figure 2. Visual representation of pre-classifier and post-classifier detection outputs against ground-truth from the inference pipeline.

This stage functions as a high-resolution "zoom-in" verification step to validate the presence of caries. The 320×320 resolution allows for caries mask, such as shape and texture, to be observed in detail compared to the global panoramic view.

Our cascaded inference framework requires each final candidate C_i to be characterized by b_i, which denotes the bounding box coordinates $[x, y, w, h]$ predicted by stage 1, and confidence scores from detection (s_{det}) and classification (s_{cls}) to satisfy the following dual-threshold criteria:

$$C_i = (b_i, s_{det}, s_{cls}) \mid (s_{det} \geq \tau_{det}) \wedge (s_{cls} \geq \tau_{cls}) \tag{3}$$

where the thresholds $\tau_{det} = 0.05$ and $\tau_{cls} = 0.4$ were determined via a grid-search on the validation set to optimize the F1-score. This allows the framework to consider low-confidence detections from the stage 1 detector, provided they are confirmed with high confidence by the stage 2 classifier.

Figure 2 qualitatively illustrates the refinement process. Candidate regions proposed by the first-stage detector are evaluated by the classifier, which effectively filters out non-carious structures (orange boxes) while preserving true carious lesions (red boxes) that align with the ground truth annotations (solid boxes).

For the final evaluation and the construction of Precision-Recall (PR) curves, candidates were ranked based on a late-fusion score, which is defined as $s_{final} = s_{det} \times s_{cls}$. While the classifier operates as a gatekeeper to enhance precision, this combined scoring strategy was employed to reflect spatial localization capability of the detection module and textural verification performance of the classification module in the final ranking.

III. Experiments

A. Dataset and Pre-processing

We experimented on a publicly available AAAI dataset [6]. This dataset comprises 6,000 panoramic dental X-ray images, with a total of 13,783 dental caries instances. The whole dataset was randomly divided into 4,320/480/1,200 for training, validation, and testing, respectively. The dataset consists of panoramic radiographs with a consistent height of 1,280 pixels, with widths ranging from 1,932 to 2,648 pixels, with approximately 65% of the images sized at 2,440 $\times$ 1,280 pixels.

For the stage 1 detection model, we generated a balanced training set of 640×640 pixel-sized patches. For robust spatial localization, ground-truth lesions were placed at random coordinates within each slice to simulate a spatially invariant distribution encountered during SAHI inference phase. To further challenge the detection, we mined *hard negatives* (HN) by extracting lesion-free patches from high-density dental structures while avoiding peripheral gum regions. By exposing these difficult anatomical features, we forced the model to learn discriminative features to distinguish between carious and healthy dental structures.

Following the bootstrapping principles established by Felzenszwalb *et al*. [8], we curated the classifier's negative dataset using first-stage false-positive detections exclusively from training and validation subsets. A 1:1 ratio of negative-to-positive samples was maintained to ensure a balanced optimization of the binary classification task. Each proposed candidate was processed into a 320 $\times$ 320 patch which centered on predicted bounding-box. Although the input resolution was smaller than that used for detection, the design allowed the YOLOv11n-cls model to focus on the fine-grained morphological details and textural contrast gradients.

B. Training Implementation

The proposed training loops were implemented using the PyTorch library and trained on an NVIDIA RTX A6000 GPU. In stage 1, the YOLOv11n detection module [2] was employed for training with images with an input size of 640 $\times$ 640 pixels. It was optimized using a Stochastic Gradient Descent (SGD) optimizer with an initial learning rate of 1×10^{-3}, a momentum of 0.937, and a weight decay of 0.0005. Standard loss function consisted of Complete IoU (CIoU) for bounding box regression and Binary Cross-Entropy (BCE) for classification [2]. To enhance the model's robustness in varying imaging conditions, we employed an automated RandAugment strategy alongside random horizontal flipping and translations [9].

In stage 2, 320 $\times$ 320 pixels patches were utilized for training YOLOv11n-cls based classifier. We employed an initial learning rate of 5×10^{-4} and a dropout rate of 0.3 to prevent overfitting on the textural features of the dental crops. To enhance the model's generalization across subtle gradients, we implemented Mixup augmentation with a coefficient of 0.2 [10]. Both stages were trained for 500 epochs with a batch size of 32, and an early stopping patience of 50 epochs to encourage optimal convergence.

C. Evaluation

Model performance was primarily assessed based on diagnostic reliability, with emphasis on precision to minimize false-positive predictions and F1-score to evaluate the balance between precision and recall. We also reported mAP at an IoU threshold of 0.5 (mAP_{50}) for benchmarking consistency. This approach allowed us to quantify precision and F1-score gains afforded by the classification-based refinement stage while maintaining high sensitivity in the initial detection stage.

We also compared our proposed framework with the following state-of-the-art (SOTA) object detection architectures, focusing on their performance in managing precision-recall trade off: Faster R-CNN [11], a two-stage detector combining region proposals with bounding box classification; YOLOv8 [12], an anchor-free detector that integrates cross stage partial (CSP) and feature pyramid network (FPN) to aggregate multi-scale features [12]; YOLOv11n [2], a lightweight detector built on its predecessor with refined feature extraction capabilities; and RT-DETR [13], a transformer-based architecture employing a hybrid encoder.

IV. Results

We compared our framework against several SOTA architectures, including Faster R-CNN, YOLOv8, YOLOv11n, and RT-DETR. As summarized in Table I, most baseline and SOTA models achieved relatively high recall but suffered from low precision and F1-score. In particular, RT-DETR exhibited the lowest precision of 47.8%, indicating a high false-positive rate.

The baseline results further demonstrate that the slicing-aided inference strategy is superior to the conventional full-image resizing approach. Compared to the YOLOv11n baseline, incorporating SAHI improved precision, F1-score, recall, and mAP_{50}, confirming its effectiveness in detecting small carious regions.

Our full framework (SAHI + Classifier) achieved the highest precision of 69.1% and the highest F1-score of

TABLE I
QUANTITATIVE PERFORMANCE AND COMPARATIVE ANALYSIS ON THE AAAI BENCHMARK. BOLD VALUES INDICATE THE BEST PERFORMANCE IN EACH COLUMN.

Method	Precision	Recall	F1	mAP_{50}
Baselines				
Full Panorama (YOLOv11n)	56.6	80.9	66.5	72.1
Full Image w/ SAHI	58.4	**88.8**	70.5	**73.8**
SOTA Comparisons				
Faster R-CNN (ResNet-50)	55.4	84.9	67.1	73.2
YOLOv8	59.6	81.4	68.8	73.5
RT-DETR	47.8	75.1	58.4	64.5
Proposed Framework				
Ours (SAHI + Classifier)	**69.1**	85.1	**76.3**	72.4

TABLE II
INFERENCE EFFICIENCY COMPARISON ON PANORAMIC RADIOGRAPHS. BOLD VALUES INDICATE OUR PROPOSED FULL PIPELINE.

Method	Total Latency (s)
Faster R-CNN	0.0920
YOLOv8n	0.0035
YOLOv11n	0.0163
RT-DETR	0.0985
Ours (Stage 1: SAHI)	0.3293
Ours (Full Pipeline)	**0.3394**

Note: Baseline SOTA models were evaluated on full-resolution panoramic radiographs using single-pass inference. Our framework's latency includes the patch generation ($\sigma = 0.2$), multi-slice detection, and classification.

76.3%. Although a marginal decrease in recall (from 88.8% to 85.1%) and mAP_{50} (from 73.8% to 72.4%) were observed after integrating the classifier, this trade-off reflects a deliberate shift toward higher diagnostic precision.

The inference efficiency of our framework and SOTA methods is reported in Table II. The proposed full pipeline achieved a total latency of 0.3394 seconds per panoramic radiograph. Although this latency is higher than that of single-stage detectors due to multi-slice processing, it remains within a practical range for clinical deployment.

V. DISCUSSION

Our two-stage cascaded framework effectively improved precision and F1-score, demonstrating clear improvement in diagnostic reliability for dental caries detection. By integrating slicing-aided detection with a classification-based refinement stage, the framework preserves high-resolution spatial information that is typically lost in single-pass detectors, while addressing the scale imbalance inherent in panoramic radiographs.

The primary performance improvement is attributed to the adoption of Slicing Aided Hyper Inference (SAHI), which partitions high-resolution panoramic images into overlapping patches. Since this increases the relative pixel coverage of small caries, it allows the detector to extract features at original resolution without spatial detail loss. Consequently, the detection stage achieves high sensitivity by capturing nearly all potential carious lesion candidates.

The second-stage classifier functions as a selective refinement mechanism that suppresses false-positive detections generated during the high-recall detection stage. Although this refinement stage introduces a slight reduction in recall and mAP_{50}, it presents a 10.7% improvement in precision compared to the detection-only baseline. This behavior reflects precision–recall trade-off, as the classifier, which operates on localized 320×320 pixel patches, may discard some true positives due to limited global anatomical context.

Our proposed framework achieves the highest F1-score of 76.3%, indicating an optimal balance for clinical application. Unlike single-stage SOTA detectors, which preserve high recall at the cost of high volume of false-positives, the cascaded design improves precision with minimal impact on sensitivity.

Overall, the results suggest that decoupling high-sensitivity localization from high-specificity verification is an effective strategy for small-lesion detection in large-scale medical images, enabling improved diagnostic confidence while maintaining practical inference efficiency.

VI. CONCLUSION

This study explored the integration of SAHI methodology with a hierarchical feature verification module to mitigate high-resolution dependency in caries detection in high-resolution dental radiographs. Compared to SOTA approaches, our results achieved superior performance in maximizing precision and F1-score by successfully distinguishing confounding dental structures. Future work will explore the generalization ability of this cascaded framework in other modalities, such as mammography and cone-beam computed tomography (CT), where minute pathological details of lesions are prone to scale imbalance and downsampling artifacts.

ACKNOWLEDGMENT

This work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (RS-2023-00219019, RS-2023-00240135, and RS-2025-00553670), and Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by MSIT (No. RS-2020-II201361, Artificial Intelligence Graduate School Program, Yonsei University).

REFERENCES

[1] V. V. Gordan, J. L. Riley III, R. M. d. Carvalho, J. Snyder, J. Sanderson Jr, M. Anderson, G. H. Gilbert, and D. C. Group, "Methods used by Dental Practice-Based Research Network (DPBRN) dentists to diagnose dental caries," *Operative dentistry*, vol. 36, no. 1, pp. 2–11, 2011.

[2] G. Jocher, J. Qiu, and A. Chaurasia. (2023, Jan.) Ultralytics YOLO. Version 8.0.0. [Online]. Available: https://github.com/ultralytics/ultralytics

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2016.

[4] M. Nikouei, B. Baroutian, S. Nabavi, F. Taraghi, A. Aghaei, A. Sajedi, and M. E. Moghaddam, "Small object detection: A comprehensive survey on challenges, techniques and real-world applications," *Intelligent Systems with Applications*, vol. 27, p. 200561, Sep. 2025.

[5] F. C. Akyon, S. Onur Altinuc, and A. Temizel, "Slicing aided hyper inference and fine-tuning for small object detection," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 966–970.

[6] B. Chen, S. Fu, Y. Liu, J. Pan, G. Lu, and Z. Zhang, "CariesXrays: Enhancing caries detection in hospital-scale panoramic dental x-rays via feature pyramid contrastive learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 20, 2024, pp. 21 940–21 948.

[7] J. Hosang, R. Benenson, and B. Schiele, "Learning non-maximum suppression," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4507–4515.

[8] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[9] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "RandAugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Jun. 2020, pp. 702–703.

[10] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," in *International Conference on Learning Representations (ICLR)*, 2018.

[11] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 28, 2015.

[12] M. Sohan, T. Sai Ram, and C. V. Rami Reddy, "A review on YOLOV8 and its advancements," in *International Conference on Data Intelligence and Cognitive Informatics*. Springer, 2024, pp. 529–545.

[13] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, 2024, pp. 16 965–16 974.

X-ray CT based Anomaly Detection by Pseudo-Volumetric Reconstruction with Channel-Disentangled Memory

Yunkun Bai, Xiang Zou, Yuxiang Xing

***Abstract*—X-ray computed tomography (CT) provides essential volumetric information for non-destructive testing. For internal defect detection in Ball Grid Array (BGA) solder joints, X-ray CT has become a critical tool for higher accuracy than conventional 2D X-ray detection. However, existing anomaly detection methods struggle to balance accuracy and efficiency. Slice-by-slice approaches miss z-axis spatial context, whereas 3D voxel-based methods are computationally prohibitive for real-time manufacturing. Moreover, the diversity and scarcity of defective samples in mature manufacturing necessitates unsupervised approaches capable of identifying various anomalies, ranging from micrometer-scale voids to macroscopic structural deformations. To address these challenges, We propose Pseudo-Volumetric Channel-Disentangled Memory Network (PV-CDMNet), a unified framework that synergistically combines hierarchically structured memory bank with reconstruction-based detection. The core contributions include: (1) Pseudo-Volumetric Data Construction strategy utilizing Gaussian Interval Sampling (GIS) to capture 2.5D spatial context within a lightweight 2D architecture, naturally adapt to different CT resolution parameters; (2) Hierarchical Channel-Disentangled Memory (H-CDM) modules that divide feature channels into independent semantic subspaces, preventing the suppression of subtle defect features by dominant background features; (3) 3D spatial gradient loss that enforces consistency between neighboring slices, enabling the detection of volumetric defects characterized by abnormal progression across slices. Extensive experiments on industrial BGA-CT data demonstrate that PV-CDMNet achieves state-of-the-art performance, enabling reliable detection of multi-scale and volumetric defects while meeting industrial runtime requirements.**

Keywords: **Anomaly detection, BGA detection, Channel-disentangled memory, Industrial computed tomography, Pseudo-volumetric, Unsupervised learning.**

I. Introduction

Ball Grid Array (BGA) detection is critical for IC quality assurance but faces challenges with conventional 2D X-ray due to depth ambiguity[1]. While CT offers volumetric insights, it introduces a dilemma: slice-wise analysis loses inter-slice context, whereas 3D image processing is computationally expensive.Compounding these challenges is the inherent data scarcity in mature manufacturing lines. With defect rates typically measured in parts per million (PPM), supervised learning is impractical due to insufficient data. Furthermore, the extreme scale diversity of anomalies ranging from micrometer-scale texture anomalies(e.g., micrometer-scale voids) to macroscopic structural deformations (e.g., solder bridging) poses a significant challenge for current unsupervised algorithm, which often fail to simultaneously detect such diverse anomalies.

Memory bank based and reconstruction based methods have emerged as two promising solutions for BGA detection. Memory bank based methods (e.g., PatchCore [2]) leverage stored normal patterns to effectively detect structural anomalies but their reliance on pre-trained representations limits transfer from RGB images to CT data, reducing sensitivity to micro-voids. Conversely, reconstruction-based methods benefit from end-to-end training, which enables stronger adaptability to CT data, but suffer from the "identity mapping" problem, leading to many missed detections.

To overcome this limitation, we propose PV-CDMNet, a unified framework that integrates memory-bank and reconstruction methods. Our main contributions are as follows:

(1) Pseudo-Volumetric Data Construction: Efficiently encoding 3D context via Gaussian Interval Sampling to handle slice variations without expensive 3D computation.

(2) Hierarchical Channel-Disentangled Memory (H-CDM): Dividing feature spaces into multiple semantic subspaces, each maintained by an independent memory bank, to prevent subtle defect features from being suppressed by dominant background features.

(3) 3D Spatial Gradient Loss: Enforcing inter-slice consistency to detect volumetric anomalies that are difficult to identify using 2D slices alone.

II. Related Work

Current unsupervised anomaly detection algorithms primarily fall into three categories: feature representation methods, reconstruction methods, and vision-language methods. Feature representation methods achieve strong performance by mapping normal patterns to compact embeddings, including strategies such as Memory Bank(e.g., PatchCore[2]), Distribution Map (e.g., PaDiM[3], FastFlow[4]), and Teacher-Student Architecture (e.g.,

Manuscript received January 28, 2026. This work was supported by the National Key Research and Development Program of China (grant No. 2022YFF0607802). (Corresponding author: Yuxiang Xing.)

Y. Bai, X. Zou, and Y. Xing are with the Department of Engineering Physics, Tsinghua University, Beijing 100084, China, and also with the Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, Beijing 100084, China(e-mail: baiyk21@mails.tsinghua.edu.cn; zouxiang@skyverse.cn; xingyx@mail.tsinghua.edu.cn).

STPM[5]). However, these methods largely rely on encoders pretrained on ImageNet, which creates a significant domain gap on CT data, often failing to distinguish subtle attenuation shifts from CT artifacts. Conversely, reconstruction methods (e.g., DRAEM [6], DiAD[7]) enable end-to-end training but remain prone to the identity mapping where macroscopic deformations are faithfully reconstructed rather than repaired, resulting in missed detections. Furthermore, emerging vision-language methods (e.g., WinCLIP[8]) lack the domain-specific semantic grounding required for industrial CT, and BGA detection is strongly influenced by metal artifacts. In the volumetric domain, while 3D anomaly detection has advanced in medical settings, direct voxel-level processing is computationally expensive and difficult to deploy in industrial production-line settings[9]. To address these limitations, we propose PV-CDMNet. Our approach utilizes pseudo-volumetric data construction to encode 3D context within efficient 2D pipelines, integrates H-CDM and a size-adaptive dynamic masking strategy to prevent identity mapping, and introduces 3D spatial gradient constraints to detect volumetric defects with abnormal slice-to-slice variations.

III. Methodology

A. Overview

Let $\mathcal{D}$ be a training set containing only normal BGA samples. For a pseudo-volumetric input tensor $\mathbf{X}$, the standard goal of Unsupervised Anomaly Detection (UAD) is to learn a reconstruction network Φ that minimizes the reconstruction error $\|\mathbf{X} - \Phi(\mathbf{X})\|_2^2$. Ideally, Φ should generalize well to normal patterns but fail to reconstruct unseen anomalies, yielding high error scores. During inference, for a test sample, the model produces an anomaly map $\mathbf{A}$ based on the reconstruction residuals to localize defects.

PV-CDMNet is an end-to-end masked reconstruction framework consisting of three coupled components. The Pseudo-Volumetric Data Construction (Sec. III-B) module transforms raw 3D CT data into inputs that preserve Z-axis structural continuity while acting as a physics-informed augmentation. The Memory-Gated U-Net Backbone extracts multi-scale features, where the skip connections are intercepted by the H-CDM module(Sec. III-C). This design introduces a strict memory-gated mechanism that purifies the normal feature memory. Finally, the Joint Optimization Objective (Sec. III-D) constrains the model simultaneously in the pixel intensity, structural gradient, and latent feature domains.The overall network architecture is illustrated in Fig. 1.

B. Pseudo-Volumetric Data Construction

To capture Z-axis structural continuity while avoiding the high cost of full 3D convolutions, we propose a Gaussian Interval Sampling (GIS) strategy. For a target slice $\mathbf{X}_z$ at depth z, we construct a pseudo-volumetric triplet input $\mathbf{X}_{\text{in}} = [\mathbf{X}_{z-\delta}, \mathbf{X}_z, \mathbf{X}_{z+\delta}]$, where the sampling interval δ follows a truncated Gaussian distribution $\delta \sim \mathcal{N}(\mu, \sigma^2)$. The mean μ is calibrated to the optimal physical distance for characterizing the cross-sectional evolution of volumetric defects. The variance σ^2 accommodates resolution variations and mechanical deviations across different acquisition devices, serving as a physics-informed data augmentation mechanism.

To prevent the network from learning trivial identity mappings, we increase the difficulty of the reconstruction task by reformulating it as a self-supervised inpainting task and designing a Size-Adaptive Dynamic Masking strategy. Specifically, binary masks $\mathbf{M}$ are generated with block dimensions $(h_{\text{m}}, w_{\text{m}})$ sampled proportionally to the estimated object size $(h_{\text{obj}}, w_{\text{obj}})$ via scaling factors α, ensuring scale-invariant difficulty. We implement an inverse correlation between mask scale and occlusion ratio: Micro-Masking (small α, high ratio) occludes fine details to enforce high-frequency texture repair without local interpolation, while Macro-Masking (large α, low ratio) covers substantial structural components to compel the inference of global topological logic. This multi-scale approach maintains consistent masking difficulty across varying solder joint geometries, effectively precluding shortcut learning.

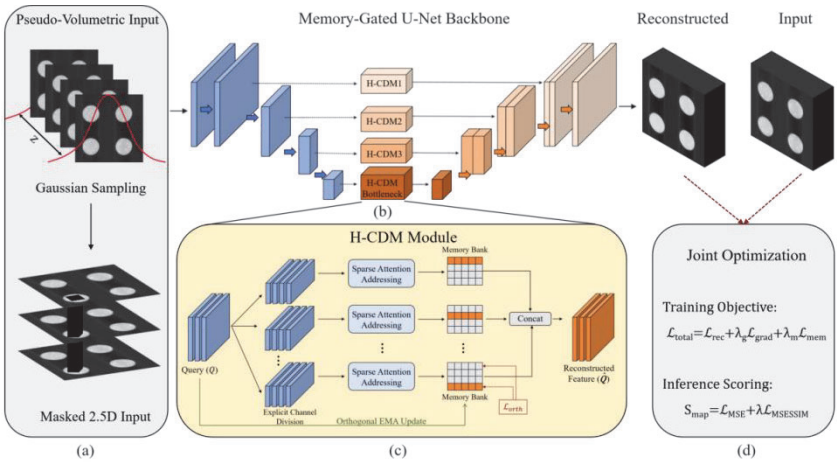


Fig. 1. The PV-CDMNet architecture. (a) Input: Pseudo-volumetric triplets via Gaussian Interval Sampling. (b) Backbone: U-Net utilizing H-CDM for feature sanitization. (c) H-CDM: Channel-disentangled memory with sparse retrieval. (d) Inference: Anomaly detection via multi-scale residuals.

C.Hierarchical Channel-Disentangled Memory (H-CDM)

Traditional memory networks suffer from background bias. This occurs because high-frequency background textures often dominate the distance metric. To address this issue, we embed the H-CDM module into the U-Net skip connections. Specifically, we divide the query feature map in the channel dimension into G independent semantic heads with separate memory banks. To encourage non-redundant representations, we apply an explicit orthogonality loss during training. This combination of physical isolation and mathematical constraint forces distinct subspaces to to become orthogonal, effectively disentangling structural geometry from fine-grained textures. As a result, it prevents background dominated channels from affecting the detection of foreground defect features.

In each subspace, sparse attention addressing is used to retrieve relevant prototypes. For brevity, we detail the process for a single semantic head. Let $\mathbf{q}$ denote a query vector from the current head. First, we calculate the Euclidean distance between $\mathbf{q}$ and the memory prototypes. To filter out noise, we retain only the Top-K nearest neighbors (i.e., smallest distances), forming the index set Ω . Subsequently, we reconstruct the feature $\widehat{\mathbf{q}}$ by computing the weighted sum of these selected prototypes:

$$\alpha_m = \frac{\exp\left(-\|\mathbf{q}-\mathbf{p}_m\|^2/\tau\right)}{\sum_{j\in\Omega}\exp\left(-\|\mathbf{q}-\mathbf{p}_j\|^2/\tau\right)} \quad (1)$$

$$\widehat{\mathbf{q}} = \sum_{m\in\Omega}\alpha_m \cdot \mathbf{p}_m \quad (2)$$

where α_m represents the normalized attention weight, and $\mathbf{p}_m$ denotes the corresponding prototype vector. The term $\|\mathbf{q}-\mathbf{p}_m\|^2$ denotes the squared Euclidean distance, and τ is the temperature parameter. The reconstructed features from all semantic heads are subsequently concatenated to form the final memory-gated output. This selective reconstruction functions as a physical denoising filter. Consequently, the output $\widehat{\boldsymbol{q}}$ is constrained to the retrieved normal prototypes, ensuring that abnormal features deviating from the normal prototype manifold, thereby suppressing abnormal deviations.

The reconstructed features replace the original encoder features on the skip connections establishing a strict information bottleneck. This ensures that only information compatible with normal prototypes reaches the decoder. Finally, to prevent the influence of inevitable anomaly contamination, we discard unstable gradient updates and use an Orthogonal Exponential Moving Average (EMA) strategy:

$$\mathbf{p}_m \leftarrow f_{\text{norm}}(\beta \cdot \mathbf{p}_m + (1-\beta)\cdot \bar{\mathbf{q}}_m) \quad (3)$$

where β is the momentum coefficient, and $\bar{\mathbf{q}}_m$ represents the mean vector of all queries assigned to prototype $\mathbf{p}_m$ within the current batch. The function $f_{\text{norm}}(\cdot)$ indicates L_2 normalization. The EMA update strategy defined in Eq. (3) functions as a temporal low-pass filter, stripping away sporadic abnormal features and CT artifacts to store only statistically stable normal patterns in training.

D.Optimization and Inference

We train the network using a hybrid objective function defined as :

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rec}} + \lambda_{\text{grad}}\mathcal{L}_{\text{grad}} + \lambda_{\text{mem}}(\mathcal{L}_{\text{orth}} + \mathcal{L}_{\text{comp}}) \quad (4)$$

The reconstruction loss $\mathcal{L}_{\text{rec}}$ combines Mean Squared Error (MSE) and Structural Similarity (SSIM) between the input volume $\mathbf{X}$ and the reconstruction $\widehat{\mathbf{X}}$, encouraging consistency in both pixel intensity and local structural details. To further detect volumetric anomalies, we introduce a 3D Spatial Gradient Loss that constrains gradients along the three spatial dimensions:

$$\mathcal{L}_{\text{grad}} = \|\nabla_{xy}\mathbf{X} - \nabla_{xy}\widehat{\mathbf{X}}\|_1 + \lambda_z\|\nabla_z\mathbf{X} - \nabla_z\widehat{\mathbf{X}}\|_1 \quad (5)$$

where ∇_{xy} and ∇_z denote gradients along the planar and depth dimensions, respectively. Specifically, ∇_z is calculated as the absolute difference between neighboring slices to constrains the Z-axis consistency. This is critical for detecting defects that appear normal in individual slices but exhibit abnormal cross-slice transitions. In addition, we introduce memory regularization, including orthogonality loss($\mathcal{L}_{\text{orth}}$) and compactness loss($\mathcal{L}_{\text{comp}}$), to constrain the latent feature space.

During inference, masking is disabled, and the input volume is reconstructed as a normal approximation $\widehat{\mathbf{X}}$. We calculate the pixel-level anomaly map $\mathbf{A}$ by aggregating intensity differences and structural distortions:

$$\mathbf{A} = \|\mathbf{X} - \widehat{\mathbf{X}}\|_2^2 + \lambda_{\text{infer}} \cdot (1 - \text{SSIM}(\mathbf{X}, \widehat{\mathbf{X}})) \quad (6)$$

where λ_{infer} balances the pixel-wise MSE error and the SSIM-based structural deviation. Powered by the learned pseudo-volumetric context, the model inherently captures z-axis inconsistencies through reconstruction residuals, without requiring explicit gradient computation in inference. The final image-level anomaly score is obtained by averaging the top-K pixel values in $\mathbf{A}$, robustly quantifying deviation s from the learned normal distribution.

IV.EXPERIMENTS

A.Experimental Setup

We employ a dual-dataset strategy to ensure industrial rigor. The Real-BGA dataset consists of 200 verified normal samples acquired from real industrial CT systems with varying resolutions. In addition, a small number of abnormal samples are collected and used exclusively for qualitative visualization and analysis For benchmarking, we constructed Syn-BGA by synthesizing defects (voids, bridging, component shifts) onto held-out normal backgrounds, creating 100 test cases with pixel-level ground truth. All experiments were conducted on a single NVIDIA RTX 3090 using the AdamW optimizer (batch size 8).

B.Quantitative Evaluation.

As detailed in Table I, our method achieves SOTA performance on Syn-BGA with 93.6% Image-level AUROC (I-AUROC) and 93.8% Pixel-level AUROC (P-AUROC). The significant margin over the strongest baseline (PatchCore, 91.0%) confirms that our approach effectively breaks the "2D ceiling" imposed by geometric ambiguity. Furthermore, with an inference speed of 38 FPS, our method is twice as fast as memory bank approaches like PatchCore (18 FPS), comfortably satisfying real-time industrial requirements .

TABLE I
QUANTITATIVE COMPARISON ON SYN-BGA.

Method	I-AUROC	P-AUROC	FPS
PaDiM	84.8	87.0	25
STFPM	86.9	89.4	55
FastFlow	88.8	91.2	32
PatchCore	91.0	92.2	18
DRAEM	88.9	93.1	22
Ours	**93.6**	**93.8**	**38**

C.Qualitative Analysis

We evaluate performance across four representative defect modalities: micro-voids, foreign materials, solder bridging, and component shifts. Fig. 2 presents a visual comparison of superimposed anomaly heatmaps, with the final red-circle localizations (last column) derived from PV-CDMNet. For texture-based anomalies like micro-voids, foreign materials, and solder bridging, PV-CDMNet demonstrates superior precision, while DRAEM and FastFlow detect voids but produce diffused heatmaps, and PatchCore suffers from activation "bleeding" into normal regions, PV-CDMNet generates concentrated responses with high structural fidelity.

The most decisive advantage of PV-CDMNet is the exclusive detection of component shifts. Due to geometric ambiguity where 2D projections of shifted balls remain circular, all baselines fail to trigger (blue heatmaps). By explicitly modeling Z-axis continuity, PV-CDMNet captures the implicit inter-slice misalignment that eludes spatial-only methods, effectively breaking the "2D ceiling" to accurately localize hidden topological anomalies.

V.CONCLUSION

In conclusion, we propose PV-CDMNet, a unified unsupervised framework for industrial CT anomaly detection. By embedding the H-CDM into a multi-scale reconstruction architecture, our method is able to captures microscopic texture anomalies as well as macroscopic structural deformations. Pseudo-volumetric data construction, combined with 3D spatial gradient constraints, allows for efficient 3D-aware processing compatible with industrial production-line throughput. Experiments on industrial BGA-CT data demonstrate that PV-CDMNet achieves state-of-the-art detection performance and meets the high-throughput requirements of real-world manufacturing lines. In the future, we will generalize these mechanisms for more volumetric detection tasks.

REFERENCES

[1] T. E. Taha, E. Emary, and K. Moustafa, "Automatic detection of ball grid array solder joints using X-ray images," Int. J. Adv. Manuf. Technol., vol. 76, no. 9, pp. 1519 – 1529, Feb. 2015.

[2] K. Roth et al., "Towards total recall in industrial anomaly detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, 2022, pp. 14318 – 14328.

[3] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "PaDiM: A patch distribution modeling framework for anomaly detection and localization," in Proc. Int. Conf. Pattern Recognit. (ICPR), Milan, Italy, 2021, pp. 475 – 489.

[4] J. Yu et al., "FastFlow: Unsupervised anomaly detection and localization via 2D normalizing flows," arXiv preprint arXiv:2111.07677, 2021.

[5] G. Wang et al., "Student-teacher feature pyramid matching for unsupervised anomaly detection," in Proc. Brit. Mach. Vis. Conf. (BMVC), 2021.

[6] V. Zavrtanik, M. Kristan, and D. Skočaj, "DRAEM - A discriminatively trained reconstruction embedding for surface anomaly detection," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Montreal, QC, Canada, 2021, pp. 8330 – 8339.

[7] H. He et al., "DiAD: A diffusion-based framework for multi-class anomaly detection," in Proc. AAAI Conf. Artif. Intell., vol. 38, no. 3, Feb. 2024, pp. 2000 – 2008.

[8] J. Jeong et al., "WinCLIP: Zero-/few-shot anomaly classification and segmentation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Vancouver, BC, Canada, 2023, pp. 19606 – 19616.

[9] P. Bergmann et al., "The MVTec 3D-AD dataset for unsupervised 3D anomaly detection and localization," in Proc. 17th Int. Jt. Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl. (VISAPP), vol. 4, 2022, pp. 202 – 213.

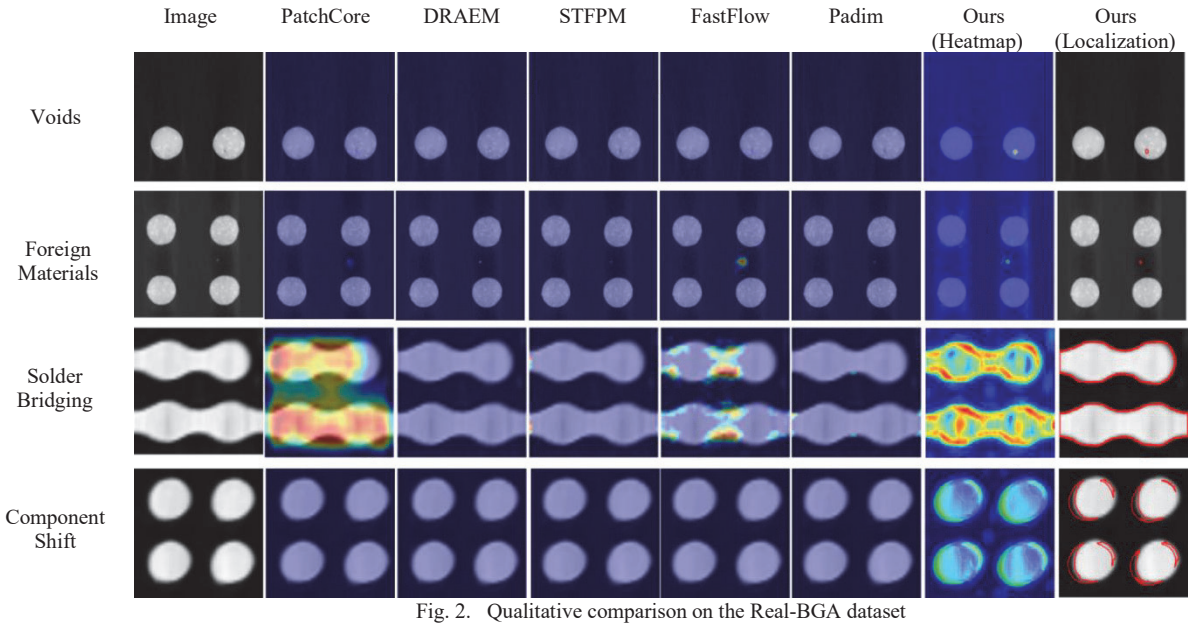


Fig. 2. Qualitative comparison on the Real-BGA dataset

Pulsed X-ray micro-CT for minimizing non-contributory irradiation during scanning

Takayuki Okamoto, Yuki Morimoto, and Hideaki Haneishi

***Abstract*—Micro X-ray computed tomography (micro-CT) enables nondestructive visualization of internal structures with micrometer-scale spatial resolution. A key limitation is the intrinsic trade-off between image quality and X-ray dose: higher signal-to-noise ratio and contrast typically require more projections and/or longer exposure time per projection, increasing the cumulative dose. This issue is particularly pronounced in vivo, where radiation exposure becomes a major constraint for repeated observations. Although iterative and deep learning-based reconstruction methods can mitigate image degradation under low-dose or sparse-view conditions, they primarily compensate for missing information and do not directly reduce the radiation delivered during imaging. Therefore, optimizing acquisition strategies, including the irradiation scheme, is essential for dose-efficient imaging in micro-CT. In this study, we propose pulsed X-ray irradiation aligned with the detector exposure window to minimize unnecessary dose delivered outside the image acquisition period. We developed a laboratory-scale micro-CT system capable of pulsed irradiation and evaluated the time-resolved dose-rate behavior and cumulative dose under continuous and pulsed emission. The time-resolved measurements showed that pulsed irradiation produced a near-background dose-rate baseline between pulses. In addition, compared with continuous emission, pulsed operation substantially reduced the cumulative dose. These results indicate that pulsed irradiation has the potential to suppress noncontributory irradiation and serve as a dose-efficient acquisition strategy for micro-CT.**

***Index Terms*—Micro-computed tomography (micro-CT), Pulsed X-ray irradiation, X-ray CT, X-ray dose rate.**

I. Introduction

MICRO X-ray computed tomography (micro-CT) is an imaging modality that enables nondestructive visualization of the internal three-dimensional structure of a specimen with micrometer-scale spatial resolution [1]. Owing to this capability, micro-CT has been widely used from basic to preclinical research, including morphological evaluation of ex vivo tissues and support for histopathological investigations [2,3], as well as longitudinal in vivo observations in small animals [4–6]. On the other hand, image quality in micro-CT is inherently traded off against X-ray dose: improving signal-to-noise ratio (SNR) and contrast typically requires increasing the number of projections and/or the exposure time per projection (or the number of averages), thereby increasing the cumulative dose. This issue is particularly pronounced in vivo, where radiation exposure becomes a major constraint for repeated observations [7]. Therefore, dose-efficient scan design that achieves the required image quality with the minimum necessary irradiation is important in micro-CT.

Representative approaches to dose reduction include (1) low-dose CT, which reduces the irradiation per projection [8], and (2) sparse-view CT, which reduces the number of projections [9]. However, low-dose CT tends to increase quantum noise, reducing SNR in projections and, consequently, degrading image quality in the reconstructed images. In contrast, sparse-view CT suffers from insufficient angular sampling, leading to severe streak artifacts in images reconstructed with analytical algorithms. Thus, straightforward dose reduction or projection-number reduction inevitably compromises reconstructed image quality.

To mitigate these challenges, numerous image quality enhancement methods have been proposed, including iterative reconstruction and deep learning-based image reconstruction techniques [10–12]. However, these approaches primarily aim to compensate for missing information under low-dose or sparse-view acquisition conditions and, therefore, are not necessarily a fundamental solution from the perspective of directly reducing the radiation dose delivered during imaging. Accordingly, to achieve both dose reduction and preservation of image quality, it is important to optimize acquisition strategies—beyond reconstruction algorithm design—by incorporating the imaging sequence and irradiation scheme.

A pulsed X-ray source capable of intermittent operation can be an effective means of addressing this issue. In conventional micro-CT, the X-ray source is typically operated with continuous X-ray emission; consequently, the specimen may continue to be irradiated even during detector readout or idle periods when the detector is not integrating, as well as during rotation-stage motion. Such irradiation during non-acquisition periods does not contribute to projection data formation, yet it increases the accumulated radiation dose. In this regard, when the X-ray output is temporally gated to synchronize with the detector exposure, irradiation is delivered only during the time windows required for projection acquisition, enabling dose-efficient imaging.

This work aims to minimize unnecessary X-ray irradiation outside the detector exposure window, which does not contribute to projection formation. We developed a laboratory-scale micro-CT system capable of pulsed X-ray irradiation. Using this system, we measured the time-resolved X-ray dose rate profiles under continuous and pulsed irradiation and quantitatively evaluated changes in cumulative

T. Okamoto is with the Center for Frontier Medical Engineering, Chiba University, Chiba 263-8522, Japan (e-mail: t_okamoto@chiba-u.jp).

Y. Morimoto is with the Faculty of Engineering, Chiba University, Chiba 263-8522, Japan.

H. Haneishi is with the Center for Frontier Medical Engineering, Chiba University, Chiba 263-8522, Japan.

dose as a function of exposure time per projection. Comparison experiments showed that pulsed irradiation produced a near-background dose-rate baseline between pulses. These findings underscore the potential utility of pulsed irradiation as an acquisition strategy that can improve dose controllability.

II. Materials and Methods

Micro-CT scanners generally consist of a micro-focus X-ray source, a two-dimensional flat-panel detector, and a rotation stage on which the object is mounted, forming a cone-beam CT geometry. These components are arranged such that the X-ray focal spot, the rotation center of the object, and the detector center are colinear, defining the system imaging axis. During scanning, X-rays transmitted through the sample are attenuated depending on its internal structure, and the transmitted intensity is recorded by the detector as projections. By rotating the sample over 360°, projections are acquired from multiple angles, and cross-sectional images are reconstructed from these projections using an image reconstruction algorithm.

Fig. 1 shows a custom-built micro-CT scanner. Our system is equipped with a sealed transmission-type micro-focus X-ray source (G-511VL-DP, Canon Anelva Corporation, Kanagawa, Japan) with a focal spot size of 2 μm, a tube voltage adjustable from 40 kV to 110 kV, and a maximum output power of 10 W. As the detector, we adopted a two-dimensional flat-panel detector (Shad-o-Box 1K HS, Teledyne DALSA, ON, Canada) with an active area of 5.7×6.4 cm^2, a matrix size of 1152 × 1130 pixels, a pixel pitch of 49.5 μm, and a maximum frame rate of 30 fps. The object is mounted on a motorized rotation stage for tomographic acquisition, and projections are acquired either in continuous rotation mode or in step-and-shoot mode.

The X-ray source supports pulsed irradiation, in which X-rays are generated by driving the tube current only for a specified pulse width. In our system, in addition to conventional continuous irradiation, detector-window-aligned pulsed irradiation is available, and the pulse width can be configured from 50 to 2000 ms in 10-ms increments. Continuous irradiation is supported in both step-and-shoot mode and continuous rotation mode, whereas pulsed irradiation is implemented only in step-and-shoot mode.

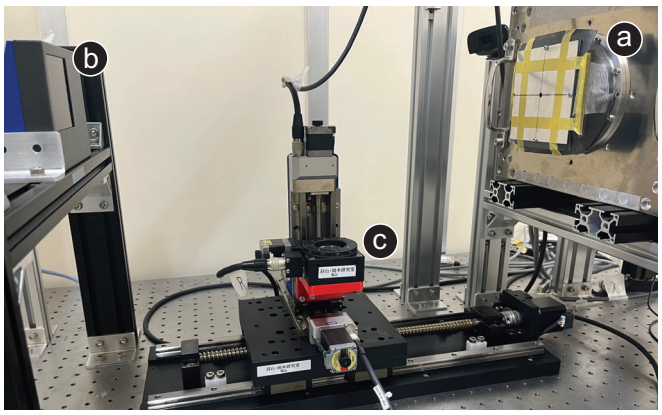


Fig. 1. Photograph of the custom-built micro-CT scanner. (a) Microfocus X-ray source with pulsed irradiation capability, (b) flat-panel detector, and (c) rotation stage.

III. Experiments

A. Measurement setup

To measure the time-resolved X-ray dose rate and cumulative dose during a scan, no object was placed on the rotation stage. Instead, a dosimeter (Cobia Flex R/F, RTI Group, Mölndal, Sweden) was attached to the surface of the flat-panel detector. The tube voltage and tube current were fixed at 50 kV and 100 μA, respectively, and the source-to-detector distance was set to 486.4 mm. For dose rate measurements, the number of projections was fixed at 300 across all scans to provide a standard acquisition condition for direct comparison.

For comparison between continuous and pulsed irradiation, the exposure time per projection was matched. In pulsed irradiation, the pulse width was set to the prescribed exposure time. In continuous X-ray irradiation, the exposure time per projection was set to $1/f$, where f is the detector frame rate. For the time-resolved dose-rate measurements of pulsed emission, pulses were generated in a standalone configuration without hardware synchronization to the detector exposure signal.

B. Evaluation protocol

The system was evaluated using the following three analyses.

(1) In step-and-shoot mode, the temporal dose-rate profiles (mGy/s) under continuous and pulsed irradiation were compared, and the dose-rate behavior was qualitatively assessed. In addition, the rise and fall times of the pulsed irradiation were calculated.

(2) In step-and-shoot mode, the exposure time was varied under both continuous and pulsed irradiation. The cumulative dose (Gy) per scan (300 projections) was measured, and the relationship between cumulative dose and exposure time was evaluated.

(3) For two conditions, continuous rotation mode with continuous irradiation and step-and-shoot mode with pulsed irradiation, the exposure time per projection was varied and the cumulative dose per scan (300 projections) was measured. The cumulative doses were then compared to evaluate differences between the two acquisition conditions.

IV. Results and Discussion

Fig. 2 shows the temporal profile of the X-ray dose rate during CT acquisition in step-and-shoot mode under continuous and pulsed irradiation with a 200-ms pulse width. Under continuous irradiation, the dose rate remained approximately constant at 0.18 mGy/s, whereas under pulsed irradiation, it decreased to a near-background level between pulses. These results indicate that, in the continuous mode, X-rays were emitted even during periods that did not contribute to the detector's effective exposure. In contrast, in the pulse mode, X-ray emission was largely confined to the prescribed pulse windows, suggesting a high potential for reducing unnecessary radiation exposure.

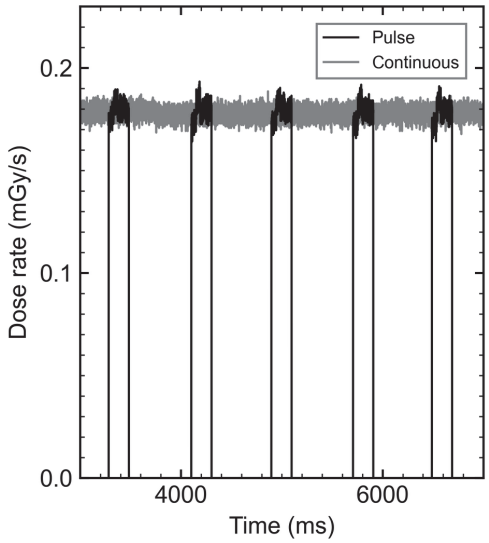


Fig. 2. Temporal profiles of X-ray dose rate (mGy/s) in step-and-shoot mode under continuous irradiation and pulsed irradiation (pulse width: 200 ms). The continuous mode exhibited an approximately constant dose rate of 0.18 mGy/s, whereas the pulse mode showed dose-rate increases only during irradiation periods and was nearly zero otherwise.

Fig. 3 presents a magnified view of the rising and falling edges of the dose-rate waveform under pulsed irradiation shown in Fig. 2. The rise and fall times were approximately 1.3 ms and approximately 0.6 ms, respectively, which were sufficiently short compared with the configurable pulse widths (50–2000 ms). Consequently, the impact of the transition intervals during ON/OFF switching, such as unintended additional irradiation or insufficient exposure, was considered small, suggesting that the X-ray output was effectively gated by the prescribed pulse width. In addition, because the total opening time of typical X-ray mechanical shutters is on the order of several to several tens of milliseconds, the rise and fall times of our system were comparable to or shorter than this timescale, indicating that the X-ray output could be gated with high temporal resolution.

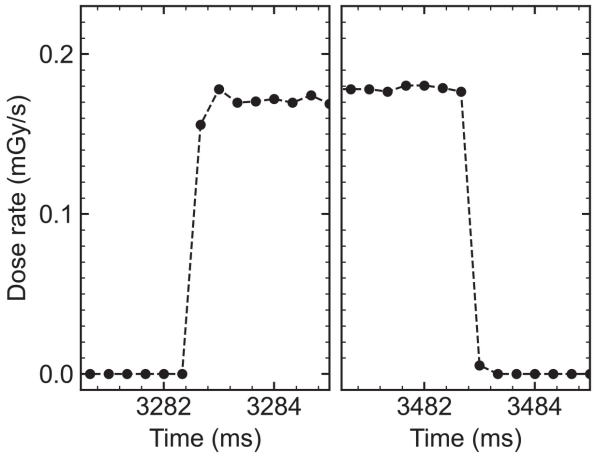


Fig. 3. Enlarged view of the rising and falling edges of the X-ray dose-rate waveform under pulsed irradiation shown in Fig. 2. The dose rate reached a steady-state level within approximately 1.3 ms after irradiation onset and decreased to nearly zero within approximately 0.6 ms after irradiation termination. Markers indicate the discrete measurements acquired by the dosimeter.

Fig. 4 shows the variation in cumulative dose per scan as a function of the exposure time per projection in step-and-shoot mode under continuous and pulsed irradiation. For all exposure times, the cumulative dose was higher under continuous irradiation, indicating that X-ray emission persisted even during periods that did not contribute to effective exposure, such as rotational motion and detector readout/idle intervals, thereby adding unnecessary dose. Under pulsed irradiation, the relationship between cumulative dose and exposure time closely followed the linear regression line, yielding a high coefficient of determination (R^2=0.99997), whereas continuous irradiation exhibited a slight deviation from the linear fit and a relatively lower R^2 (R^2=0.94746). These results suggest that pulsed irradiation improves dose controllability by making the cumulative dose more directly dependent on the prescribed pulse width while reducing the cumulative dose compared with continuous irradiation.

Table 1 compares the cumulative dose between the continuous rotation mode with continuous irradiation and the step-and-shoot mode with pulsed irradiation. In both conditions, the cumulative dose increased linearly with exposure time, while the cumulative dose was slightly higher in the continuous-rotation, continuous-irradiation condition. Fig. 5 shows the percentage reductions in cumulative dose for step-and-shoot mode with pulsed irradiation, using continuous rotation mode with continuous irradiation as the reference. At exposure times of 100, 200, 400, 500, 625, and 1000 ms, the corresponding dose reductions were 23.85%, 11.59%, 5.64%, 5.22%, 3.86%, and 1.83%, respectively. Note that, because the pulse width was configurable only in 10 ms increments, 620 ms was used as the closest available setting to 625 ms in this comparison.

On the other hand, the scanning time was shorter in the continuous rotation mode because projections can be acquired without stopping the rotation, thereby reducing the overhead associated with stepwise rotation and settling in step-and-shoot acquisition. In contrast, step-and-shoot acquisition records each projection at a stationary angle and is therefore less susceptible to motion blur caused by mechanical vibration and

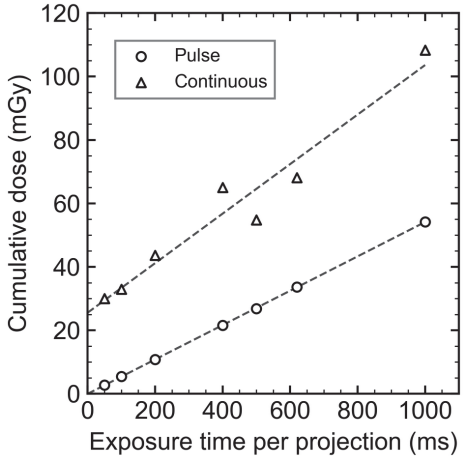


Fig. 4. Cumulative dose per scan (300 projections) as a function of exposure time per projection in the step-and-shoot mode under continuous and pulsed irradiation. Dotted lines represent independent linear regression fits for each irradiation mode.

TABLE I
CUMULATIVE DOSE AND DOSE REDUCTION FOR STEP-AND-SHOOT (PULSED X-RAY) RELATIVE TO CONTINUOUS ROTATION (CONTINUOUS X-RAY).

Exposure time (ms)	Continuous rotation (continuous X-ray)	Step-and-shoot (pulsed X-ray)	Dose reduction (%)
100	7.15	5.45	23.85
200	12.25	10.83	11.59
400	22.86	21.57	5.64
500	28.35	26.87	5.22
625 (620)	34.97	33.62	3.86
1000	55.21	54.20	1.83

Since pulsed X-ray irradiation was configurable only in 10-ms increments, 620 ms was used as the closest setting to 625 ms.

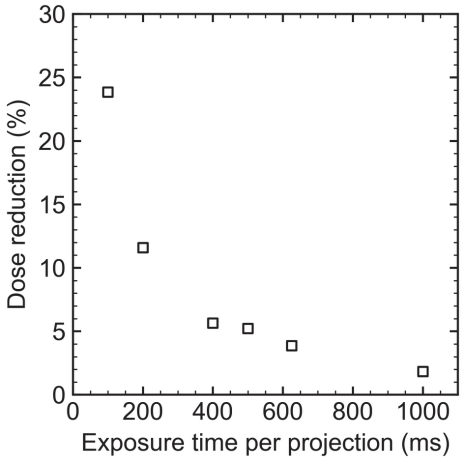


Fig. 5. Percentage reduction in cumulative dose as a function of exposure time per projection for step-and-shoot acquisition with pulsed irradiation, using continuous-rotation acquisition with continuous irradiation as the reference.

rotational-speed fluctuations. Accordingly, step-and-shoot acquisition may be advantageous when high projection accuracy is required. Moreover, although the time-resolved dose-rate measurements were performed in a standalone pulsed-irradiation configuration, the results suggest that pulsed irradiation can be extended to detector-window-aligned acquisition. This approach may reduce irradiation outside the effective exposure period, although small timing fluctuations should be considered in practical implementation. Overall, despite the longer acquisition time, this approach offers a practical trade-off by improving dose efficiency while preserving high projection accuracy.

V. CONCLUSION

We developed a micro-CT system capable of pulsed X-ray irradiation and evaluated the time-resolved behavior of dose rate and cumulative dose under continuous and pulsed irradiation. Time-resolved measurements showed that standalone pulsed irradiation produced a near-background dose-rate baseline between pulses. In step-and-shoot mode, pulsed irradiation reduced the cumulative dose compared with continuous irradiation by limiting X-ray emission to the prescribed irradiation periods. Moreover, relative to continuous-rotation mode under continuous irradiation, the step-and-shoot mode with pulsed irradiation achieved a comparable or slightly lower cumulative dose while preserving the advantages of stationary projection acquisition. These results indicate that the pulsed irradiation strategy has the potential to enable dose-efficient micro-CT by minimizing unnecessary exposure to the greatest extent possible while maintaining image quality.

ACKNOWLEDGMENT

This study was supported by AMED, Grant Number JP25hma922033h0001; JST, ACT-X, Grant Number JPMJAX23K5; and JSPS Program for Forming Japan's Peak Research Universities (J-PEAKS), Grant Number JPJS00420230002. We wish to thank Takashi Tomitsuka from Voxel Works Corporation (Tokyo, Japan) for his technical assistance in the construction of our micro-CT system.

REFERENCES

[1] A. du Plessis, C. Broeckhoven, A. Guelpa, and S. G. le Roux, "Laboratory X-ray micro-computed tomography: a user guideline for biological samples," *GigaScience*, vol. 6, no. 6, gix027, Jun. 2017.

[2] O. L. Katsamenis, M. Olding, J. A. Warner, D. S. Chatelet, M. G. Jones, G. Sgalla, *et al.*, "X-ray micro-computed tomography for nondestructive three-dimensional (3D) X-ray histology," *Am. J. Pathol.*, vol. 189, no. 8, pp. 1608–1620, Aug. 2019.

[3] R. Tang, J. M. Buckley, L. Fernandez, S. Coopey, O. Aftreth, J. Michaelson, *et al*., "Micro-computed tomography (Micro-CT): a novel approach for intraoperative breast cancer specimen imaging," *Breast Cancer Res. Treat.*, vol. 139, pp. 311–316, May 2013.

[4] J. H. Waarsing, J. S. Day, J. C. van der Linden, A. G. Ederveen, C. Spanjers, N. De Clerck, *et al*., "Detecting and tracking local changes in the tibiae of individual rats: a novel method to analyse longitudinal in vivo micro-CT data," *Bone*, vol. 34, no. 1, pp. 163–169, Jan. 2004.

[5] E. Namati, J. Thiesse, J. C. Sieren, A. Ross, E. A. Hoffman, and G. McLennan, "Longitudinal assessment of lung cancer progression in the mouse using in vivo micro-CT imaging," *Med. Phys*., vol. 37, no. 9, pp. 4793–4805, Sep. 2010.

[6] D. W. Holdsworth and M. M. Thornton, "Micro-CT in small animal and specimen imaging," *Trends Biotechnol.*, vol. 20, no. 8, pp. S34–S39, Aug. 2002.

[7] D. P. Clark and C. T. Badea, "Micro-CT of rodents: State-of-the-art and future perspectives," *Phys. Med.*, vol. 30, no. 6, pp. 619–634, Sep. 2014.

[8] C. H. McCollough, A. N. Primak, N. Braun, J. Kofler, L. Yu, and J. Christner, "Strategies for reducing radiation dose in CT," *Radiol. Clin. North Am*., vol. 47, no. 1, pp. 27–40, Jan. 2009.

[9] H. Kudo, T. Suzuki, and E. A. Rashed, "Image reconstruction for sparse-view CT and interior CT—introduction to compressed sensing and differentiated backprojection," *Quant. Imaging Med. Surg.*, vol. 3, no. 3, pp. 147–161, Jun. 2013.

[10] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.

[11] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, *et al*., "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 188–203, Jan. 2020.

[12] T. Okamoto, T. Ohnishi, and H. Haneishi, "Artifact reduction for sparse-view CT using deep learning with band patch," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 6, no. 8, pp. 859–873, Nov. 2022.

Quantitative measurement of 3D resolution in cone beam CT using touching spheres

Matthew Andrew. Author, *Carl Zeiss X-ray Microscopy*, Johannes Schule, *Carl Zeiss*, Konrad Lingel, *Carl Zeiss*

***Abstract*—Quantitatively assessing three-dimensional spatial resolution in cone-beam CT remains difficult due to anisotropic system behavior and the dependence of traditional methods on dedicated phantoms that do not reflect realistic imaging conditions. This work introduces a fully data-driven technique that utilizes naturally occurring contacts between touching spherical structures as intrinsic probes of system blur. By interpreting each sphere–sphere interface as an empirical edge-spread function, the method directly estimates the effective point-spread function within reconstructed volumes, enabling a direction-resolved 3D modulation transfer characterization without requiring specialized calibration objects. Through both simulation studies and experimental data acquired on a commercial X-ray microscope, we demonstrate strong consistency between this 3D spatial-domain metric and established 2D resolution measurements, while also highlighting the method's responsiveness to key factors such as X-ray source spot size, imaging trajectory, and reconstruction-induced blurring. Additional tests confirm positional invariance within practical samples, underscoring the robustness of the approach. Overall, this framework enables realistic, application-specific resolution assessment and provides a powerful tool for benchmarking, optimizing imaging parameters, and retrospectively evaluating image quality across diverse CT conditions.**

***Index Terms*—X-ray tomography, Image resolution**

I. Introduction

QUANTIFYING true three-dimensional (3D) spatial resolution in cone-beam computed tomography (CBCT) remains a longstanding challenge. Traditional modulation transfer function (MTF) measurements rely on calibration phantoms—such as wires, slanted edges, or bead arrays—that often fail to reflect the complex, heterogeneous structures found in practical imaging scenarios. Moreover, classical 2D MTF approaches measure response only along specific orientations and cannot capture anisotropic resolution behaviors characteristic of cone-beam geometry [1], [2], [3].

In many real volumetric datasets, however, touching or nearly touching spherical structures occur naturally—for example in powders, calibration bead packs, or biological tissues. These sphere–sphere contacts form narrow boundary transitions whose apparent thickness in the reconstructed volume arises entirely from blur introduced by the imaging and reconstruction pipeline. This work leverages such natural contacts as intrinsic resolution probes, enabling a fully data-driven estimation of the effective point-spread function (PSF) without reliance on dedicated phantoms, or indirect PSF inference [4].

We present a generalized spatial-domain framework for measuring 3D resolution by interpreting sphere contacts as empirical edge-spread functions (ESFs). By sampling many such interfaces across a volume, the method yields a direction-resolved, statistically robust description of system blur—effectively a 3D MTF. We validate this approach using simulation and physical experiments on a commercial X-ray microscope and demonstrate consistency with established 2D methods, as well as sensitivity to factors such as spot size, imaging trajectory, and reconstruction-induced blur.

II. Conceptual resolution measurement workflow

The proposed method operates entirely in the spatial domain. Whereas classical MTF procedures assess system resolution by Fourier-transforming edge or line-spread functions, the present framework evaluates the grey-value transitions that naturally occur between neighboring objects in the reconstructed volume. These transitions encode the effective PSF of the imaging chain and thus enable resolution to be inferred directly from the data itself without requiring dedicated test phantoms or explicit frequency-domain processing. Every contact between two segmented objects forms a local resolution probe. By analysing how intensity varies across such interfaces, the method quantifies the local blur along the interface normal. Repeating this process for many contacts throughout the volume yields a statistical, direction-resolved description of the systems spatial performance, providing a volumetric analogue of classical resolution characterization. The complete workflow consists of four stages: data preparation, geometry extraction, local response measurement, and aggregation.

A. Data Preparation

Establishes reliable object boundaries through denoising and segmentation. This step ensures that all subsequent geometric reasoning operates on physically meaningful structures rather than noise-driven artefacts.

B. Geometry Extraction

Computes signed distance fields for all segmented objects and identifies touching pairs. The midpoint between opposing interfaces defines the aperture field, which provides a one-dimensional coordinate axis aligned with the local interface normal.

All authors are with Carl Zeiss X-ray Microscopy, 5300 Central Parkway, Dublin, CA, USA (Contact e-mail: author@nrim.go.jp).

C. *Local Response Measurement*

Samples the reconstructed intensity along this aperture axis and normalizes the resulting transition with respect to surrounding material intensities. The outcome is a normalized one-dimensional transition curve representing the local PSF in physical space.

D. *Aggregation*

Compiles the individual normalized transition curves into a direction resolved characteristic function. The median curve forms a robust spatial-domain analogue of the classical MTF, from which concise blur descriptors such as the 10/90 % width, FWHM, or Gaussian σ can be derived.

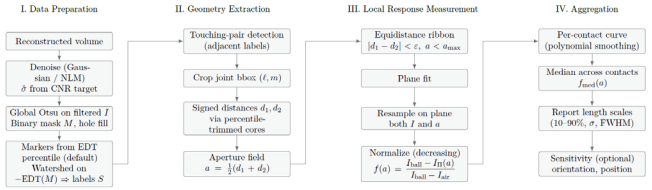


Fig 1. Schematic overview of the 3D-MTF workflow.

An example of two simulated touching balls are shown below in figure 2.

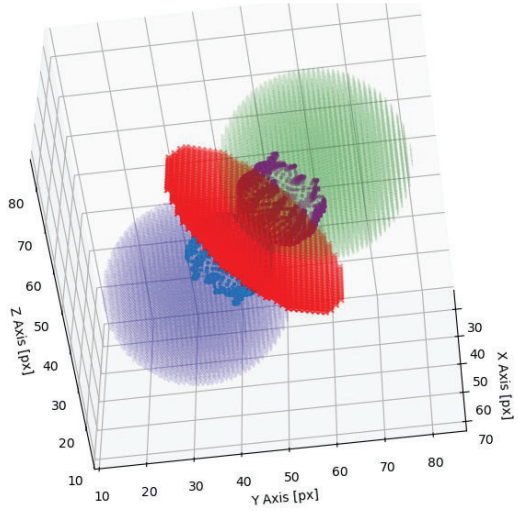


Fig 2. Two discretized touching balls, showing the separated volumes in blue and green, coupled with the equidistant sampling plane (red).

III. Assessment & Benchmarking - Simulation

Simulation was used to provide quantitative benchmarking between 3D resolution measurement and 2D resolution measurement. Note that this does not account for tomographic resolution loss associated with filtering, back projection, interpolation and reconstruction, but is more to compare response measurement techniques in a nominal imaging system. For this purpose a well described PSF (Gaussian blur) was used. 2D resolution measurement was performed by taking a nominal aperture (as in between two balls) and blurring it with a Gaussian PSF of varying sizes (figure 4) then measuring response cutoffs as a function of PSF size. As the Gaussian MTF is well described, it is possible to relate the theoretical MTF cutoffs with measured cutoffs from the touching spheres approach (figure 5).

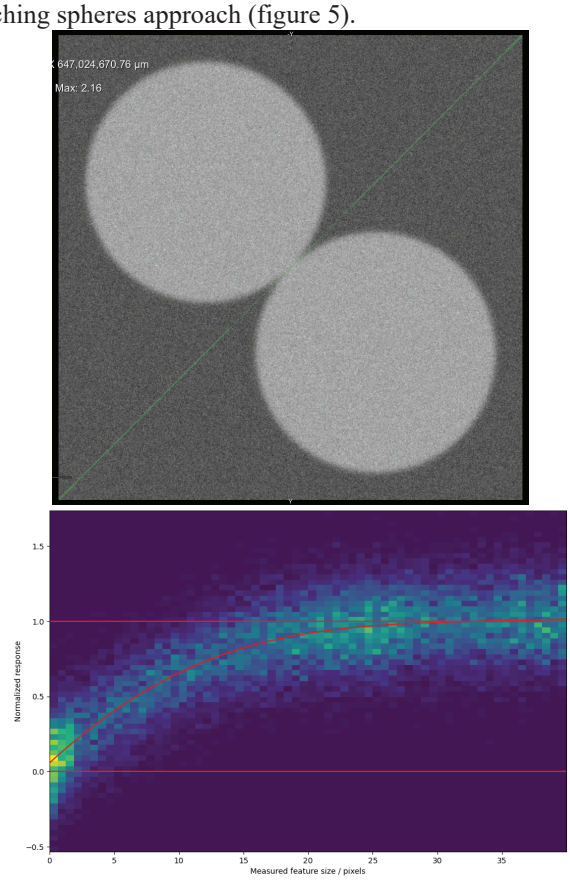


Fig 3. Cross sectional plane through a simulated 3D touching ball volume. The sampling plane is shown in green. The measured period response function 2D histogram is shown below, with its corresponding fitted response (in red) changing from 0 for very small features to 1 for very large features.

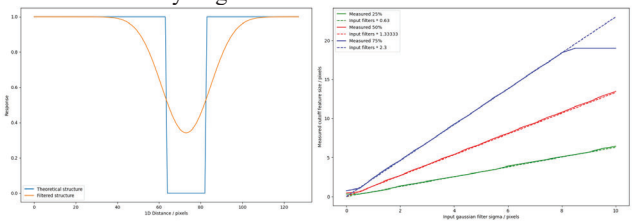


Fig. 4. Nominal aperture, blurred by a gaussian. Relationship between measured PSF cutoffs and input gaussian size (sigma) (blue = 75%, red = 50%, green = 25%).

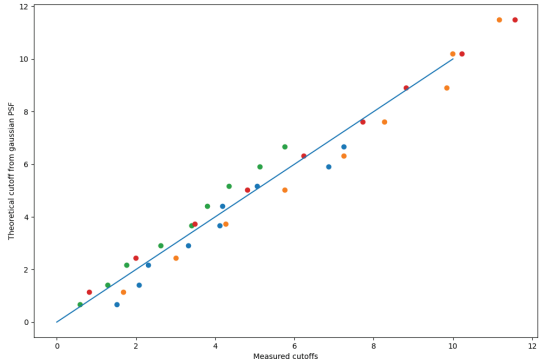


Fig 5. Comparison of measured MTF cutoffs from touching sphere approach (X axis) and theoretical cutoffs for a nominal aperture (Y axis, figure 4), showing good agreement. Blue corresponds to the 1:1 line.

This technique can then be used compare system resolution under varying conditions. This was performed using xrm_digital_twin, a simulation software specifically designed to simulate Zeiss X-ray microscopes. Two example experiments are included, one looking at the relationship between 3D spatial resolution and X-ray source spot size (figure 6) and the relationship between 3D spatial resolution and elevation angle in a rotational laminography acquisition trajectory (figure 7). Digital touching sphere samples were generated using a volumetric exclusion technique. Note that in these studies the system resolution is physically simulated on 2D projections, which are then reconstructed back into the volume domain using an FDK algorithm [5], so are subject to incremental blurring from reconstruction, in contrast to the measurement validation approaches used above.

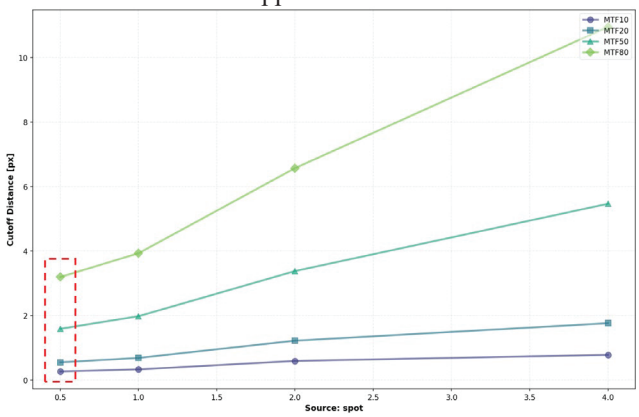


Fig 6. Relationship between 3D resolution and source spot size.

In the experiment relating 3D resolution and source spot size, there is a linear monotonic relationship between MTF cutoff and spot size, apart from at the smallest spot size (Gaussian sigma = 0.5 pixels). At this point cutoffs are slightly higher than we would otherwise expect based on extrapolating the observed trends at larger spot sizes. This is interpreted as being due to the impacts on resolution of filtering, back projection and interpolation during reconstruction, as well as sampling effects as the PSF size approaches the system sampling frequency.

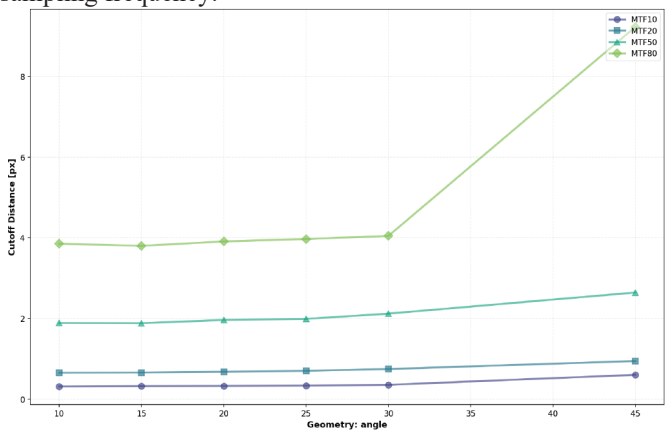


Fig 7. Relationship between 3D resolution and elevation angle in a rotational laminography trajectory.

The assessment of resolution in rotational laminography is complex as the non Tuy-complete nature of the imaging creates significant imaging artefacts [6], particularly as laminographic elevation angles increase beyond around 10 degrees. Initial studies of aggregate resolution (fig. 7) show very little impact of measured resolution as angle increases at low to intermediate angles. This is expected – elevation does not in itself impact resolution. At the highest angles (45 degrees in this study) significant changes in reported resolution are reported. It is likely that this is caused by the breakdown of the image processing workflow, particularly object segmentation & separation (workflow step II.A-B above). It would also be expected that measured resolution would be highly dependent on the orientation of the identified tangential sampling plane for a particular ball touching point. A detailed examination of this effect is beyond the scope of this study and would be an interesting topic for further investigation.

IV. Assessment and Benchmarking - Measurement

Physical touching sphere samples were created using commercially available mono-disperse SiO_2beads in the size range 63-75µm (www.cospheric.com). These were packed into a 1mm inner diameter Kapton (polyimide) tube, which were mounted onto commercially available 1mm diameter graphite rods. A range of experiments were conducted to characterize system resolution on a Versa XRM 730 under a range of conditions. Two experiments are reported here: first a baseline limiting resolution test using the high resolution 40X-P objective, examining positional invariance in the measurement across the sample. Then a resolution measurement in a high geometric magnification condition to examine how changes in source focus current significantly change measured system resolution.

The first experiment examined positional invariance in the measurement. The sample was imaged at 60kV with 3201 projections, 15mm source-to-sample distance, 5mm sample-to-detector distance, a 4s exposure time and a high resolution 40X-P objective at 4 locations axially distributed through the volume, separated by 1mm. These data were reconstructed to a nominal voxel size around 0.25um. Results are shown in figure 8.

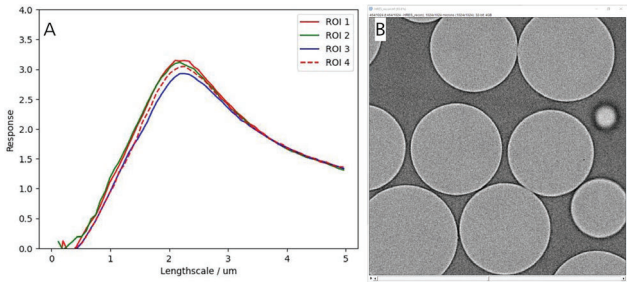


Fig. 8. Translational invariance of 3D resolution measurement.

The shape of the measured normalized imaging response function is remarkable in that it grows significantly above 1.0. This is expected, but counterintuitive, and is caused by the propagation phase contrast effect amplifying specific wavelengths beyond the low frequency (absorption contrast dominated) response used for MTF normalization. This effect

is enhanced further when two phase fringes overlap for nearby adjacent interfaces (e.g. the walls of a crack) which can give a strong response at those specific frequencies, in this case up to around a factor of 3 above the response seen at the lowest spatial frequencies.

The measurements from all four regions broadly agree in the shape, length scale, and distribution. This shows that the resolution measurement does not depend on the specific spheres used to make the measurement (translational invariance). Finally, while a detailed analysis of these results are beyond the scope of this paper, the limiting cutoff of these 3D feature response measurements (~5-10%) support the specified system resolution of 0.45µm. As the resolution specification and 3D measurement differ significantly in approach and definition, further study is needed to understand the relationship between different resolution metrics.

The second experiment shows how measured system resolution is a function of spot size, as determined by source focus current. To make the system strongly spot limited in its resolution, the sample was imaged at 60kV with 1601, 7mm source to sample distance, 40mm sample to detector distance, a 3s exposure time and a high resolution 4X objective. The relatively high geometric magnification (~6.7) makes the resulting system resolution strongly dependent on source spot size. The same region was imaged under multiple source focus conditions, controlled by increasing the source focus current from its optimal value by 0.25%, 0.5% and 0.75% (figure 9).

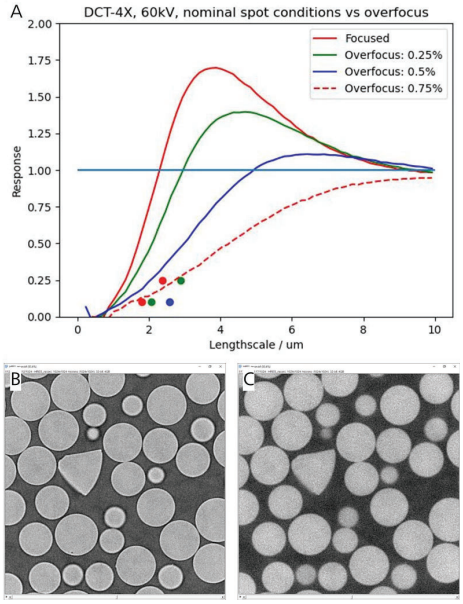


Fig 9. Impact of source overfocus on spot limited 3D resolution.

Increasing the focus current by 0.5% significantly reduces measured spatial resolution from ~1um 25% response cutoff to ~2.3um 25% response cutoff. Interestingly it also seems to significantly depress the measured phase response: a smaller spot has a much larger degree of coherency than a larger spot, greatly increasing the amount of observed phase phenomena, with the normalized response function maximum in the focused current case of ~1.75, and the over-focused case of ~1. This effect can also be qualitatively seen in reconstructed cross sections (figure 9).

V. Conclusions

This work presents a generalizable and fully data-driven approach for quantifying three-dimensional spatial resolution in cone-beam CT using touching spherical structures. By treating sphere–sphere contacts as intrinsic resolution probes, the method avoids reliance on specialized phantoms and instead derives effective point-spread behavior directly from reconstructed volumes. Through simulation and experimental validation, we demonstrate strong agreement between this 3D spatial-domain metric and established 2D resolution measurement techniques, as well as its sensitivity to system parameters such as source spot size, imaging trajectory, and reconstruction effects.

The framework enables robust, orientation-resolved resolution characterization under realistic imaging conditions and shows practical invariance across spatial positions and sample configurations. As such, it provides a powerful foundation for system benchmarking, parameter optimization, and retrospective image quality assessment.

Future work may further explore its behavior in complex acquisition geometries, particularly in regimes—such as high-angle laminography—where segmentation and sampling challenges become significant.

References

[1] H. Fujita et al., “A simple method for determining the modulation transfer function in digital radiography,” IEEE Transactions on Medical Imaging, vol. 11, no. 1, pp. 34–39, 1992.
[2] E. Samei et al., Acceptance Testing and Quality Control of Digital Radiographic Imaging Systems. American Association of Physicists in Medicine, 2017.
[3] P. Wu, J. M. Boone, A. M. Hernandez, M. Mahesh, and J. H. Siewerdsen, “Theory, method, and test tools for determination of 3d mtf characteristics in cone-beam ct,” Medical Physics, vol. 48, no. 6, pp. 2772–2789, 2021.
[4] C. Lee, H-D Song, and J. Baek, “3D MTF estimation using sphere phantoms for cone-beam computed tomography systems”, Medical Physics, vol. 47, no. 7, pp. 2838-2851, 2020
[5] L. Feldkamp, L Davis, and J Kress, “Practical cone-beam algorithm” Journal of the Optical Society of America A, vol. 1, no. 6, pp. 612-619, 1984
[6] Schulze, Ralf, et al. "Artefacts in CBCT: a review." Dentomaxillofacial Radiology, vol. 40, no. 5, pp. 265-273, 2011

SynthCaries: Dental Caries Generation on Panoramic Radiograph using Diffusion

Yoonha Eo*, Eunmin Lee*, and Jongduk Baek†

***Abstract*—Dental caries detection from panoramic radiographs remains challenging due to the scarcity of lesion-level annotations and the high cost of expert labeling. To address these limitations, we propose SynthCaries, a diffusion-based framework that generates realistic caries on panoramic radiographs. Our approach localizes regions of potential caries using tooth segmentation masks generated by the Segment Anything Model 3 (SAM3). After anatomically regulated preprocessing, a mask generator produces candidate caries location masks. These masks are used as conditioning inputs to guide anatomically plausible caries synthesis within a Stable Diffusion-based inpainting framework. During inference, synthetic caries are selectively generated within plausible regions while preserving surrounding structures. Experiments on downstream caries classification demonstrate that models trained solely on synthetic data achieve performance comparable to models trained on real data, underscoring the effectiveness of SynthCaries as a scalable data augmentation strategy.**

***Index Terms*—Diffusion Inpainting, Lesion Synthesis, Dental Caries, Panoramic X-ray**

I. Introduction

DENTAL caries remains one of the most prevalent chronic diseases worldwide, yet early detection through panoramic radiographs is hindered by the scarcity of annotated training data for deep learning models. While lesion synthesis has emerged as a powerful data augmentation strategy in other medical imaging domains, such as chest X-rays [1] and brain MRI [2], to the best of our knowledge, no lesion-level synthesis approaches exist specifically for dental panoramic radiographs. Existing generative research in dental imaging has primarily focused on whole panoramic X-ray or individual teeth generation rather than precise local caries lesion synthesis. Earlier generative approaches relied on GAN-based methods such as Pano-GAN [3] for panoramic radiograph synthesis, along with other tooth-level generation approaches [4]. Recently, diffusion models have been explored for panorama synthesis, exemplified by PanoDiff-SR [5], which generates low-resolution panoramas followed by super-resolution upsampling, and ToothMaker [6], which disentangles control for realistic dental X-rays. However, these methods focus on entire image structures or tooth morphologies, lacking the precision required for caries-specific lesion placement, shape, and density control within anatomically correct tooth contexts.

Diffusion-based inpainting uniquely addresses these limitations through iterative denoising conditioned on precise anatomical masks and caries attributes. Unlike GANs, which struggle with mode collapse and lack fine-grained control over lesion morphology, diffusion-based inpainting models progressively refine lesion textures while preserving surrounding structures [7]. This iterative refinement is particularly crucial for caries, where realistic radiopacity gradients and high-frequency textural details must coexist with complex tooth geometries and overlapping structures.

Our approach thus leverages diffusion inpainting's superior controllability and texture fidelity specifically for caries lesion synthesis, bridging the gap between entire panorama generation and the pixel-level precision required for realistic pathology augmentation in panoramic radiographs.

II. Method

A. Caries Mask Generation

The overall caries mask generation pipeline is illustrated in Fig. 1(a). Individual tooth regions were first extracted as polygon masks using a pretrained Segment Anything Model 3 (SAM3) [8] with a "tooth area" prompt. To prevent the overlap with existing pathological regions, any tooth polygon containing annotated caries pixels (red boxes) was excluded from the candidate set, ensuring that synthetic caries did not spatially coincide with real lesions.

For each remaining candidate tooth (blue overlay), a caries seed location was sampled from a bounded subregion within the tooth crown to ensure anatomically plausible placement. Since SAM3 provides segmentation only at the tooth level, the crown region was approximated heuristically as a constrained subregion of the tooth mask.

Starting from the sampled seed, an irregular caries shape was generated using a stochastic blob-growing algorithm [9]. Lesion growth was guided by a two-dimensional probability grid incorporating a base growth probability, a spatial likelihood field, and distance from the seed. Eligible pixels were iteratively aggregated within an 8-connected neighborhood to form a non-uniform cluster. Gaussian smoothing, morphological closing, and small-hole filling were subsequently applied to improve shape continuity while preserving natural boundary irregularity. Finally, the synthesized lesion mask was clipped to the corresponding tooth polygon to ensure confinement within the anatomical tooth boundary.

B. Training LoRA for Caries Inpainting

Next, instead of training the Stable Diffusion Inpainting model [10] from scratch, which requires huge computational cost, we finetuned the model using Low-Rank Adaptation

*These authors contributed equally to this work.
† Corresponding author
Department of Artificial Intelligence, College of Computing, Yonsei University, Republic of Korea (e-mail: yoonhaeo@yonsei.ac.kr; eunmin_lee@yonsei.ac.kr; jongdukbaek@yonsei.ac.kr).

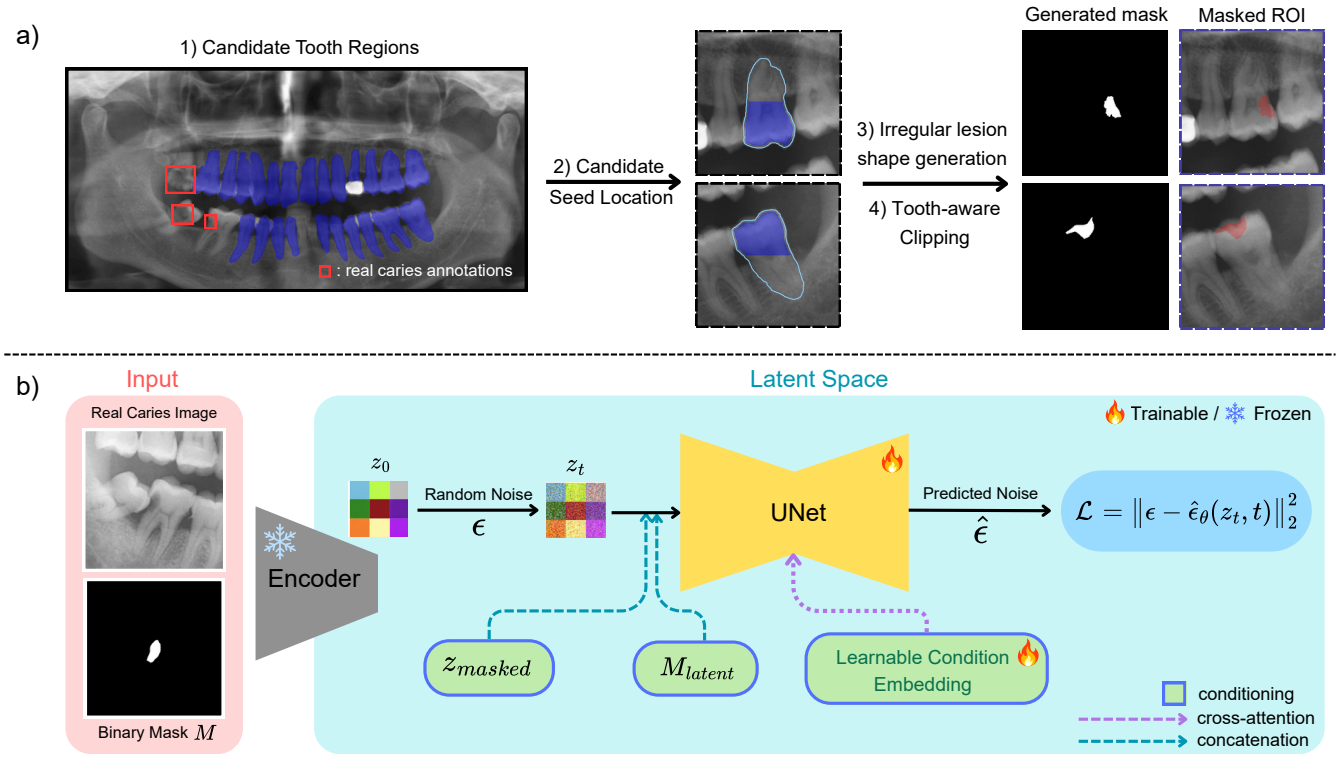


Fig. 1. The overall pipeline of (a) Caries Mask Generation and (b) Fine-tuning the Diffusion Inpainting Model.

(LoRA) by only training 13.6M parameters, which is approximately 1.6 % of the entire model. Moreover, the training was performed on 512×512 crop-wise region of interest (ROI) from 1500×3000 resized full images to focus the resource to reconstruct masked area of local tooth anatomy. The overall fine-tuning pipeline is illustrated in Fig. 1(b).

Given a crop-wise ROI and its binary caries mask M, the input image was encoded into a clean latent representation z_0 using the frozen VAE encoder. The original CLIP text encoder [11] was replaced with a learnable caries class embedding, which is injected into the U-Net via cross-attention for domain-specific semantic conditioning.

Following the Denoising Diffusion Probabilistic Model (DDPM) formulation, Gaussian noise $\epsilon_t \sim \mathcal{N}(0, I)$ was added at timestep t:

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon_t, \tag{1}$$

where $\bar{\alpha}_t$ denotes the cumulative product of the noise schedule.

Spatial conditioning followed the standard Stable Diffusion inpainting pipeline. A masked latent was constructed as

$$z_{\text{masked}} = (1 - M_{\text{latent}}) \odot z_0, \tag{2}$$

where M_{latent} denotes the binary caries mask M resized to the latent resolution.

The noisy latent z_t, masked latent z_{masked}, and mask M_{latent} are concatenated channel-wise and provided as input to the U-Net, allowing the network to synthesize caries within the masked regions by leveraging the surrounding anatomical context.

Finally, the network was trained to predict the injected noise using a mask-weighted loss, which assigns higher weights to pixels inside the caries mask to emphasize accurate lesion reconstruction while maintaining stability in the surrounding regions.

C. Synthetic Caries Generation

For inference, caries inpainting was performed on a selected ROI using a user-generated mask, which was generated following the pipeline shown in Fig. 1(a).

To reduce border effects and artifacts near mask boundaries, the binary caries mask M was converted into a continuous-valued soft mask M_{soft} using Gaussian blurring, scaled by a controllable factor to adjust the inpainting strength. This soft mask modulated the degree of corruption and regeneration in the diffusion process, enabling anatomically coherent synthesis.

Starting from z_t, the latent representation was iteratively denoised using the trained U-Net. We adopted Denoising Diffusion Implicit Models (DDIM) for inference, which enables efficient and deterministic sampling:

$$z_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\hat{z}_0 + \sqrt{1 - \bar{\alpha}_{t-1}}\epsilon_\theta(z_t, t). \tag{3}$$

Here, $\hat{z}_0$ denotes the estimated clean latent, $\epsilon_\theta(z_t, t)$ is the noise predicted by the U-Net, and $\bar{\alpha}_{t-1}$ represents the cumulative noise schedule. The U-Net was conditioned on the masked latent and the latent-space soft mask M_{soft}, which are concatenated with z_t along the channel dimension during inference.

During inference, the network reconstructed the ROI image, after which only pixels within the caries mask are replaced with the synthesized output, while all surrounding regions

are retained from the original image. This compositing step ensured strict preservation of non-target regions and pixel-level fidelity outside the caries area.

III. Experiments

A. Datasets

We conducted experiments on two publicly available dental panoramic radiograph datasets: DC1000[12] was used to train the caries-aware diffusion model, while AAAI Caries[13] was used to evaluate cross-dataset generalization and synthetic caries generation.

1) DC1000 Dataset: DC1000 consists of 597 panoramic radiographs with expert-annotated pixel-level caries polygons. For diffusion inpainting, 512×512 patches centered on caries regions were extracted, resulting in 1,081 image–mask pairs.

2) AAAI Caries Dataset: The AAAI Caries dataset contains 6,000 panoramic radiographs with bounding-box caries annotations. The diffusion model trained on DC1000 was applied to synthesize caries on caries-free teeth using the proposed mask generation method. To reduce redundancy, 1,000 images were randomly sampled, yielding 4,507 synthetic caries, which were used exclusively for downstream classification experiments.

B. Real Caries Classification Using Synthetic Data

To evaluate the availability of synthesized caries, we performed a downstream binary classification experiment (caries vs background) using YOLO26n-cls [14]. From the 4,507 generated caries, 760 high-quality samples were selected based on a confidence score of 1.00 predicted by a classifier trained on real data, and 320×320 patches centered on caries were extracted to construct the training set.

Across all experiments, the number of training caries samples was fixed to 760, consisting of either synthetic-only data or mixtures of synthetic and real caries at varying ratios. Background patches were sampled from non-caries regions, ensuring that at least 50% of each patch contains tooth structures. This prevented the model from trivially distinguishing caries and background based solely on the presence of teeth.

All validation and test sets contained only real data, each comprising 250 caries and 250 background samples.

IV. Results

A. Inpainting Results

Fig. 2 presents various inpainting results on panoramic radiographs. The proposed method generated visually realistic caries that are well-integrated with surrounding tooth structures, exhibiting smooth boundaries and anatomically plausible textures. Notably, the synthesized caries resembled real caries in terms of shape, contrast, and spatial distribution, despite being generated on teeth originally free of caries.

B. Classification Results

As shown in Table I, training with synthetic data achieved an F1-score of 0.9. Notably, this setting yielded the highest precision among all configurations, indicating a low false-positive rate when trained solely on synthetic caries. As the proportion of real data increased, recall and F1-score consistently improved, with the best performance obtained using only real data (F1-score 0.924). Mixed training setups showed intermediate performance, with the 5:5 synthetic-to-real ratio achieving an F1-score of 0.919, close to that of the real-only setting.

TABLE I
Classification Performance on AAAI Dataset under different synthetic-to-real train data ratios (Caries:Background = 1:1)

Train Data Ratio	Recall	Precision	F1-score	Accuracy
Synthetic Only	0.901	0.899	0.9	0.901
Syn : Real 8:2	0.928	0.889	0.908	0.906
Syn : Real 7:3	0.948	0.871	0.908	0.904
Syn : Real 5:5	0.968	0.874	0.919	0.914
Real Only	0.972	0.880	0.924	0.920

V. Discussion

The qualitative and quantitative results demonstrate the effectiveness of the proposed diffusion-based caries synthesis framework. As shown in Fig. 2, inpainted caries exhibit smooth boundaries, realistic contrast, and anatomically plausible textures that are well integrated with surrounding tooth structures, despite being generated in teeth that were originally caries-free.

The classification results further support the practical usage of the synthesized data. Models trained exclusively on synthetic caries achieve strong generalization to real test data (F1-score of 0.9), indicating that the synthesized lesions capture key discriminative characteristics of real caries. While incorporating real data yields the highest overall performance, mixed training settings result in only marginal degradation compared to the real-only baseline, with the 5:5 synthetic-to-real ratio achieving an F1-score of 0.919.

These findings suggest that synthetic caries can effectively complement or partially replace real annotations under limited labeling resources, substantially reducing the reliance on costly expert annotations while maintaining competitive performance in downstream diagnostic tasks.

VI. Conclusion

We presented a diffusion-based framework for controllable caries synthesis on dental panoramic radiographs. By combining anatomically guided mask generation with conditional diffusion-based inpainting, the proposed method generates localized and realistic caries conditioned on the surrounding tooth structures.

Experimental results demonstrate that the synthesized caries are diagnostically meaningful and effective for downstream classification, with minimal performance degradation even when replacing a substantial portion of real annotations. Overall, this work establishes diffusion-based caries synthesis as a practical solution to alleviate annotation scarcity in dental imaging and provide a scalable foundation for future caries-related learning tasks, including detection and segmentation.

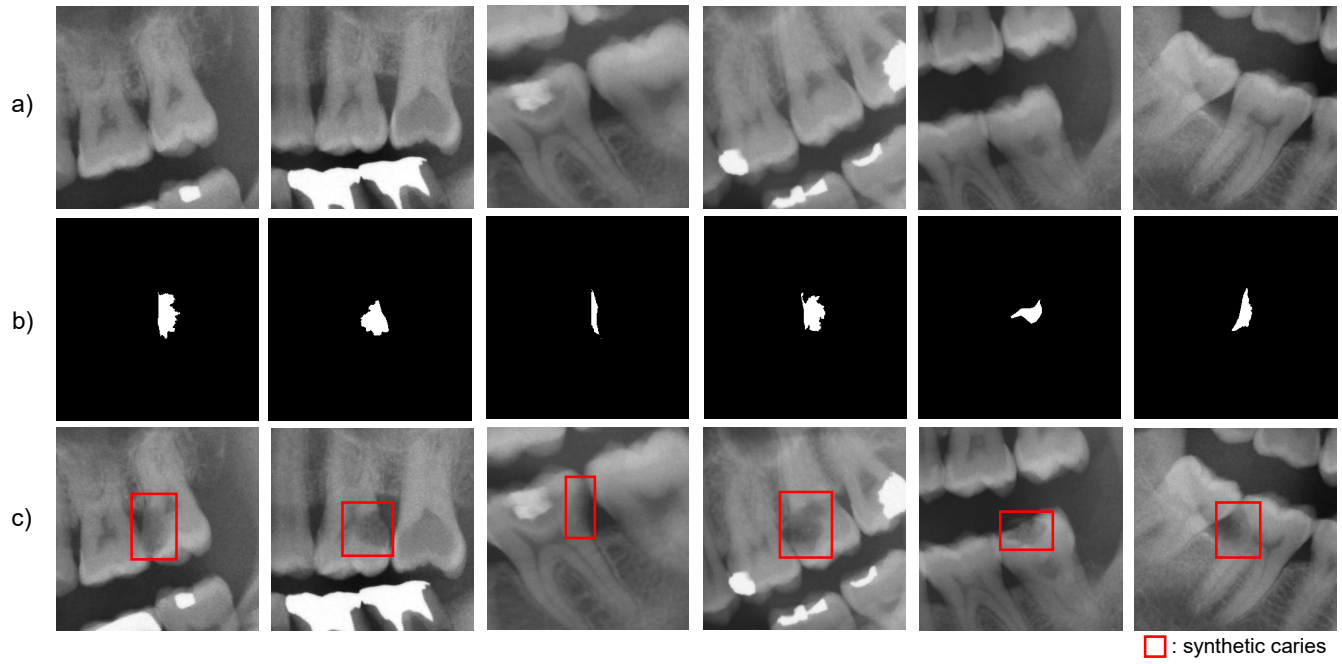


Fig. 2. (a) Original Images, (b) Generated Caries Masks, (c) Synthesized Caries-inpainted Images.

REFERENCES

[1] S. Singh, A. Goyal, A. Mittal, P. Kumar, P. Putha *et al.*, "A diffusion-driven fine-grained nodule synthesis framework for enhanced lung nodule detection from chest radiographs," in *Medical Imaging with Deep Learning*.

[2] O. Zamzam, H. Akrami, A. Joshi, and R. Leahy, "Modeling and reversing brain lesions using diffusion models," *arXiv preprint arXiv:2507.05670*, 2025.

[3] S. Pedersen, S. Jain, M. Chavez, V. Ladehoff, B. N. de Freitas, and R. Pauwels, "Pano-gan: A deep generative model for panoramic dental radiographs," *Journal of Imaging*, vol. 11, no. 2, p. 41, 2025.

[4] R. Nambiar and R. Nanjundegowda, "Tooth image synthesis using customized gan architectures: A deep learning-based solution for data scarcity in dentistry," in *2025 International Conference on Computing Technologies & Data Communication (ICCTDC)*. IEEE, 2025, pp. 1–7.

[5] S. Jain, B. N. de Freitas, A. Basse-OConnor, A. Iosifidis, and R. Pauwels, "Panodiff-sr: Synthesizing dental panoramic radiographs using diffusion and super-resolution," *arXiv preprint arXiv:2507.09227*, 2025.

[6] W. Yu, X. Guo, W. Li, X. Liu, H. Chen, and Y. Yuan, "Toothmaker: Realistic panoramic dental radiograph generation via disentangled control," *IEEE Transactions on Medical Imaging*, vol. 44, no. 12, pp. 5233–5244, 2025.

[7] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. Van Gool, "Repaint: Inpainting using denoising diffusion probabilistic models," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 11 461–11 471.

[8] N. Carion, L. Gustafson, Y.-T. Hu, S. Debnath, R. Hu, D. Suris, C. Ryali, K. V. Alwala, H. Khedr, A. Huang *et al.*, "Sam 3: Segment anything with concepts," *arXiv preprint arXiv:2511.16719*, 2025.

[9] R. Adams and L. Bischof, "Seeded region growing," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 16, no. 6, pp. 641–647, 1994.

[10] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2022, pp. 10 684–10 695.

[11] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark *et al.*, "Learning transferable visual models from natural language supervision," in *International conference on machine learning*. PmLR, 2021, pp. 8748–8763.

[12] X. Wang, S. Gao, K. Jiang, H. Zhang, L. Wang, F. Chen, J. Yu, and F. Yang, "Multi-level uncertainty aware learning for semi-supervised dental panoramic caries segmentation," *Neurocomputing*, vol. 540, p. 126208, 2023.

[13] B. Chen, S. Fu, Y. Liu, J. Pan, G. Lu, and Z. Zhang, "Cariesxrays: Enhancing caries detection in hospital-scale panoramic dental x-rays via feature pyramid contrastive learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 20, 2024, pp. 21 940–21 948.

[14] G. Jocher and J. Qiu, "Ultralytics yolo26," 2026. [Online]. Available: https://github.com/ultralytics/ultralytics

Gradient Step Plug-and-Play Model for Dental Cone-Beam CT Reconstruction

Idris Tatachak, Luis Kabongo, Nicolas Papadakis, Xavier Ripoche, Simon Rit

***Abstract*—The goal of this work is to reduce the effect of photon noise in dental cone-beam CT reconstruction. We consider an inverse problem formulation and develop a data-based prior. To this end, we simulate fan-beam acquisitions and add photon noise to the projection data. The prior is obtained by training a gradient-step denoiser using reconstructed simulated acquisitions. The trained model is integrated into a plug-and-play gradient-step algorithm to reconstruct images from simulated projections. Experiments on synthetic data demonstrate the denoising capabilities of the trained model, while qualitative evaluations on real images showcase the algorithm's performance and generalization ability.**

***Index Terms*—Computed Tomography Reconstruction, Deep-Learning, Plug-and-Play.**

I. Introduction

In computed tomography (CT), regularization plays a major role in improving the image quality of reconstructions from CT acquisitions. The reconstruction problem can be formulated as:

$$x^* \in \arg\min_x f(x) + \lambda g(x), \tag{1}$$

with $f(x)$ the data fidelity term, and $g(x)$ the regularization term weighted by the regularization parameter $\lambda \geq 0$. In the case of iterative CT reconstruction, the data fidelity term is usually taken as a quadratic loss as follows:

$$f(x) = \frac{1}{2}||Ax - p||_2^2. \tag{2}$$

Many techniques and regularization functions have been proposed to treat degraded data (due to under-sampling or limited angle scan) and correct artifacts (e.g. metal or photon starvation artifacts) present in CT data. These regularization techniques act as priors in our problem, injecting prior information about the object we aim to reconstruct. Recently, significant contributions have been made in the direction of data-based priors used in the context of model-based iterative reconstruction (MBIR) algorithms. These approaches leverage a prior $g(x)$ that is either explicitly defined (e.g., total variation) or implicitly given by an available model like an image denoiser. With the development of deep learning, the number and effectiveness of implicit models have increased drastically in the past years.

A notable class of algorithms used in MBIR is Plug-and-Play (PnP) algorithms [1], [2]. PnP algorithms use off-the-shelf image denoisers to incorporate an implicit prior in an iterative reconstruction algorithm. First proposed using classical denoisers such as BM3D [3], PnP methods have increasingly adopted deep denoisers [4], [5].

The performance and convergence behavior of iterative PnP algorithms are inherently tied to the properties of the denoiser D being employed. A main question in the development of PnP models thus concerns the properties required by the denoiser D to ensure the convergence of the iterative process providing the reconstruction. Many constraints have been proposed such as 1-Lipschitz constraint of the denoiser [6], which limits the network expressiveness, or firmly non-expansive constraint [7] that corresponds to a 1-Lipschitz operator $(2D - \mathrm{Id})$, in order to ensure the convergence of the iterates. This last constraint leads to a more expressive network, but it is not associated to an explicit regularization function g.

In this work, we leverage on the *Gradient Step Denoiser* [8] to build an iterative CT reconstruction algorithm. A key advantage of this denoiser is that it does not require an explicit penalization of the Lipschitz constant of its Jacobian, which can be challenging in practice [7]. In this work, we first train a deep gradient step denoiser using simulated fanbeam acquisitions and demonstrate its performance on real dental cone-beam CT (CBCT) acquisitions. Section II describes the simulated database generation process, the model training framework, and the reconstruction algorithm used. Section III presents reconstruction results obtained with our framework on real dental CT data. We finally conclude in section IV, pointing to potential improvements of the presented method.

II. Materials and Methods

In this section, we first present the Gradient Step Model used for regularization and then described the proposed approach, which consists of three sequential steps: database generation, model training for the general denoising task (independent of reconstruction), and image reconstruction using the trained model.

A. Gradient Step Denoiser

The Gradient Step Denoiser [8] is designed to minimize an explicit regularization term g defined as

$$g(x) = \frac{1}{2}||x - N_\sigma(x)||^2, \tag{3}$$

TATACHAK Idris and RIT Simon are with INSA-Lyon, Universite Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69373, Lyon, France. (email: simon.rit@creatis.insa-lyon.fr)

PAPADAKIS Nicolas is with Univ. Bordeaux, CNRS, Inria, Bordeaux INP, IMB, UMR 5251, F-33400 Talence, France. (email: nicolas.papadakis@math.u-bordeaux.fr)

RIPOCHE Xavier, KABONGO Luis, and TATACHAK Idris are with ACTEON Group, France.

Corresponding email: idris.tatachak@creatis.insa-lyon.fr

where $N_\sigma(x)$ is a differentiable image-to-image neural network depending on a noise level parameter σ, whose Jacobian is noted J_{N_σ}. The gradient of the regularization term ∇g is

$$\nabla g(x) = x - N_\sigma(x) - J_{N_\sigma}(x)^T(x - N_\sigma(x)) \tag{4}$$

and the final gradient step denoiser is defined as

$$D_\sigma = \mathrm{Id} - \nabla g = N_\sigma(x) + J_{N_\sigma}(x)^T(x - N_\sigma(x)). \tag{5}$$

This denoiser therefore modifies the original architecture N_σ to ensure it represents the gradient of an explicit function. This is a crucial property guaranteeing convergence when integrated into an iterative reconstruction process [8].

B. Data Generation

PnP models are usually trained on natural images, which has the advantage of providing a large amount of training data. Here, we train the model on images close to the distribution of real dental CT images to learn a prior that is better suited to our problem. To this end, we performed simulations of polychromatic fan-beam acquisitions of 2D slices of the XCAT phantom [9] using RTK [10]. Noise was then added to the simulated projections following a compound Poisson distribution, where the rate parameter was estimated from air acquisitions of real CBCT projections.

As a result, our dataset is constructed to provide paired examples required for training the model:

- **Input:** a reconstructed image obtained from a noisy acquisition.
- **Ground truth:** the corresponding reconstructed image obtained from a noise-free acquisition.

1) Phantom Preparation: Simulations were performed using the XCAT phantom [9], which provides a full-body material map. In total, 1490 slices of the XCAT phantom were extracted, each with a thickness of 1.1 mm, starting from the bottom of the phantom. Each slice has an in-plane pixel spacing of 0.14 mm, resulting in images of size 2092×2092 pixels.

To augment the dataset, patches of size 888×888 pixels were extracted from these slices. The patches were selected such that the field-of-view (FOV) contained at least 20% non-air material and at least 10% air, in order to avoid the interior problem (where projections are truncated on both sides, leading the solution to be non-unique [11]). Patches not satisfying these criteria were discarded, resulting in a total of 5,715 samples.

2) CT simulation: Acquisitions were simulated using the central slice geometry of a dental CBCT scanner. Line integrals were computed using the RTK CudaRayCast operator on the different materials m (e.g. water, muscle, cartilage, skull bone, ...), and a given machine spectrum. The measured intensity is computed as:

$$I = \sum_E S(E)\, \exp\left(- \sum_m \int_L \mu_m(x, E)\, \mathrm{d}x \right) \tag{6}$$

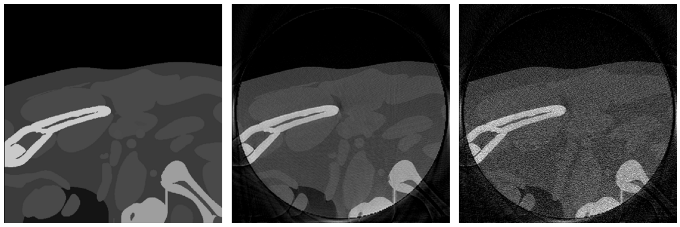


Fig. 1: Example slices of the XCAT phantom, displayed with the same window (0, 0.15) mm^{-1}. Left: Effective attenuation map of the region to be reconstructed, computed as the linear attenuation coefficient at the average spectrum energy. Center: Reconstructed image from simulated acquisitions with no simulated noise. Right: Idem with noise.

where $S(E)$ denotes the effective spectrum (emitted spectrum multiplied by the detector sensitivity) at energy level E, sampled from 0 to 90 keV in 0.5 keV intervals, for a tube voltage of 90 kV and 12.5 mAs. The index m denotes a material from the set of materials defined in the XCAT phantom.

3) Noise model: The noise is restricted to photon noise without electronic noise as it is dominant at the selected dose level. The photon noise in the simulated projections is modeled using a compound Poisson distribution resulting from the sum of Poisson realizations at each energy level,

$$N \sim \sum_E \mathcal{P}\left(S(E) \exp\left(- \sum_m \int_L \mu_m(x, E)\, \mathrm{d}x \right) \right). \tag{7}$$

4) Reconstruction and ground-truth: From the aforementioned simulated acquisitions, fan-beam CT images were reconstructed using the same geometry with 592×592 pixels with 0.15 mm spacing. Truncation of the projections resulting from the patient being larger than the FOV, common in dental CBCT, was handled by correcting the projections before reconstruction. The truncation correction was performed according to the approach based on the subtraction of the out-of-FOV forward projection of a coarse grid CT image, as detailed in [12], [13], with a coarse grid of 592×592 pixels of 0.4 mm.

For the ground truth, we used the reconstructed images from acquisitions without simulated noise (see Fig. 1). This choice was made instead of using the effective attenuation map in order to prevent the model from learning the correction of beam hardening and truncation artifacts.

5) Real acquisitions: Real dental CBCT acquisitions were used to qualitatively validate the reconstruction algorithm. No pre-processing or post-processing was applied to the data. In particular, calibration, dead-pixel correction, and patient motion correction were not applied to the examples presented here. Because the study focuses on 2D denoising, slices were extracted from the volumes to get 2D data.

C. Network and training

We trained a denoiser based on the DRUNet architecture [5], made of four blocks, with one input channel and one output channel, for an image denoising task.

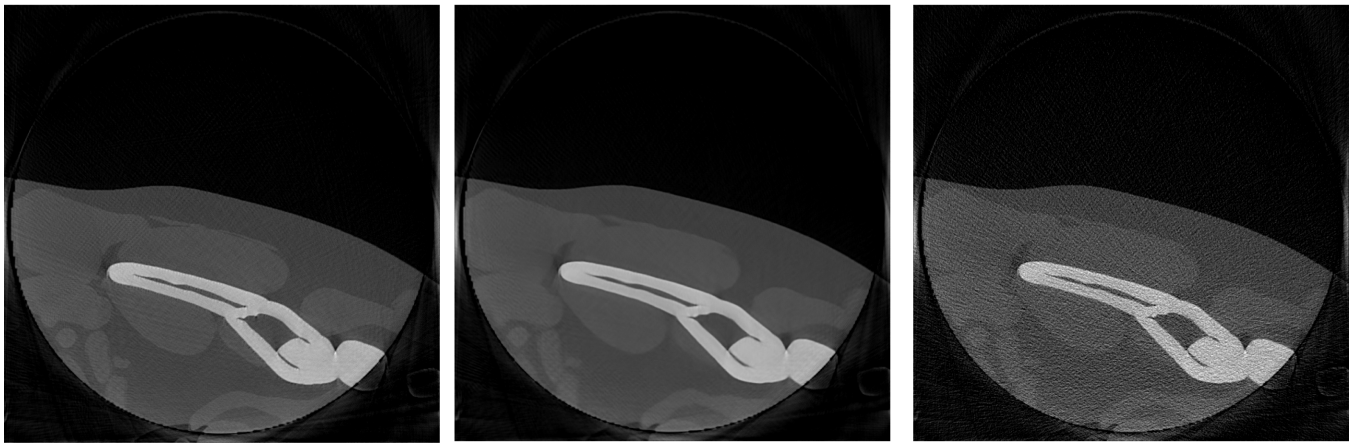


Fig. 2: Example of application of the denoiser on a denoising task for a simulated test-set image at the jaw level, in post-processing. Left: Reference image. Middle: Image denoised by the trained model (PSNR = 34.77 dB). Right: Noisy image (PSNR = 23.49 dB).

The dataset of 5715 images was divided into 4317 training samples ($\sim$ 76%), 600 validation samples ($\sim$ 10%), and 743 test samples ($\sim$14%), the jaw region of the XCAT phantom being reserved exclusively for the test set, as it is our region of interest.

To initialize the model weights, we used the pretrained weights from [8]. Additionally, we replaced the DRUNet's activation layer with a softplus function instead of ELU, to ensure continuous differentiability with respect to the input [14].

The model was fine-tuned for 63 epochs with a learning rate of 5×10^{-5} and a batch size of 8. The loss consists of the L1 distance between the network output and the ground-truth image:

$$\mathcal{L}(x, \hat{x}) = |x - \hat{x}|. \tag{8}$$

D. Gradient Step Plug-and-Play (GS-PnP) reconstruction algorithm

To reconstruct images from the simulated acquisitions, we solved problem (1) with an iterative reconstruction algorithm, the steepest gradient descent:

$$x_{k+1} = x_k - \tau\big(\nabla f(x_k) + \lambda \nabla g(x_k)\big), \tag{9}$$

where τ denotes the step size, and λ the regularization parameter. The regularization gradient ∇g is computed at each iteration following the denoiser defined in relation (5):

$$\nabla g(x_k) = x_k - D_\sigma(x_k). \tag{10}$$

The algorithm was initialized with $x_0 = \mathbf{0}$ and ran for 1500 iterations. The algorithm was implemented using DeepInverse [15]. Several values of the regularization parameter λ were tested to evaluate their effect on reconstruction quality.

III. Results

We first report the results of model training on the denoising task in Subsection III-A. Then, we present in Section III-B the reconstruction results obtained using the trained denoiser.

A. Model Testing

The trained denoiser was evaluated on a test set of 798 images obtained from the simulated acquisition of the XCAT phantom. The model was directly applied as a post-processing step to the reconstructed noisy images to assess its performance in this task. The model achieved a mean PSNR of 34.7 dB on the test set when comparing the denoised output with the non-noisy reconstruction over the entire image, whereas the noisy input achieved a PSNR of 22.95 dB.

Figure 2 presents a qualitative illustration in which the denoised image (Fig. 2, middle) appears noticeably cleaner while preserving the same contrast as in the clean reconstruction (Fig. 2, left). However, some fine structures visible in the ground truth, such as the small circular feature to the left of the jawbone, were obscured by noise in the input image (Fig. 2, right) and were not fully recovered by the model.

B. Performance on PnP Reconstruction

A quantitative evaluation was performed by testing the GS-PnP algorithm on a CT reconstruction task. The evaluation used 258 simulated acquisitions from the test set in the jaw region of the XCAT phantom. The ground truth corresponded to reconstructions without simulated noise, and the regularization parameter was set to $\lambda = 30$. We obtained a mean PSNR of 32.1 dB compared to 22.2 dB without the application of the prior obtained with the conjugate gradient algorithm.

Qualitative results were obtained by applying the GS-PnP algorithm to reconstruct real acquisitions extracted from CBCT slices. These results were compared with reconstructions obtained using (i) a Total Variation (TV) prior, and (ii)

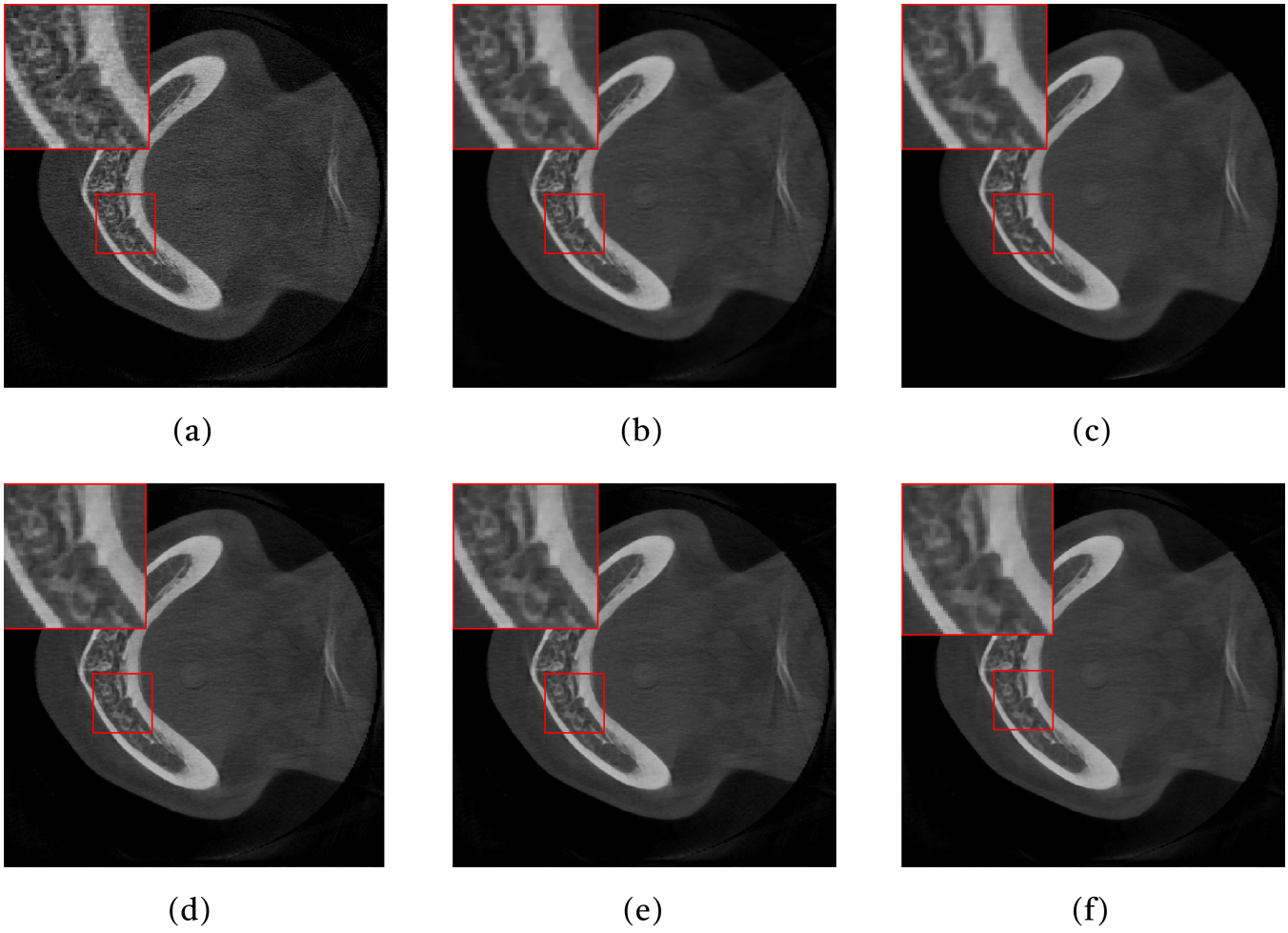


Fig. 3: Example of application of the Gradient-Step denoiser as a prior in a reconstruction task from a real acquisition, using 1500 iterations, compared to TV and a Denoiser from [5]. **(a)** Reconstruction without prior using gradient descent ($\lambda = 0$). **(b)** Reconstruction using a TV prior with PGD; over 20 iterations, with a regularization parameter of $\lambda = 0.05$. **(c)** Reconstruction using a DRUNet prior with α-PGD; the denoiser replacing the proximal operator used an averaging parameter $\alpha = 0.004$. **(d)** Reconstruction using the trained model as a prior with $\lambda = 15$. **(e)** Same as (d), but reconstructed with $\lambda = 20$. **(f)** Same as (d), but reconstructed with $\lambda = 30$.

the Gaussian denoiser from [5] plugged as a proximal operator in an α-Proximal Gradient Descent (α-PGD) scheme [16]. We used the Gaussian denoiser provided by the DeepInverse library [15], which has been trained on natural images.

Tests were performed using different values for the hyper-parameter λ controlling the regularization strength. Figure 3 illustrates that increasing the regularization parameter from $\lambda = 15$ to $\lambda = 30$ leads to a smoother result, though some fine details get attenuated by the model. We also observe than the results from our data-based model (Fig. 3.d) is sharper than reconstruction from the TV prior (Fig. 3.b) and the α-PGD Gaussian denoiser (Fig. 3.c). Nevertheless, small low-attenuation regions located between high-attenuation areas seem to suffer from reduced contrast, particularly for large values of λ.

IV. Conclusion

In this work, we trained a data-driven prior for CT reconstruction that effectively reduces photon noise in reconstructed images. The model is trained as a gradient-step denoiser, minimizing an explicit potential. Crucially, this potential is learned directly from the noise distribution encountered in CT images, rather than assuming Gaussian noise, making it better suited to the problem.

Challenges still persist in the reconstruction of some fine and clinically relevant structures. These limitations may be attributed to the use of a phantom with limited fine-detail features. More representative phantoms for training the denoiser could help improve reconstruction accuracy.

A major limitation of the currently presented approach is the slow convergence of the algorithm used, the steepest gradient descent being slow to converge. Future work will focus on implementing our approach using faster algorithms that are better suited to the reconstruction problem at hand.

ACKNOWLEDGMENT

This work was supported by the French National Association for Technical Research (ANRT) and the France 2030 BPI project Cone Beam AI and was performed within the framework of the SIRIC LYriCAN+ (INCa-DGOS-INSERM-ITMO cancer 18003) and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program 'Investissements d'Avenir' (ANR- 11-IDEX-0007) operated by the ANR. Simon Rit is a Royal Society Wolfson Visiting Fellow at the University College of London funded by the Wolfson Foundation and Royal Society.

REFERENCES

[1] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-play priors for model based reconstruction," in *2013 IEEE global conference on signal and information processing*. IEEE, 2013, pp. 945–948.

[2] U. S. Kamilov, C. A. Bouman, G. T. Buzzard, and B. Wohlberg, "Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications," *IEEE Signal Processing Magazine*, vol. 40, no. 1, pp. 85–97, 2023.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[4] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[5] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. Van Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Trans. on Pattern Analysis and Mach. Intel.*, vol. 44, no. 10, pp. 6360–6376, 2021.

[6] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang, and W. Yin, "Plug-and-play methods provably converge with properly trained denoisers," in *International Conference on Machine Learning*, 2019, pp. 5546–5557.

[7] J.-C. Pesquet, A. Repetti, M. Terris, and Y. Wiaux, "Learning maximally monotone operators for image recovery," *SIAM Journal on Imaging Sciences*, vol. 14, no. 3, pp. 1206–1237, 2021.

[8] S. Hurault, A. Leclaire, and N. Papadakis, "Gradient step denoiser for convergent plug-and-play," *CoRR*, vol. abs/2110.03220, 2021.

[9] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Medical Physics*, vol. 37, no. 9, pp. 4902–4915, 2010.

[10] S. Rit, S. Brousmiche, J. Finet, G. C. Sharp, and P. Steininger, "Reconstruction Toolkit (RTK) v2, an Insight Toolkit (ITK) module for tomographic reconstruction," in *International Conference on the use of Computers in Radiation therapy (ICCR)*, Lyon, France, 2024.

[11] K. T. Smith, D. C. Solmon, S. L. Wagner, and C. Hamaker, "Mathematical aspects of divergent beam radiography," *Proceedings of the National Academy of Sciences*, vol. 75, no. 5, pp. 2055–2058, 1978. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.75.5.2055

[12] H. Dang, J. W. Stayman, A. Sisniega, and W. Zbijewski et al., "Multi-resolution statistical image reconstruction for mitigation of truncation effects: application to cone-beam CT of the head," *Phys Med Biol*, vol. 62, no. 2, pp. 539–559, 2016.

[13] H. S. Park and K. Jeon, "An iterative reconstruction method for dental cone-beam computed tomography with a truncated field of view," *arXiv preprint arXiv:2508.07618*, 2025.

[14] S. Hurault, A. Leclaire, and N. Papadakis, "Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization," in *Int. Conference on Machine Learning*, vol. 162, 2022, pp. 9483–9505.

[15] J. Tachella, M. Terris, S. Hurault, and A. Wang et al., "Deepinverse: A python package for solving imaging inverse problems with deep learning," *J. of Open Source Software*, vol. 10, no. 115, p. 8923, 2025.

[16] S. Hurault, A. Chambolle, A. Leclaire, and N. Papadakis, "A relaxed proximal gradient descent algorithm for convergent plug-and-play with proximal denoiser," in *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2023, pp. 379–392.

Super-Resolution µCT of Complete Pathological Tarsal Bones

Sascha Senck, Lukas Nepelius, Johann Kastner, Jonathan Glinz, Klemens Trieb, Elena Kranioti, Bernhard Plank, Patrick Weinberger

***Abstract*—Microcomputed tomography (µCT) enables detailed visualization of trabecular bone microarchitecture but is fundamentally constrained by specimen size, leading to reduced spatial resolution for large anatomical structures such as complete hindfoot bones. This work focuses on the scalability and robustness of a deep-learning-based super-resolution (SR) pipeline when applied to substantially larger and clinically relevant foot bones. A dry pathological talus–calcaneus complex exhibiting talonavicular coalition, characterized by abnormal osseous fusion and altered load transfer, was used as a challenging target specimen. The SR model was trained on real paired µCT data from 20 dried os cuneiforme mediale specimens acquired at 20 µm and 80 µm voxel size and applied without retraining to the larger fused talus–calcaneus specimen scanned at 80 µm voxel size. A 3D U-Net architecture was employed, and SR inference was performed using a sliding-window strategy to reconstruct a virtual 20 µm dataset of the entire fused hindfoot. Qualitative assessment based on µCT slice images demonstrates improved delineation of trabecular structures across the full volume without obvious reconstruction artifacts. The results indicate that SR trained on small, high-quality µCT datasets can be transferred to large pathological bones, supporting future translation toward trabecular-aware clinical CT applications in orthopedics.**

***Index Terms*— Microcomputed tomography, foot bones, super-resolution, talonavicular coalition, trabecular bone**

I. Introduction

HIGH-resolution characterization of trabecular bone architecture is essential in pre-clinical imaging and orthopedic research, particularly for conditions that alter load distribution, joint congruency, and bone remodeling [1]. Laboratory µCT provides high spatial resolution for trabecular bone analysis but is constrained by object size, limiting its applicability to complete bones such as the talus and calcaneus. For these structures, voxel sizes using 2k flat-panel detectors are typically in the range of 80-100 µm, resulting in partial volume effects, blurred edges, and loss of microarchitectural detail such as thin trabeculae. In comparison, standard medical CT systems offer a slice thickness of approximately 400 µm, e.g. Siemens Force, recent systems featuring single photon counting detectors offer slice thicknesses down to 200 µm, e.g. in Standard UHR mode for a field of view of approximately 20 cm [2].

Talonavicular coalition is a congenital or acquired condition characterized by abnormal fibrous, cartilaginous, or osseous fusion between the talus and adjacent tarsal bones. Clinically, it is associated with restricted joint motion, altered hindfoot biomechanics, pain, and early degenerative changes. From an orthopedic perspective, understanding the three-dimensional trabecular microarchitecture in such fused regions is of relevance, as it reflects long-term load adaptation, stress redistribution, and pathological bone remodeling [3]. However, current clinical imaging modalities lack the spatial resolution required to visualize these microstructural changes in detail.

Super-resolution (SR) methods based on convolutional neural networks offer a computational pathway to enhance image resolution beyond hardware-imposed limits [4, 5]. Whether SR models trained on small bones can be meaningfully applied to large, anatomically complex, and pathological foot bones is part of current research. This study addresses this question by applying an established SR pipeline to a fused talus-calcaneus specimen. The work is a feasibility study, emphasizing the benefits and limits of CNN-based SR and its translational relevance in orthopedic science.

II. Materials and Methods

A. Specimens and training data

Super-resolution training was performed using µCT data from 20 dried os cuneiforme mediale (C2) specimens. These bones were selected because their small dimensions allow complete high-resolution µCT acquisition at 20 µm isotropic voxel size. For each C2 specimen, a corresponding low-resolution scan at 80 µm voxel size was acquired, yielding real, paired LowRes - HiRes datasets. Fig.1 shows a medical CT slice image in sagittal view indicating the C2, talus, and calcaneus bones.

The inference target was a dry pathological hindfoot specimen consisting of a fused talus and calcaneus consistent with talonavicular coalition. The specimen represents a clinically relevant but challenging case due to its larger

This work was supported by the projects “sustaiNDT” and “HyperMAT”.

Sascha Senck is with the University of Applied Sciences Upper Austria, Wels, Austria and with Paracelsus Medical University, Salzburg, Austria (e-mail: sascha.senck@fh-wels.at).

Lukas Nepelius, Jonathan Glinz, Bernhard Plank, and Johann Kastner are with the University of Applied Sciences Upper Austria, Wels, Austria.

Patrick Weinberger is with the University of Applied Sciences Upper Austria, Campus Wels, Wels, Austria, and also with Johannes Kepler University, Linz, Austria.

Klemens Trieb is with Paracelsus Medical University, Salzburg, Austria, and with Danube Private University, Krems, Austria.

Elena Kranioti is with the University of Crete, Iraklio, Greece.

volume compared to the smaller bones used in the training data, heterogeneous trabecular architecture, and long-term pathological remodeling.

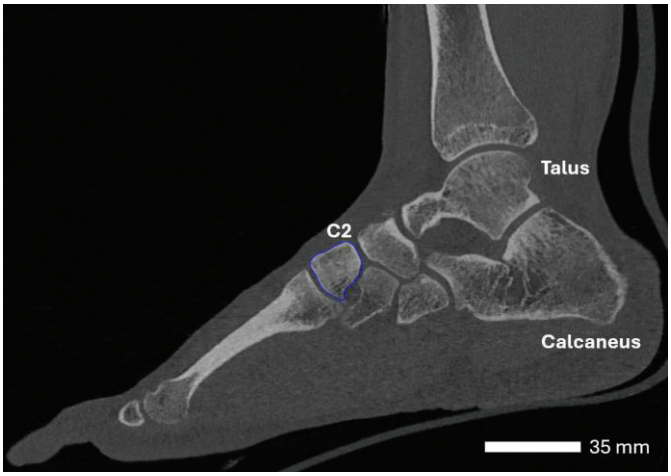


Fig. 1. Medical CT slice image of a healthy foot indicating the bones used for training (C2) and inference (talus and calcaneus).

Fig. 2 shows the target specimens characterized by an extensive coalition of talus and calcaneus and an extensive bone resorption area at the proximal part of the talus. All dry bones were provided by the Institute of Forensic Medicine at the University of Crete in Heraklion (Greece) and scanned using X-ray microcomputed tomography (µCT).



Fig. 2. Dry pathological talus–calcaneus complex exhibiting talonavicular coalition and an avascular necrosis of the talus.

The fused talus-calcaneus specimen was scanned at 80 µm voxel size to ensure complete coverage within a single µCT scan. The resulting dataset comprised approximately 667× 1168×1235 voxels. Training data and the target specimens reference data were acquired using a laboratory µCT system (Procon Alpha 240). Voxel size for the target specimens was 80 µm with a volume dimension of 667×1168×1235 voxels, stored at 16-bit grayscale depth. The µCT scans were performed at a tube voltage of 220 kV and a tube current of 400 µA, providing sufficient X-ray penetration for the dense, dry talus–calcaneus specimen. A physical copper filter with 0.6 mm thickness was used to reduce beam hardening artefacts. A calibration phantom was used to correct the voxel size. Mitos XAid was used for CT data reconstruction.

B. Super-resolution model and inference strategy

A 3D U-Net architecture was employed for super-resolution reconstruction. The used model was trained for 15 epochs using exclusively real paired µCT datasets of C2 bones acquired at 20 µm and 80 µm voxel size. Sixteen C2 specimens were used for training and four for testing as described in [4].

Due to the dimensions of the target volume, SR inference was performed using a sliding-window inference approach [6], which is an efficient approach for applying SR to large three-dimensional µCT datasets. In this approach, the low-resolution input volume is subdivided into smaller three-dimensional sub-volumes, which are sequentially processed by the pretrained neural network on the GPU. Each sub-volume serves as an independent input patch for the SR model. The reconstructed high-resolution sub-volumes are then transferred back to system memory, where they are reassembled into the full-resolution volume. This separation of GPU-based inference and CPU-based volume assembly is essential, as GPU memory is typically the primary bottleneck, whereas substantially larger memory resources are available in system RAM.

The sliding-window strategy further allows the use of overlapping sub-volumes, typically on the order of 25-50%. Overlap between adjacent sub-volumes improves reconstruction continuity and significantly reduces grid or stitching artifacts that may otherwise arise at patch boundaries. During reassembly, overlapping regions are blended, resulting in smoother transitions and a more homogeneous final volume. This concept follows established sliding-window inference strategies for volumetric medical imaging, as implemented in frameworks such as MONAI [6].

The importance of this approach is evident when considering the data scaling inherent to SR. With a super-resolution factor of 4, the voxel count increases by a factor of 4 in each spatial direction (X, Y, and Z), corresponding to a 64-fold increase in total data volume. Consequently, an original µCT dataset of approximately 2 GB at 80 µm voxel size expands to more than 120 GB at a upscaled voxel size of 20 µm. While such scaling is relatively manageable in two-dimensional image processing where only two spatial dimensions are involved and file sizes typically remain in the megabyte range, this is a substantial computational and memory challenge for full three-dimensional µCT volumes.

To accommodate this data expansion, inference was performed on a high-memory workstation equipped with 1024 GB system RAM and two NVIDIA RTX A5000 GPUs with 24 GB of GPU memory respectively.



Fig. 3. µCT slice images of C2 at voxel sizes of 80 µm (LowRes) and 20 µm (HiRes: middle; SR: bottom).

Sub-volume size, overlap, and batch size were empirically optimized to balance GPU memory usage, inference throughput, and reconstruction quality. After optimization, the complete SR inference of the fused talus–calcaneus dataset (final volume size 2668×4672×4940 voxels with 123 GB) required approximately 90 minutes of computation time. Evaluation for this feasibility study was strictly qualitative since HiRes scans of the target bones at 20 µm were not possible.

III. RESULTS

Fig. 3 illustrates the super-resolution (SR) reconstruction of complete C2 bones using the model trained on paired C2 µCT data as described in [1]. The comparison between HiRes, and SR reconstructions demonstrates that the network is able to recover fine trabecular structures and internal connectivity that are strongly affected by partial volume effects in the LowRes input.

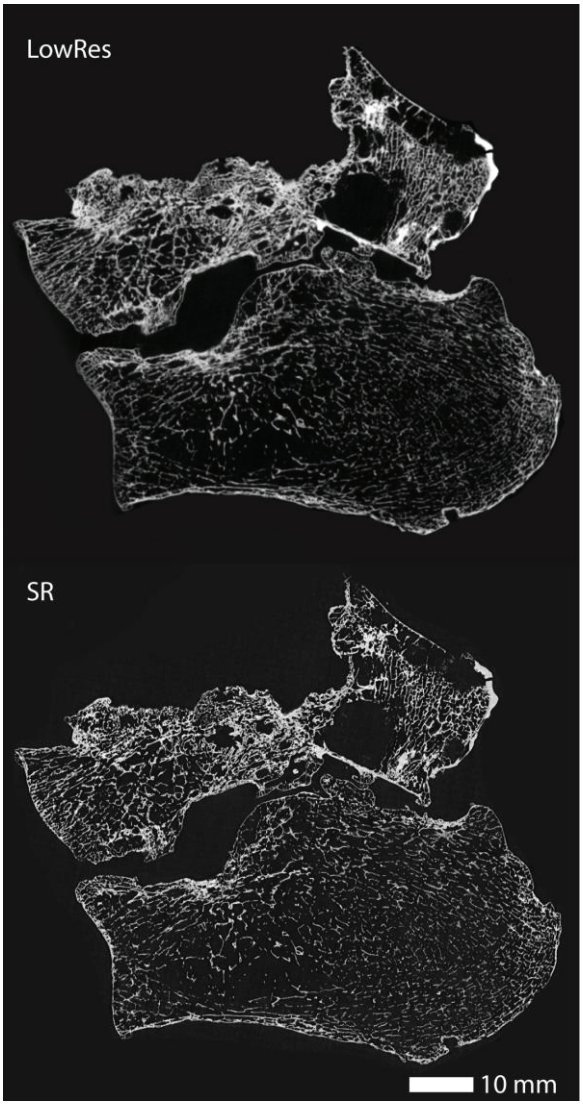


Fig. 4. µCT slice images of the fused talus and calcaneus bones at voxel sizes of 80 µm (LowRes: top) and 20 µm (SR: bottom).

Fig. 4 shows the corresponding SR prediction applied to the fused talus-calcaneus target specimen, representing a substantially larger and anatomically more complex bone volume than used during training. The SR reconstruction of the fused talus–calcaneus specimen resulted in a spatially continuous high-resolution volume without visible stitching

artifacts at sub-volume boundaries, despite the use of sliding-window inference. Compared to the original 80 µm dataset, SR images exhibit a markedly improved visualization of trabecular struts and predominant trabecular orientation patterns across the entire specimen. Fig. 5 provides a detailed view of the fused region characterized by pathological bone changes, where SR reconstruction enhances the visibility of trabecular organization in the coalition zone. No obvious SR-related artifacts, such as checkerboard or grid patterns, are observed. Both trabecular and cortical bone structures appear anatomically plausible and are reconstructed without apparent geometric distortions.

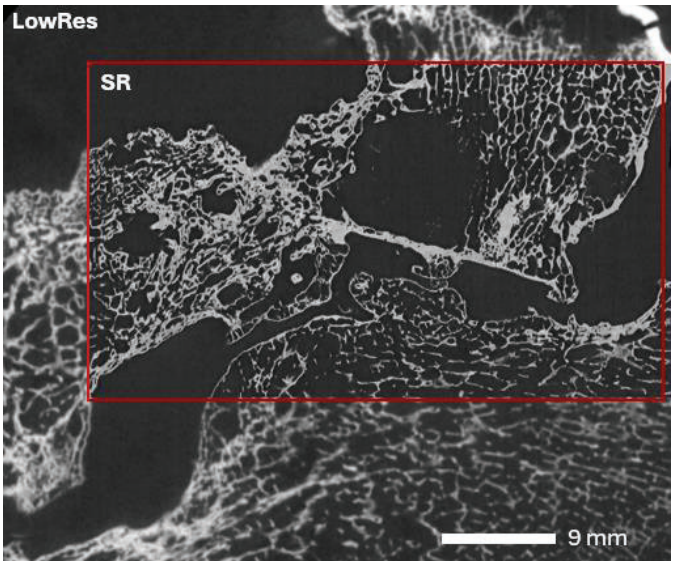


Fig. 5. Detailed view of the fused talus–calcaneus region exhibiting pathological bone changes. The overlay highlights the comparison between the original µCT data (LowRes, 80 µm voxel size) and the corresponding SR reconstruction (in red box: voxel size 20 µm).

IV. DISCUSSION

This study demonstrates that a SR pipeline trained on small tarsal bones using real, paired µCT data can be transferred to substantially larger, pathological foot bones without retraining or fine-tuning. The fused talus-calcaneus specimen represents a demanding test case due to its larger volume, complex three-dimensional geometry, and pathological bone alterations associated with talonavicular coalition. The qualitative results indicate that the trained model generalizes beyond the anatomical scale and morphology represented in the training data. From an orthopedic perspective, the improved visualization of trabecular microarchitecture in pathological conditions is of clinical relevance since an improved depiction of trabecular organization in fused and load-altered regions may support a better understanding of long-term mechanical adaptation, stress redistribution, and bone remodeling. Such information has the potential to complement clinical assessment by providing additional structural context that is not accessible with standard-resolution CT imaging. While the present study is strictly qualitative and does not aim at quantitative morphometric validation, it illustrates how SR-enhanced imaging can bridge the gap between conventional CT and high-resolution µCT for larger skeletal structures.

The presented approach does not introduce a new machine-learning model or task-specific optimization. Instead, it demonstrates the robustness and scalability of an existing, basic SR framework when applied to a challenging and clinically relevant specimen. This aspect is important for translational imaging workflows, where methodological stability, reproducibility, and predictable behavior of trained models are essential prerequisites for broader applicability, e.g. for SR of medical CT data.

V. CONCLUSION

The presented results indicate that super-resolution µCT trained on small, high-quality tarsal bone datasets can be successfully applied to large, pathological foot bones such as a fused talus–calcaneus specimen. Despite the substantial increase in data volume and anatomical complexity, the SR pipeline produces anatomically plausible reconstructions with improved visualization of trabecular structures and reduced partial volume effects compared to the original low-resolution data. The study further highlights that applying SR to large three-dimensional µCT datasets requires dedicated methodological considerations, such as sliding-window inference, which represent a key technical prerequisite for scaling super-resolution methods to larger volumes.

VI. OUTLOOK

Future work will focus on extending the present qualitative feasibility study toward a quantitatively validated framework for SR µCT of larger skeletal structures. A next step is the acquisition of high-resolution reference data for larger hindfoot bones, which is currently limited by geometric constraints of laboratory µCT systems. To address this, region-of-interest (ROI) tomography will be employed to obtain localized high-resolution scans (20 µm voxel size) within selected subregions of the talus and calcaneus. However, this approach prevents the use of fiducial markers, as these cannot be placed inside the bone without destructive preparation, thereby limiting precise spatial registration between LowRes and HiRes (training) datasets. Therefore, Senck et al. [4], utilized smaller os cuneiforme intermediate bones that can be embedded in dedicated sample holders with fiducial markers, enabling accurate paired acquisitions without altering the biological hard tissue.

Based on these ROI datasets, quantitative morphometric analyses will be performed, including trabecular thickness (Tb.Th.) and bone volume fraction (BV/TV). In addition, nominal-actual comparisons will be used to assess the degree of information loss in LowRes data and the extent to which SR reconstruction restores fine relevant microstructural features like thin trabeculae. This will allow a systematic evaluation of SR performance beyond visual inspection, including error characterization and bias analysis in relation to partial volume effects. Furthermore, we will expand the dataset to include additional non-pathological and

pathological foot bone samples, e.g. from patients with diabetic neuropathic osteoarthropathy (DNOAP, Charcot foot), which is characterized by pathological fractures, progressive bone destruction, and pronounced microstructural alterations such as increased cortical porosity, bone resorption, and sclerosis [7]. To further improve the robustness and balance of the training data, additional anatomical regions such as femoral heads will be included, thereby increasing structural variability and enabling a more comprehensive assessment of the generalizability of the trained SR model.

ACKNOWLEDGMENT

The work was supported by the project “sustaiNDT” [grant nr. 909801] via the Austrian Research Promotion Agency (FFG) funded within the framework "Production & Materials" of Austrian Research Promotion Agency (FFG) and by the project “HyperMAT” (IBW/EFRE & JTF 2021-2027), funded by the European Just Transition Fund and the Upper Austrian Government.

REFERENCES

[1] F. Wichlas et al., “Significantly increased bone volume in a critical-sized defect model in the rat animal model by transplantation of a stand-alone vascularized periosteal flap,”Eur J Trauma Emerg Surg., vol. 28, no. 51(1), pp. 121, Feb. 2025; DOI: 10.1007/s00068-025-02770-5.

[2] Joël Greffieret al. “Image quality and dose reduction with photon counting detector CT: Comparison between ultra-high resolution mode and standard mode using a phantom study,” Diagnostic and Interventional Imaging, vol. 106, no. 9, pp. 320-326, Sep. 2025. DOI: 10.1016/j.diii.2025.03.009.

[3] S. A. Brennan, et al. “Talonavicular synostosis with lateral ankle instability - A case report and review of the literature,” Foot and Ankle Surgery, vol. 18, no. 3, pp. e34-e36, Sept. 2012, DOI: 10.1016/j.fas.2012.01.002

[4] S. Senck et al., “Optimizing µCT resolution in tarsal bones: A comparative study of super-resolution models for trabecular bone analysis,” Tomogr. Mater. Struct., vol. 8, p. 100063, 2025, DOI: 10.1016/j.tmater.2025.100063.

[5] T. J. Chan and C. S. Rajapakse, “A Super-Resolution Diffusion Model for Recovering Bone Microstructure from CT Images,” Radiol. Artif. Intell., vol. 5, no. 6, p. e220251, Sep. 2023, DOI: 10.1148/ryai.220251.

[6] M. J. Cardoso et al., “MONAI: An open-source framework for deep learning in healthcare,” 2022. DOI: 10.48550/ARXIV.2211.02701.

[7] S. Senck et al., “Visualization of local cortical defects in Charcot foot using microcomputed tomography”. Orthopäde 44, pp. 8–13, 2015, https://doi.org/10.1007/s00132-014-3053-0.

Clinical Feasibility of Fully Autonomous Cardiac CT Using Pulsed Mode Projections

P. Wu,[1] E. Haneda,[1] I. Heukensfeldt Jansen,[1] S. Araujo,[1] A. Hsiao,[2] E. McVeigh,[3] and B. De Man[1]

[1]GE HealthCare, One Research Circle, Niskayuna, NY, USA 12309
[2]Dept. of Radiology, Medicine, Radiology at University of California San Diego, CA, USA
[3]Dept. of Bioengineering, Medicine, Radiology at University of California San Diego, CA, USA

***Abstract*—** Cardiac CT (CCT) is of vital importance in heart disease diagnosis but is conventionally limited by its complex workflow that requires dedicated phase and bolus monitoring devices. To address this issue, an autonomous cardiac CT workflow was presented in which the acquisition is prospectively timed directly from minimal-dose pulsed-mode projections (PMPs), removing dependence on ECG leads and separate bolus-tracking scans. In this work, a new dual-output multi-channel CNN–LSTM network was proposed to jointly estimate cardiac phase (RR%) and iodine contrast level from the same PMP stream. Clinical feasibility of this workflow was evaluated for the first time in a prospective human subject study, where the network estimated phase and iodine contrast were compared with ones obtained from the conventional workflow (i.e., ECG and timing bolus).

On a simulated test set, the dual-output formulation achieved accuracy comparable to or slightly exceeding that of two single-task networks while substantially reducing parameters and inference time, indicating that the dominant projection-domain features for phase and contrast are widely shareable. In real human subject data, phase estimation exhibited minor domain shift (compared to simulated data), especially for human subjects under breath-hold. On the other hand, contrast estimation showed a larger difference relative to performance on simulated dataset. Future work will expand the virtual data generator to explicitly model respiratory motion to narrow this domain gap and improve clinical robustness.

***Index Terms*—** Computed Tomography, Cardiac Phase Estimation, Iodine Contrast Estimation, Deep Learning

I. Introduction

Cardiovascular disease is the leading cause of morbidity and mortality in the US.[1] As one of the most widely used noninvasive heart imaging techniques, Cardiac CT (CCT) delivers high-resolution anatomy and functional insights, but its clinical robustness hinges on precise scan timing: selecting a motion-quiet cardiac phase and an optimal iodine contrast level. In current practice, this timing relies on ECG gating and contrast-tracking protocols (e.g., bolus tracking or test bolus), which add setup time, increased human and financial cost, and potential failure modes (e.g., poor ECG contact, electromagnetic interference), and can reduce patient comfort.

We recently introduced an autonomous CCT paradigm in which a minimal-dose pulsed-mode projection (PMP) stream is analyzed in real time to drive scan timing without external devices.[2] In this new autonomous exam paradigm, the scan starts with a minimal-dose monitoring stage where the PMPs are continuously acquired and analyzed in real-time by deep learning (DL) techniques to determine the required information over time, namely cardiac phase and iodine contrast.

One drawback of our previous studies is that they rely purely on realistic simulations/emulations (adding contrast to 4D CCT data) to establish technical feasibility. To address this issue, here we report the first human-subject validation of this concept in a clinical setting. In this study, we evaluate whether PMP-based timing can replicate clinical decisions currently made with ECG (for phase) and test-bolus methods (for contrast), using each as the reference standard.

Additionally, in our prior study, phase and contrast were estimated by two independent networks operating on the same PMP stream. This duplicated preprocessing and temporal modeling, increased latency/compute, and importantly, neglected to model any coupling and did not exploit the natural correlation between myocardial motion and contrast kinetics. To overcome these drawbacks, in this work, we propose a new single dual-output temporal network that produces jointly estimated phase (RR%) and contrast estimates in one forward pass. Our hypothesis is that shared spatiotemporal representations in latent space allow the model to exploit the physiological coupling between cardiac motion and contrast kinetics, improving stability while reducing complexity versus running two separate models. In other word, there is interaction between (i) cardiac phase, (ii) respiratory phase, (iii) contrast phase, and (iv) PMP, so they need to be incorporated into a concurrent solution.

In this work, we report the development of a projection domain multiple output DL-based method that simultaneously estimates phase and iodine contrast for autonomous CT. The performance of the proposed method was evaluated with physics-based emulated datasets (with

perfect ground truth) and human-subject data (with surrogate ground truth). Collectively, our goal is to advance autonomous CCT from simulation to clinical feasibility.

II. Methods

A. Autonomous Cardiac CT Protocol

Figure 1 shows the flowchart of the proposed autonomous CCT protocol. Following intravenous contrast initiation, the scanner enters a low-dose monitoring stage in which pulsed-mode projections (PMPs) are acquired with the collimator already opened to the diagnostic scan range. To preserve temporal fidelity while minimizing dose, we acquire 1 to 2 views per rotation (rotation time is 0.28 s) at fixed angles identified in prior optimization studies as informative for cardiac motion and bolus dynamics.[2] The resulting PMP stream is analyzed online by a single dual-output temporal network that produces, at each update, a joint estimate of cardiac phase (RR%) and contrast level. While contrast level can be estimated for different cardiac chambers, we solely use the contrast level from the left ventricle as example in this work. Phase estimates are temporally regularized with a dynamic programming step [3] to enforce periodic consistency over short windows (i.e., heart rate should be relatively stable over short period of time), whereas contrast estimates are accumulated into a running bolus curve.

An autonomous trigger is asserted when both timing conditions are satisfied, i.e., RR% is within a user-specified window (likely around end-diastole) and contrast is at or near the predicted peak for the target chamber. At this point the system switches immediately from pulsed monitoring to the diagnostic acquisition. The monitoring duration (~10–20 s) adds negligible dose (as only 1 or 2 low-dose PMP per rotation) while providing sufficient temporal context for robust dual-criterion triggering.

B. Cardiac Phase and Contrast Estimation

We employ a time-distributed multi-channel CNN–LSTM architecture that produces joint estimates of cardiac phase and contrast from the same pulsed-mode projection (PMP) stream. This is an extension of our previous PhaseNet design, as shown in Fig. 2. Each time point is represented by a multi-channel PMP group that is first processed by a shared ResNet feature extractor; the resulting feature sequence is then modeled by a bidirectional LSTM to capture temporal dependencies before two task-specific heads generate outputs. The final stage consists of two separate task-specific heads (lightweight FC–activation–FC stack), for cardiac phase (RR%) and iodine contrast.

Our hypothesis is that since cardiac phase and contrast are driven by similar underlying physiology and anatomy, learning them jointly lets a shared CNN–LSTM trunk capture motion-sensitive features, heart-region localization, and short-range temporal consistency that benefit both tasks. Moreover, contrast regression (arguably the simpler task) could potentially provide a smoother, more stable training signal that regularizes early spatiotemporal features, from which the sharper, locally ambiguous phase estimation benefits (i.e., auxiliary-task regularization [4]).

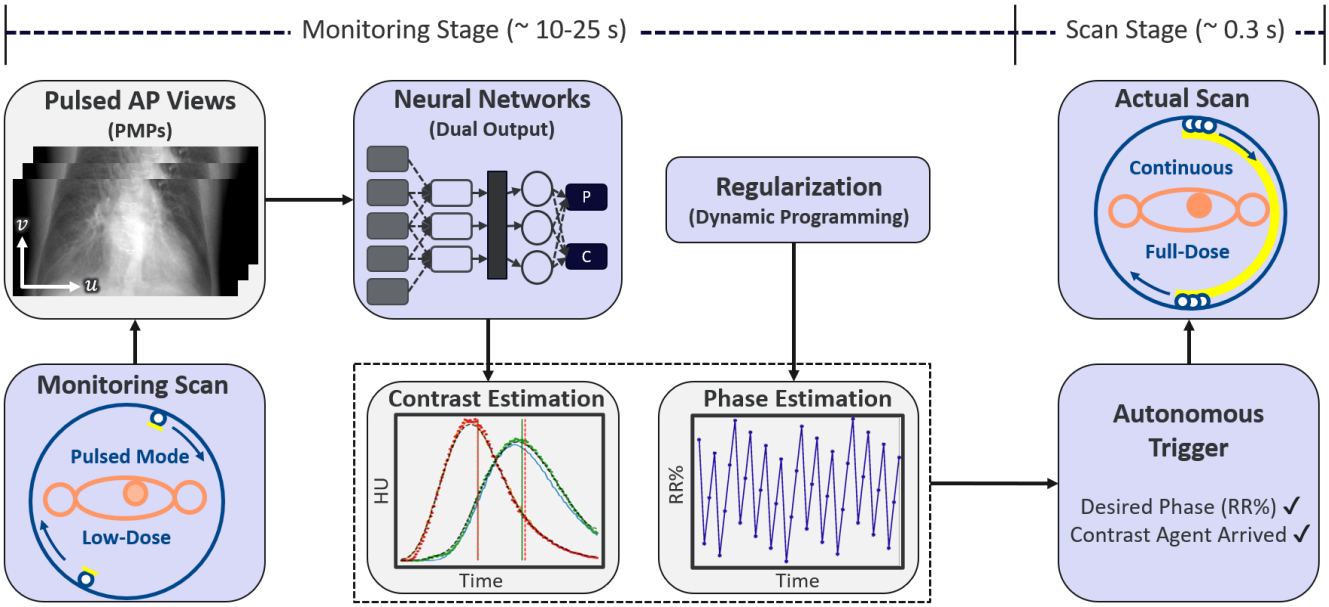


Figure 1. Flowchart of the autonomous cardiac CT protocol through projection domain cardiac phase and contrast estimation.

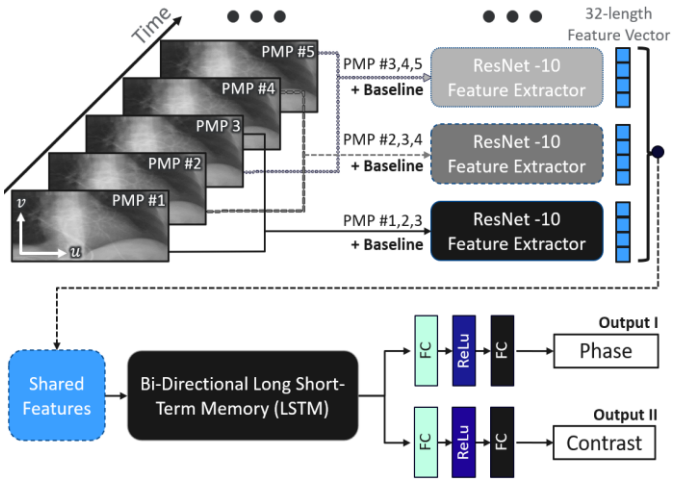


Figure 2. Architecture of the proposed dual-output network.

The PMP groups are generated in a sliding window fashion to capture minor cardiac motions between adjacent PMPs, which are crucial for phase estimation. Importantly, to enable reliable contrast analysis, every PMP group is prepended with a fixed baseline projection, i.e., a pre-bolus PMP acquired at the beginning of monitoring. This design ensures that the feature extractor receives both the current observation and a stable, subject-specific reference for contrast estimation, while preserving the discriminative cues needed for phase estimation.

As in our prior work, the model predicts the state of the last PMP in the window, enabling prospective, per-view updates with sub-rotation latency; the LSTM component remains essential for stabilizing estimates under heart-rate variability and view sparsity.

C. Clinical Validation

A human-subject study was conducted at the University of California San Diego (UCSD) under IRB approval. Twenty human subjects were scanned using a clinical CT system (GE HealthCare Revolution CT™). Subjects were acquired with three respiratory conditions during the PMP monitoring stage: free breathing (n=10), breath-hold (n=8), and shallow breathing (n=2). For all participants, we collected two subsequent monitoring data: (i) standard test bolus timing data to establish clinical references for contrast timing, and (ii) PMP data for autonomous analysis. A multi-lead ECG/EKG was recorded throughout every scan to provide ground-truth cardiac phase (RR%) for evaluation of phase estimation.

As illustrated in Fig. 3, we performed a test-bolus acquisition using continuous gantry rotation and a narrow collimation (i.e., limited z-coverage) to minimize dose yet preserve temporal sampling (i.e., GE HealthCare SmartPrep). The test bolus consisted of iodinated contrast mixed with saline, adjusted so that the total injection volume matched the planned diagnostic bolus. For the research pulsed acquisition, the x-ray source operated in pulsed mode with two fixed angle locations per rotation, producing a sparse projection time series suitable for prospective analysis by the dual-output network.

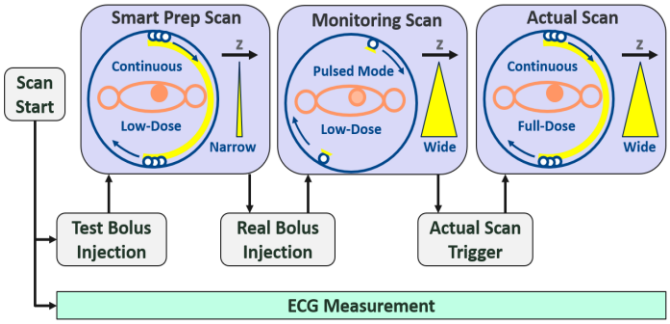


Figure 3. Acquisition protocol used in the clinical study (to obtain test data as well as standard ground truth).

Because our training data generation assumed no respiratory motion, the primary evaluation in this report is restricted to the 10 non-free-breathing cases (8 breath-hold + 2 shallow breathing).

D. Dataset and Experiment Settings

To generate training and testing data, we used a virtual imaging framework [5] to simulate realistic pulsed-mode projections (PMPs) with variability in anatomy, heart rate, cardiac phase, and contrast dynamics. Briefly, the framework forward-projects 4D cardiac CT models constructed from an ensemble of multi-phase clinical cardiac CT scans, enabling arbitrary phase/contrast profiles. In this study, 115 subject models were instantiated to create 300 simulated "monitoring" exams. Each exam comprised 35 s of PMP acquisition with two PMPs per gantry rotation at fixed angles 150°/330°, yielding 75,000 PMPs in total. Because the 4D cardiac CT does not include respiratory motion, these data represent a breath-controlled scenario. The simulated cases were split 80/10/10 % into training/validation/testing sets with no subject overlapping across splits.

All networks were trained using Adam (initial learning rate 3×10^{-5}, batch size 32). For cardiac phase, periodicity was handled via a wrapped error and performance was reported using RR% RMSE (computed after wrap-around alignment to the ground truth), Pearson correlation coefficient, and RMSE of the first order difference of RR% ($RMSE_d$). For contrast, the results were evaluated using peak detection error (difference between predicted and reference peak-time).

III. Results

A. Evaluation on emulated Data

As shown in Table 1, on the emulated PMP test set, the dual-output multi-channel CNN–LSTM achieved comparable or slightly better accuracy than a pipeline using two separate single-output networks (one for phase, one for contrast), when trained with identical data and hyper-parameters. At the same time, the dual-output design substantially reduced model size and per-update latency because a single shared encoder–LSTM is executed once per PMP update rather than twice. These findings indicate that the spatiotemporal features extracted from sparse projections are largely shareable across the two tasks (phase and contrast), and that a unified estimator streamlines inference without sacrificing accuracy. Consistent with prior observations on time-series modeling, the LSTM component remained critical for stabilizing both outputs under heart-rate variability and view sparsity.

Networks	RMSE (%)	$RMSE_d$	Pearson Coeff
Single Output	7.953 ± 0.762	5.323 ± 0.564	0.972 ± 0.003
Dual Output	7.425 ± 0.572	5.325 ± 0.510	0.973 ± 0.003

Networks	Peak Diff (s)	# Parameters
Single Output	0.179 ± 0.723	14.6 M
Dual Output	0.166 ± 0.661	7.4 M

Table 1. Contrast and phase estimation performance for single and dual output networks.

B. Human-subject Evaluation

We next evaluated the dual output model on human-subject data using the non-free-breathing cohort. For cardiac phase, the domain shift from simulated to clinical data was shown to be small, with an average RR% RMSE of only 6.98% (compared to ~7.42% in simulated data). Overall agreement with ECG-derived RR% remained high as shown in Fig. 4, with notably larger errors in shallow-breathing cases than in breath-hold cases, and one outlier currently under review involving a very low heart rate compared with the rest of the cases.

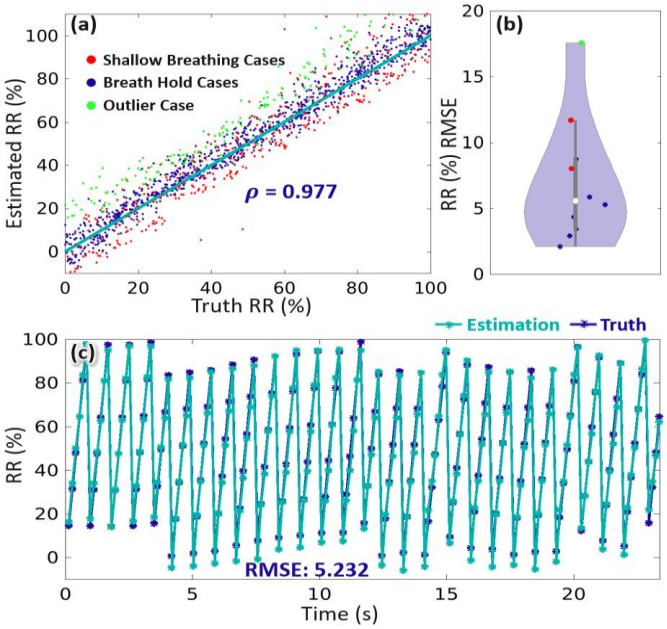


Figure 4. DL-based phase estimation performance on real data. (a) Estimated vs. truth RR% for all PMPs. (b) Averaged RR% RMSE for each subject. In (a) and (b), green, red, and blue dots are for the outlier case, shallow breathing cases, and breath holding cases respectively. (c) Cardiac phase estimation overtime in an example testing case.

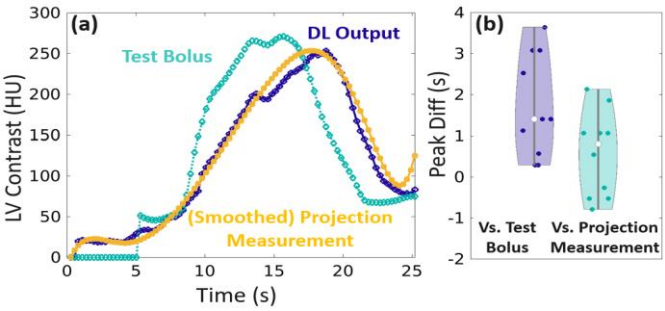


Figure 5. DL-based contrast estimation performance on real data. (a) Estimated left ventricle (LV) contrast vs. test bolus result ("ground truth) and smoothed projection measurement (surrogate ground truth). (b) Peak timing differences of DL estimated contrast curve, vs. test bolus and vs. projection measurement.

However, contrast timing exhibited a larger degradation relative to simulation, most evident in peak-time detection error. Several physiological factors may contribute: (i) different cardiac output between test bolus and PMP monitoring scans; (ii) a higher instantaneous iodine concentration and increased viscosity during the diagnostic injection (i.e., PMP monitoring scans) may modify transit times. This results in our network-predicted contrast curve having peaks almost always later than those in the test bolus (as shown in Fig. 5).

Supporting the network estimates, a direct projection-domain measurement (fixed cardiac ROI intensity

vs. time, without learning) showed good agreement with the model-predicted contrast curve as shown in Fig. 5(a), suggesting that at least some of the discrepancy arises from reference-curve differences between test-bolus/SmartPrep and the PMP monitoring context rather than from network failure (i.e., imperfect ground truth). Average absolute peak timing difference is ~1.73s when compared with the test bolus data, and ~0.98s when compared with direct projection measurements.

IV. Discussion and conclusions

We demonstrated an autonomous, projection-driven cardiac CT workflow in which a dual-output multi-channel CNN–LSTM provides joint phase and contrast estimates from pulsed-mode projections. This workflow is evaluated for the first time on human-subject data. On the simulated set, the dual-output model achieved comparable or slightly better accuracy than two separate single-task networks while reducing parameters and inference time, indicating that spatiotemporal features for phase and contrast are largely shareable. In human-subject testing, phase translated well with minor domain shift (best performance under breath-hold), whereas contrast peak timing degraded slightly relative to the simulation study. This could be due to domain shift or could partially arises from reference-curve differences between test-bolus and the PMP monitoring context (i.e., imperfect ground truth), which we will keep investigating in future work. If there is indeed a difference between test bolus and ground truth, the value of PMP data is increased because it is obtained immediately before the imaging data.

To address the domain gap related to breathing motion, we will expand the virtual data generator to explicitly model respiratory motion in future work, by simulating PMP sequences with parameterized breathing (amplitude, frequency), chest-wall and diaphragm-driven cranio-caudal motion; and using a warped forward projector accounting for said breathing motion. Additionally, future work will focus on improving model robustness on clinical data, followed by prospective development and testing of an autonomous triggering algorithm that fuses phase and contrast outputs (and their uncertainties) into a safety-aware decision rule.

Acknowledgements

Research reported in this publication was supported by the NIH/NHLBI grant R01HL153250. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

V. References

[1] J.-E. Scholtz and B. Ghoshhajra, *Cardiovasc Diagn Ther*, vol. 7, no. 5, pp. 439–451, Oct. 2017, doi: 10.21037/cdt.2017.06.07.

[2] P. Wu *et al.*, *IEEE Transactions on Medical Imaging*, 2025, Vol. 44, no. 6.

[3] A. Viterbi, *IEEE transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.

[4] S. Shin, H. Do, and Y. Son, *ECCV 2024*, vol. 15138.

[5] E. Haneda *et al.*, in *The 7th International Conference on Image Formation in X-Ray Computed Tomography*, 2022.

Mixed Prior Diffusion Posterior Sampling for High Resolution CT Reconstruction

Peiqing Teng, Xin Wang, Jian Zhou, Liang Cai, Tzu-Cheng Lee, Ruoqiao Zhang, J. Webster Stayman

***Abstract*—Diffusion models have a demonstrated ability to estimate complex prior distributions, including classes of image data. Diffusion Posterior Sampling (DPS) has recently emerged as a powerful framework for incorporating such learned priors into inverse problems; however, the performance of DPS is limited by the ability to train high-quality priors. This is particularly true in x-ray computed tomography, where patient data must balance noise and spatial resolution within acceptable x-ray dose limits. In this work, we propose Mixed Prior Diffusion Posterior Sampling (Mixed Prior DPS) that integrates global and local diffusion priors in the reconstruction. The global diffusion model is trained on full CT volume with modest spatial resolution to capture large-scale anatomical structures and lower frequency content. A local diffusion model trained on high-resolution patches focuses on fine-scale, high-frequency details. The two priors are combined in the Fourier domain to form a mixed posterior mean, and a masking strategy is employed to apply prior mixing selectively within an appropriate local region-of-interest while employing standard-resolution DPS elsewhere. This framework permits the use of modest resolution CT datasets for global context in conjunction with specialized high-resolution datasets (e.g. organ or tissue specific scans based on specialized protocols or on *ex vivo* samples where high-resolution, low-noise images can be collected without radiation dose constraints). In this work, we created a high-resolution lung training dataset for the local prior model using a CBCT benchtop system and a (plastinated) physical lung specimen. Experimental results on simulated low-dose CT demonstrate that the proposed method improves spatial resolution and data consistency as compared with filtered backprojection, standard DPS, and DPS with re-sampling, while preserving fine lung structures. These results highlight the effectiveness of multi-resolution prior integration for low-dose CT reconstruction.**

I. Introduction

DIFFUSION models [1]–[3] have demonstrated remarkable performance in generative modeling tasks and have recently been applied to inverse problems through diffusion posterior sampling (DPS). [4] A key advantage of DPS is that a prior model can be learned through one-time unsupervised training and subsequently applied to a wide range of measurement models without retraining. This flexibility makes DPS particularly well suited for x-ray computed tomography (CT) reconstruction, where imaging protocols can vary substantially across patients and scanners, including differences in tube current, scan geometry, axial coverage, and pitch. Recent studies have demonstrated the effectiveness of DPS for CT reconstruction across diverse imaging protocols [5], and several acceleration strategies have been proposed to enable efficient processing of clinically sized datasets [6]–[8]. Together, these developments highlight DPS as a promising and data-efficient reconstruction framework.

Despite these advantages, DPS performance is fundamentally limited by the availability and quality of training data. In particular, large-scale high-resolution CT datasets are difficult to obtain in clinical settings, and existing high-resolution scans are often noisy, making them suboptimal for training diffusion priors. As a result, most diffusion-based CT priors are trained on standard clinical datasets that provide modest spatial resolution. Although such data are effective for capturing global anatomical structure, they offer limited support for reconstruction methods that seek to enhance spatial resolution beyond current clinical standards. Existing strategies to address this limitation include super-resolution preprocessing of training data [9]–[11] or the incorporation of priors trained on general-purpose image datasets. However, these approaches are only indirectly related to the true high-resolution anatomical features that may be present in CT images.

To address this challenge, we consider an alternative paradigm in which a clinically derived, nominal-resolution anatomical prior is incorporated but additional high-resolution prior information is introduced from a complementary dataset. For example, public clinical CT databases can provide abundant full-field anatomical data, while microCT or specialized high-resolution scans of *ex vivo* organs or tissue samples can offer much better structural detail for specific anatomical regions with low noise since there are no patient dose constraints. Such high-resolution priors are region-dependent since fine-scale anatomical structures can vary significantly across organs and tissue types. Thus, we employ strategies to apply tissue-specific priors in appropriate anatomical regions of interest. This combination has the potential to provide a globally consistent anatomical prior together with region-specific high-resolution information, enabling improved reconstruction in targeted regions of interest (ROIs).

Our proposed approach is called Mixed Prior Diffusion Posterior Sampling (Mixed Prior DPS) and integrates both global and local diffusion priors for low-dose CT reconstruction. The method employs a global diffusion model trained on full-field, nominal-resolution CT images to capture large-scale anatomical structure and low-frequency content, alongside a patch-based diffusion model trained on high-resolution local image patches to model fine-scale, high-frequency details. These two priors are combined through a frequency-domain fusion strategy, and a masking mechanism is used to selectively apply the mixed prior within appropriate anatomical ROIs while maintaining global consistency elsewhere. We develop a mathematical formulation of the proposed approach

This work was supported, in part, by Canon Medical Systems and by NIH grant R01EB035908. P. Teng, Electrical and Computer Engineering, Johns Hopkins University (e-mail:pteng3@jh.edu); X. Wang and J. Stayman, Biomedical Engineering, Johns Hopkins University (e-mail:web.stayman@jhu.edu); J. Zhou, L. Cai, T. Lee, and R. Zhang, Canon Medical Research USA, Inc.

and present an efficient algorithm for posterior sampling with mixed priors. Simulation results demonstrate that the proposed method is stable and effectively utilizes complementary information from both priors, enabling enhanced regional spatial resolution while preserving data consistency.

II. Methodology

A. Diffusion Posterior Sampling

1) Diffusion Model: A diffusion model is commonly formulated using a stochastic differential equation (SDE) [2], which progressively transforms the real data distribution $p_{data}(x)$ into a standard Gaussian distribution $N(0, I)$. In particular, the variance-preserving forward SDE is defined as

$$dx = -\frac{1}{2}\beta(t)xdt + \sqrt{\beta(t)}dw, \tag{1}$$

where $\beta(t)$ denotes a predefined noise schedule and w represents Brownian motion.

The time-dependent score function $\nabla_{x_t} \log p_t(x_t)$ is approximated by a neural network $s_\theta(x_t, t)$. Starting from Gaussian noise, generative samples are obtained by evolving the system toward the data distribution using the corresponding reverse-time SDE,

$$dx = [-\frac{1}{2}\beta(t)x - \beta(t)s_\theta(x_t, t)]dt + \sqrt{\beta(t)}d\bar{w}. \tag{2}$$

2) Diffusion Posterior Sampling: Conditional posterior sampling can be achieved by modifying the unconditional score in reverse SDE in (2) to the conditional score $\nabla_{x_t} \log p_t(x_t|y)$, where y denotes the measurement data. Applying Bayes' rule, the conditional score can be decomposed into the sum of the unconditional score and the gradient of the log-likelihood:

$$\nabla_{x_t} \log p_t(x_t|y) = \nabla_{x_t} \log p_t(x_t) + \nabla_{x_t} \log p_t(y|x_t). \tag{3}$$

The unconditional score $\nabla_{x_t} \log p_t(x_t)$ is approximated by the neural network $s_\theta(x_t, t)$. To approximate the likelihood gradient term, Chung et al. [4] proposed replacing the time-dependent variable x_t in the likelihood $\log p_t(y|x_t)$ with the posterior mean $\hat{x}_{0,t}$. This posterior mean is computed using Tweedie's formula:

$$\hat{x}_{0,t} = \mathbb{E}[x_0|x_t] = \frac{1}{\sqrt{\bar{\alpha}(t)}}(x_t + (1 - \bar{\alpha}(t))s_\theta(x_t, t)) \tag{4}$$

Substituting the resulting conditional score into the reverse SDE yields "standard" Diffusion Posterior Sampling (DPS).

3) Acceleration Strategies for DPS: To accelerate DPS, we adopt several efficiency strategies proposed in previous work [6]. First, a jump-start initialization is used to initialize the reverse diffusion process from a filtered backprojection (FBP) reconstruction rather than pure Gaussian noise. The FBP image provides informative structural priors at minimal computational cost. Although FBP reconstructions in low-dose or sparse-view settings may contain artifacts, effective initialization only requires a reasonable approximation of the image in the forward process. Specifically, Gaussian noise is added to the FBP reconstruction as

$$x_{\text{FBP},t} = \sqrt{\bar{\alpha}(t)}x_{\text{FBP}} + \sqrt{1 - \bar{\alpha}(t)}\epsilon. \tag{5}$$

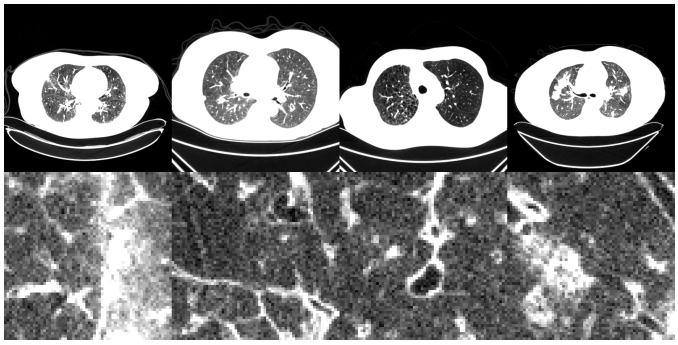


Fig. 1: Illustration of samples from global training dataset and local training dataset. The top row shows samples from global dataset. The bottom row shows samples from local dataset.

where $\epsilon \sim N(0, I)$. For a sufficiently large initialization time t_0, the FBP image after adding Gaussian noise approximates the distribution of x_{t_0},

$$x_{\text{FBP},t_0} \approx x_{t_0} \sim N(\sqrt{\bar{\alpha}(t_0)}x_0, (1 - \bar{\alpha}(t_0))I). \tag{6}$$

By the Markov property of diffusion models, the reverse process can therefore be initialized at t_0, skipping early diffusion steps and significantly reducing sampling time while maintaining stability and reconstruction quality.

Other strategies in [6] including gradient approximation, and accelerated optimization are also employed for efficient DPS computation.

B. Mixed Prior Diffusion Posteior Sampling

1) Global Prior Model and Local Prior Model: We employ two diffusion models to estimate the anatomical prior distribution. The global prior model is trained on full CT images and is used to capture the global data distribution. This model primarily captures large-scale information, including overall anatomy, object boundaries, and low-frequency image content.

The local prior model is trained separately on high-resolution image patches extracted from tissue specifc ROIs, such as the lung. By focusing on localized patches rather than full images, this model emphasizes fine-scale, high-frequency details that are often blurred or ignored in the global diffusion model.

Together, these two priors provide complementary information during reconstruction. The global prior enforces anatomical consistency and structural coherence, while the local prior enhances detailed textures and fine structures.

2) Histogram-Based Intensity Normalization: The global and local diffusion priors may be trained on CT datasets acquired under different imaging conditions, including variations scanning protocols, beam quality, corrections, etc. Consequently, the attenuation coefficient distributions of the two datasets may differ, introducing a domain shift that can hinder effective integration of the priors during posterior sampling.

To address this issue, we perform a histogram-based distribution alignment to match the intensity statistics of the global and local datasets. This preprocessing step preserves relative contrast while aligning the overall intensity range and

statistical properties. By operating in a consistent intensity space, the global and local diffusion priors can be better combined during posterior sampling. We have observed that this alignment improves stability and convergence, enabling effective fusion of global anatomical structure with local high-resolution detail.

3) Prior Mixing: We employ two prior models within the high-resolution ROI to guide reconstruction by jointly estimating the posterior mean of the ROI. To integrate information from both priors, we combine their posterior means in the Fourier domain, where the global prior contributes low-frequency components and the local prior contributes high-frequency details. Specifically, the mixed posterior mean $\hat{x}_{0,\text{mixed}}$ is defined as

$$\hat{x}_{0,\text{mixed}} = \mathcal{F}^{-1}\Big(\mathcal{L}\big(\mathcal{F}(\hat{x}_{0,\text{global}})\big) \; + \; \mathcal{H}\big(\mathcal{F}(\hat{x}_{0,\text{local}})\big)\Big) \quad (7)$$

where $\hat{x}_{0,\text{global}}$ and $\hat{x}_{0,\text{local}}$ denote the posterior means of ROI predicted by the global and local diffusion models, respectively; $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the Fourier and inverse Fourier transforms; and, $\mathcal{L}(\cdot)$ and $\mathcal{H}(\cdot)$ represent the low-pass and high-pass filters, respectively. This mixing allows each prior to operate at the spatial scale where it is most reliable, combining global structural consistency with locally detailed texture.

Directly mixing posterior means in the Fourier domain may introduce inconsistencies that accumulate during the reverse diffusion process and lead to unstable sampling. To solve this problem, we use a resampling [12] step after each likelihood update during posterior sampling. This resampling ensures that samples are drawn from a marginal distribution conditions on the current posterior mean, preventing error accumulation and stabilizing the reconstruction process.

4) Region-Dependent Prior Mixing with Masked DPS: During posterior sampling, prior mixing is applied only within the ROI for which the high-resolution prior is available, while DPS using only the global diffusion model is retained for the remaining image regions. This region-dependent prior estimation is enforced using a binary masking method, following the approach in [13]. Specifically, a binary mask $M \in \{0,1\}^{H\times W}$ is constructed to extract the ROI from the image, where $M(i,j) = 1$ indicates pixels located inside the ROI and 0 otherwise. At each posterior sampling step, the posterior mean of the full image is updated by combining the ROI-specific update with the global update as

$$\hat{x}_0 = M \odot \hat{x}_{0,\text{mixed}} + (1 - M) \odot \hat{x}_{0,\text{global}} \quad (8)$$

where $\hat{x}_{0,\text{mixed}}$ is computed using the mixed prior that combines global and local patch-based diffusion models, and $\hat{x}_{0,\text{global}}$ is obtained from the global diffusion model. While patch-based training alone may not provide sufficient global coherence, this limitation is effectively addressed in our reconstruction process. In particular, the physical measurement model enforced through the likelihood term and the global structural constraints imposed by the global diffusion prior.

In addition to the frequency-domain prior mixing described above, we introduce a temporal prior mixing strategy that operates across reverse diffusion time steps. During the reverse sampling process, the choice of prior model for computing the score function is region-dependent. Outside the ROI, the global diffusion model is used exclusively to compute the score at every time step. Within the ROI, however, the global and local diffusion models are alternately employed at different reverse sampling time steps. This alternating scheme allows the reconstruction within the ROI to benefit from both the global prior, which provides large-scale structural coherence and anatomical consistency, and the local prior, which contributes fine-scale, high-frequency detail. By interleaving the two priors temporally, this strategy complements the frequency-domain mixing of posterior means, offering an additional mechanism to integrate multi-resolution prior information throughout the reverse diffusion process. Together, the frequency-domain fusion and the temporal prior mixing enable a more comprehensive utilization of complementary prior knowledge, improving both stability and reconstruction quality within the ROI.

Additionally, the jump start strategy [6] initializes the reverse process with strong posterior restriction which significantly constrains the sampling solution space. These constraints allow the patch-based diffusion model to reliably predict local posterior means without introducing global inconsistencies. As a result, the mixed prior selectively enhances fine-scale structures within the ROI, while the global prior maintains anatomical consistency in the background, leading to improved stability, reconstruction quality, and computational efficiency during posterior sampling. The overall algorithm and pseudocode is presented in Algorithm 1.

Algorithm 1 Mixed Prior DPS

T': Jump start time step, x_{fbp}: FBP reconstruction, A, y: Forward projector and measurement
η: Step size, N: Number of iteration per reverse step, $m = 0$, $v = 0$: First moment vector and second moment vector, $\gamma_1 = 0.9$, $\gamma_2 = 0.999$: Adam momentum coefficients
β_t, α_t,σ_t and $\bar{\alpha}_t$: Scheduler, where $\alpha_i = 1 - \beta_i$, $\sigma_t = \beta_t$ and $\bar{\alpha}_i = \prod_{j=1}^{i} \beta_i$.
M: The mask of the ROI.
$x_{T'} = \sqrt{\bar{\alpha}_{T'}} x_{\text{FBP}} + \sqrt{1 - \bar{\alpha}_{T'}}\epsilon$, $\epsilon \sim N(0, I)$
for $t = T'$ to 1 **do**:
    $\hat{x}_{0,\text{global}} = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t + (1 - \bar{\alpha}_t)\epsilon_{\theta,\text{global}}(x_t, t))$
    $\hat{x}_{0,\text{local}} = M \odot (\frac{1}{\sqrt{\bar{\alpha}_t}}(x_t + (1 - \bar{\alpha}_t)\epsilon_{\theta,\text{local}}(x_t, t)))$
    $\hat{x}_{0,\text{mixed}} = \mathcal{F}^{-1}\Big(\mathcal{L}\big(\mathcal{F}(M \odot \hat{x}_{0,\text{global}})\big) \; + \; \mathcal{H}\big(\mathcal{F}(\hat{x}_{0,\text{local}})\big)\Big)$
    $\hat{x}_0 = M \odot \hat{x}_{0,\text{mixed}} + (1 - M)\hat{x}_{0,\text{global}}$
    for $i = 1$ to N **do**:
        $\hat{x}_0 = \text{Adam}(\hat{x}_0, \nabla_{\hat{x}_0} \|y - B\exp(-A\hat{x}_0)\|_2^2, \eta, \gamma_1, \gamma_2)$
    end for
    $x'_t = \sqrt{\bar{\alpha}_t}\hat{x}_0 + \sqrt{1 - \bar{\alpha}_t}\xi$, $\xi \sim N(0, I)$
    $z \sim N(0, I)$
    if t is odd **then**
        $\epsilon_\theta(x'_t, t) = M \odot \epsilon_{\theta,\text{global}}(x'_t, t) + (1 - M) \odot \epsilon_{\theta,\text{local}}(x'_t, t)$
    else
        $\epsilon_\theta(x'_t, t) = \epsilon_{\theta,\text{global}}(x'_t, t)$
    end if
    $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x'_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x'_t, t)) + \sigma_t z$
end for

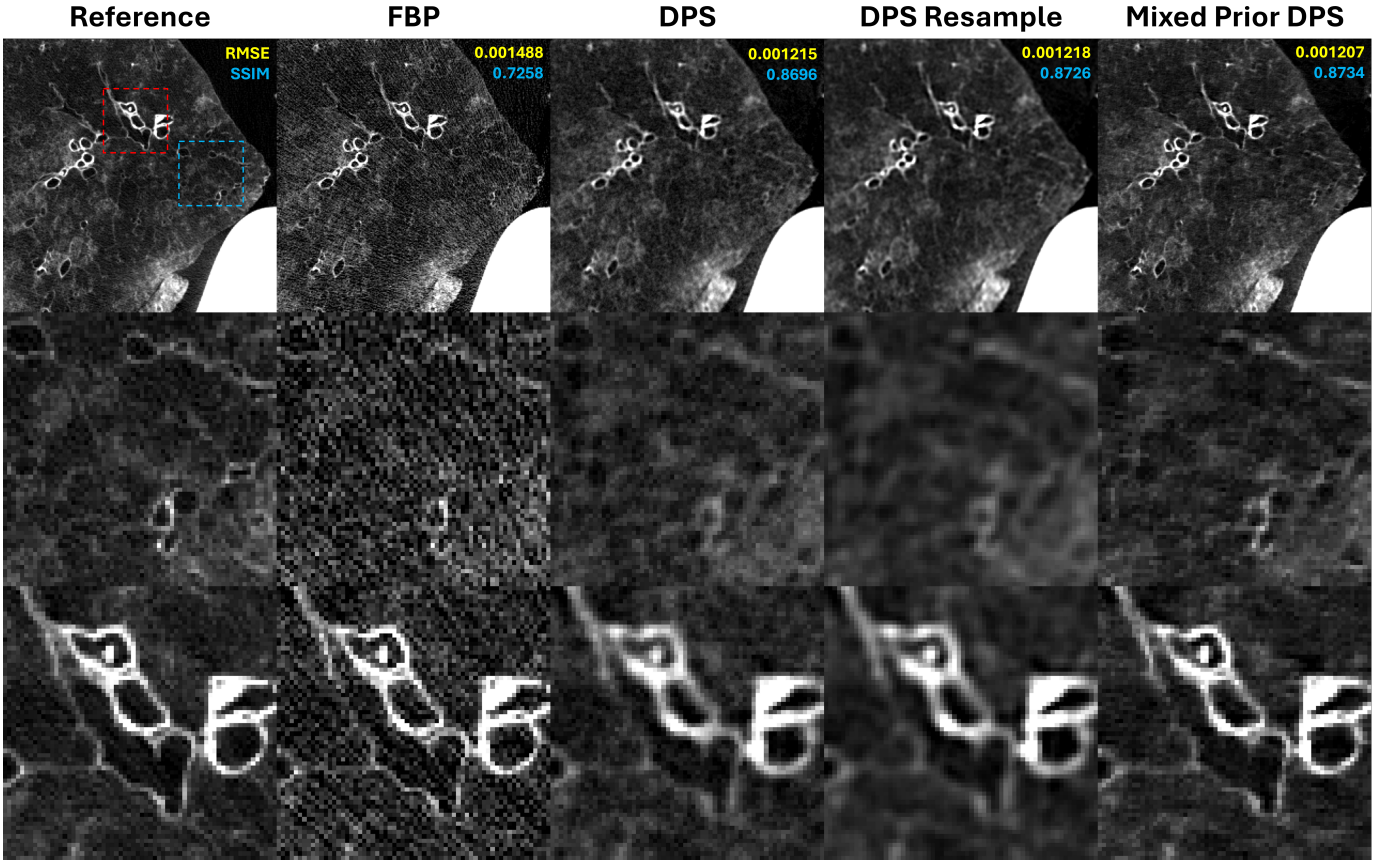


Fig. 2: Illustration of Mixed Prior DPS with comparisons to FBP, "standard" DPS, and DPS with resampling. The first rows shows different reconstructions of the plastinated lung. The bottom two rows show zoomed areas of the lung.

III. Experiments

A. Network Training and Data Acquisition

We employed a U-Net architecture [14] with an attention layer at the bottleneck to estimate both the global and local prior distributions. The diffusion models follow the DDPM framework [1] with a linear noise schedule defined by $\beta_1 = 0.0001$ and $\beta_T = 0.02$.

The global diffusion models were trained on the CT Lymph Nodes dataset [15]. The full image size is 512×512 with a voxel size of 0.75 mm. The training data were Fourier upsampled to 1024×1024 grids to match the high-resolution reconstruction. We selected the lung as the high-resolution ROI target for local diffusion model training and created lung masks using open source lung segmentation tool [16]. The high-resolution dataset used for training the local diffusion model was acquired using a flat-panel, cone-beam CT (CBCT) bench by scanning a plastinated lung. CBCT data were collected using 120 kVp, 120 mAs, and 1280 views. The detector pixel size was 0.2 mm $\times$ 0.2 mm, and the voxel size was 0.25 mm $\times$ 0.25 mm $\times$ 0.25 mm. Training patches of size 64×64 were extracted from multiple reconstructed volumes with a size of $30 \times 1024 \times 1024$. The training samples from two datasets are presented in Fig. 1.

B. Low-Dose CT Reconstruction and Hyper-parameters

We use the proposed Mixed Prior DPS to perform low-dose CT reconstruction. We adopt a standard monoenergetic CT system model, $y = I_0 \exp(-Ax)$, where A is the linear forward projection matrix, I_0 is the incident photon intensity, and y denotes the measured data.

For evaluation, we scanned a body phantom with a plastinated lung inserted using our CBCT bench system to simulate a high-resolution scan. The scan parameters were 120 kVp, 10 ms, and 30 mA. The reference image was obtained by averaging four independent scans. Low-dose CT data were simulated by applying binomial thinning to the Poisson-distributed photon counts with $p = 0.5$, corresponding to half the dose of a single scan. The proposed Mixed Prior DPS was then evaluated on the resulting low-dose CT reconstruction.

The CT system geometry uses a source-to-axis distance of 872.96 mm and a source-to-detector distance of 1351.3 mm, with a detector pixel size of 0.278 mm, voxel size of 0.35 mm and 1080 views covering $360°$. We compare the proposed Mixed Prior DPS with filtered backprojection (FBP), standard DPS [4], and standard DPS with resampling [12]. The hyperparameters in the acceleration strategies [7] for all methods, including the jump-start parameter T', the maximum-likelihood optimizer step size η, and the number of optimizer iterations per step N, are fine-tuned as follows: (i) Standard DPS: $T' = 30$, $\eta = 4 \times 10^{-5}$, $N = 14$; (ii) Standard DPS

with resampling: $T' = 30$, $\eta = 2 \times 10^{-4}$, $N = 24$; (iii) Mixed Prior DPS: $T' = 30$, $\eta = 2 \times 10^{-4}$, $N = 24$.

IV. Results

The reconstruction results are presented in Fig. 2. Compared with the baseline methods, the proposed Mixed Prior DPS achieves improved spatial resolution and better data consistency in the reconstructed images. In particular, fine anatomical structures within the lung region are more clearly resolved, with sharper boundaries and reduced noise. In contrast, standard DPS exhibits noticeable noise and blurry, while DPS with resampling tends to oversmooth high-frequency details. These results demonstrate that combining global and local priors enables Mixed Prior DPS to better balance structural fidelity and detail preservation in low-dose CT reconstruction.

We evaluated image quality using quantitative metrics, including root-mean-square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM), for the proposed Mixed Prior DPS and the comparison methods. As summarized in Fig. 2 the first row, Mixed Prior DPS achieves the lowest RMSE and the highest PSNR and SSIM among all methods. Specifically, Mixed Prior DPS reduces RMSE compared with standard DPS and DPS with resampling, while improving PSNR and SSIM, indicating enhanced reconstruction accuracy and data consistency.

V. Conclusion

We proposed Mixed Prior Diffusion Posterior Sampling, a reconstruction method that integrates global and local diffusion priors for low-dose CT reconstruction. By combining a global diffusion model trained on full-size, nominal-resolution CT images with a local diffusion model trained on high-resolution patches, the proposed method leverages complementary prior information. The frequency-domain fusion strategy and temporal prior mixing enable effective integration of low-frequency anatomical structure with high-frequency detail, while the masked DPS ensures that high-resolution enhancement is applied selectively within targeted anatomical regions. Experimental results on low-dose CT data reconstruction demonstrate that Mixed Prior DPS achieves improved spatial resolution, reduced noise, and better data consistency compared with FBP, standard DPS, and DPS with resampling. This method offers a flexible paradigm for incorporating specialized high-resolution datasets, such as those acquired from *ex vivo* specimens, into clinical CT reconstruction pipelines, and may be extended to other anatomical regions and imaging applications in future work.

References

[1] Ho, J., Jain, A., and Abbeel, P., "Denoising diffusion probabilistic models," *Advances in neural information processing systems* **33**, 6840–6851 (2020).

[2] Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., and Poole, B., "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456* (2020).

[3] Song, Y. and Ermon, S., "Generative modeling by estimating gradients of the data distribution," *Advances in neural information processing systems* **32** (2019).

[4] Chung, H., Kim, J., Mccann, M. T., Klasky, M. L., and Ye, J. C., "Diffusion posterior sampling for general noisy inverse problems," *arXiv preprint arXiv:2209.14687* (2022).

[5] Li, S., Jiang, X., Tivnan, M., Gang, G. J., Shen, Y., and Stayman, J. W., "Ct reconstruction using diffusion posterior sampling conditioned on a nonlinear measurement model," *Journal of Medical Imaging* **11**(4), 043504–043504 (2024).

[6] Jiang, X., Li, S., Teng, P., Gang, G., and Stayman, J. W., "Strategies for ct reconstruction using diffusion posterior sampling with a nonlinear model," *ArXiv* , arXiv–2407 (2024).

[7] Teng, P., Jiang, X., Cai, L., Lee, E., Zhang, R., Zhou, J., and Stayman, J. W., "3d diffusion posterior sampling for ct reconstruction," in [*Medical Imaging 2025: Physics of Medical Imaging*], **13405**, 202–207, SPIE (2025).

[8] Wang, X., Shi, G., Sivakumar, A., Ye, T., Sylvester, A., Stayman, J., and Zbijewski, W., "A conditional generative diffusion model of trabecular bone with tunable microstructure," in [*Medical Imaging 2025: Clinical and Biomedical Imaging*], **13410**, 79–84, SPIE (2025).

[9] Fok, W. Y. R., Fieselmann, A., Herbst, M., Ritschl, L., Kappler, S., and Saalfeld, S., "Deep learning in computed tomography super resolution using multi-modality data training," *Medical Physics* **51**(4), 2846–2860 (2024).

[10] Xuan, K., Si, L., Zhang, L., Xue, Z., Jiao, Y., Yao, W., Shen, D., Wu, D., and Wang, Q., "Reducing magnetic resonance image spacing by learning without ground-truth," (2020).

[11] Wang, X., Song, Z., Zhu, Y., Wang, S., Zhang, L., Shen, D., and Wang, Q., "Inter-slice super-resolution of magnetic resonance images by pre-training and self-supervised fine-tuning," in [*2024 IEEE International Symposium on Biomedical Imaging (ISBI)*], 1–5 (2024).

[12] Song, B., Kwon, S. M., Zhang, Z., Hu, X., Qu, Q., and Shen, L., "Solving inverse problems with latent diffusion models via hard data consistency," *arXiv preprint arXiv:2307.08123* (2023).

[13] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., and Van Gool, L., "Repaint: Inpainting using denoising diffusion probabilistic models," in [*2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*], 11451–11461 (2022).

[14] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).

[15] Roth, H., Lu, L., Seff, A., Cherry, K. M., Hoffman, J., Wang, S., Liu, J., Turkbey, E., and Summers, R. M., "A new 2.5 d representation for lymph node detection in ct (ct lymph nodes)," (2015).

[16] Hofmanninger, J., Prayer, F., Pan, J., Röhrich, S., Prosch, H., and Langs, G., "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental* **4** (Aug. 2020).

Diffusion Posterior Sampling for Angiographic Reconstruction using Limited Angle and Sparse View data for 4D-DSA

Shudong Li, Suyu Liao, Peiqing Teng, J. Webster Stayman, Grace J. Gang

***Abstract*—Four-dimensional digital subtraction angiography (4D-DSA) is an effective modality for the diagnosis and treatment planning of neurovascular interventions. Single time point 4D-DSA Projection data is usually underdetermined for a 3D reconstruction. Current image formation therefore relies on approximations using a 3D-DSA. In this work, we investigate application of advanced reconstruction on limited angle and sparse view data acquired at a single time point in 4D-DSA. We adopted the diffusion posterior sampling (DPS) algorithm which combines a physics-based data fidelity with an anatomically realistic diffusion prior. The diffusion prior was trained using axial slices extracted from 3D liver vasculature generative model. A total variation (TV) penalty along the z direction was applied to encourage continuity between slices. We compared DPS with a model-based approach using a directional TV penalty (DTV). We simulated a single-arc geometry on a single-plane C-arm and inverse geometry tomosynthesis system as well as an orthogonal-arc geometry achievable using a bi-plane C-arm systems. The DPS algorithm is able to reconstruct high-fidelity vascular anatomy under the challenging geometry of 10 views over a single arc of 40 degrees. As expected, vessels with thinner branches and/or branches in directions corresponding to the missing frequencies are more challenging to reconstruct. Orthogonal arcs improve the reconstruction quality of both algorithms, indicating the importance of more distributed angular sampling. DPS outperforms DTV under all conditions evaluated. The DPS algorithm provides an effective approach to reconstruction 4D-DSA data acquired using limited angle and sparse view geometries.**

***Index Terms*—CT reconstruction, 4D-DSA, deep learning reconstruction, diffusion posterior sampling.**

I. Introduction

THE diagnosis and treatment of vascular abnormalities require a high spatial and temporal resolution imaging modality for accurate visualization of pathology [1]. Conventional 2D-DSA offers high temporal resolution but does not allow visualization of the complex 3D vasculature; 3D-DSA enables volumetric visualization but provides no temporal information [2]. Four-dimensional digital subtraction angiography (4D-DSA) has been developed as a promising imaging technique that overcomes the limitations of both 2D and 3D-DSA. Multiple clinical studies have reported the effectiveness of 4D-DSA in improving the diagnosis and treatment planning of neurovascular applications, including arteriovenous malformation (AVM) evaluation and flow measurements [3].

Shudong Li is in the Electrical and Computer Engineering Department at Johns Hopkins University, baltimore, MD, 21219, USA, e-mail: sli248@jhu.edu

Suyu Liao is in the Radiology Department at University of Pennsylvania, Philadelphia, PA 19104, USA, e-mail: suyuliao@seas.upenn.edu

Peiqing Teng is in the Electrical and Computer Engineering Department at Johns Hopkins University, baltimore, MD, 21219, USA, email: pteng3@jhu.edu

J. Webster Stayman is in the Biomedical Engineering Department at Johns Hopkins University, baltimore, MD, 21219, USA, e-mail: web.stayman@jhu.edu

Grace J. Gang is in the Radiology Department at University of Pennsylvania, Philadelphia, PA 19104, USA, email: grace.j.gang@pennmedicine.upenn.edu

The current implementation of 4D-DSA produces multiple time-resolved 3D angiographic volumes from the standard 3D-DSA rotational acquisitions on C-arm angiography systems[4]. Alternatively, 4D-DSA has also been investigated using limited angle, tomosynthesis data[5]. To achieve high temporal resolution, it is challenging with current hardware to acquire projection data spanning a large enough angular range necessary for a complete 3D reconstruction. To circumvent this problem, current methods for 4D-DSA image formation typically rely on a 3D DSA to constrain the locations of the contrast-enhanced vessels. The contrast distribution at a particular time point can then be approximated using the backprojection of the corresponding projection(s)[6]. With advances in image reconstruction techniques in recent years, the first goal of this work is to investigate whether more sophisticated reconstruction methods can improve image quality for 4D-DSA images without relying on a high quality 3D-DSA. Simultaneously, the commercial availability of inverse geometry tomosynthesis systems[7] presents opportunities for ultra-fast acquisition of limited angle, sparse view tomosynthesis data that can be leveraged for 4D-DSA applications. The second goal of this work is to investigate the feasibility of such systems for 4D-DSA when combined with advanced reconstruction.

The limited angle and sparse view reconstruction problem is under-determined. While reconstructing the full patient anatomy using such data is extremely challenging, the structures we seek in DSA are fairly simple. Model-based reconstruction with sparsity-enforcing priors has been applied to this class of problems [8]. Recently, diffusion posterior sampling (DPS) has been proposed as an efficient framework for solving inverse problems in CT [9]. The DPS algorithm relies on a data fidelity term coupled with an anatomically realistic prior trained using a generative diffusion model. This method has been successfully applied to mitigate artifacts associated with limited angle or sparse view data in CT [10].

In this work, we applied DPS algorithm to reconstruct a volumetric angiographic image from limited angle, sparse view projections corresponding to one time point in a 4D DSA sequence. We trained a diffusion prior representing the 3D contrast-enhanced vasculature in the liver. Different geometry configurations including the angular arc and number of views were investigated. Image quality was compared among images reconstructed using a model-based approach with directional total variation (TV) regularizer, and DPS.

II. Method

A. Forward model

We adopt the following tomographic forward model where measurement data, $\mathbf{y}$, are approximated as independent Gaussian random variables:

$$\mathbf{y} \sim N(\bar{\mathbf{y}}, \mathbf{D}\{\bar{\mathbf{y}}\}) \text{ where } \bar{\mathbf{y}} = I_0 \exp(-\mathbf{A}\mathbf{x_0}), \quad (1)$$

Here, $\mathbf{A}$ is the forward projector defined for the given CT imaging system, $\mathbf{x_0}$ denotes the image of linear attenuation coefficients, and I_0 a scalar gain term presuming uniform x-ray fluence and detector sensitivity. Measurements y are Gaussian distributed with mean and variance both equal to their mean, $\bar{\mathbf{y}}$. The line integrals, $\mathbf{l}$, can also be approximated as Gaussian:[11]

$$\mathbf{l} \sim N(\bar{\mathbf{l}}, \mathbf{D}^{-1}\{\mathbf{y}\}) \text{ where } \bar{\mathbf{l}} = -\ln(\frac{\mathbf{y}}{I_0}), \quad (2)$$

Data for DSA are acquired using a protocol consisting of two acquisitions, one before contrast injection and one after. Under the assumption of negligible patient motion between the two acquisitions, their line integrals can be subtracted to yield the line integrals resulting from the contrast-enhanced vessels only, $\mathbf{l_v}$:

$$\mathbf{l_v} \sim N(\bar{\mathbf{l}}_\mathbf{v}, \mathbf{\Sigma_v^{-1}}) \text{ where } \bar{\mathbf{l}}_\mathbf{v} = \bar{\mathbf{l}}_\mathbf{c} - \bar{\mathbf{l}}_\mathbf{a}, \quad (3)$$

where $\bar{\mathbf{l}}_\mathbf{c}$ and $\bar{\mathbf{l}}_\mathbf{a}$ are the post- and pre-contrast sinograms, respectively. The pre- and post contrast sinograms are statistically independent. The covariance of the vessel sinogram, $\mathbf{\Sigma_v^{-1}}$, is therefore:

$$\mathbf{\Sigma_v^{-1}} = \mathbf{D}^{-1}\{\mathbf{y_c}\} + \mathbf{D}^{-1}\{\mathbf{y_a}\}. \quad (4)$$

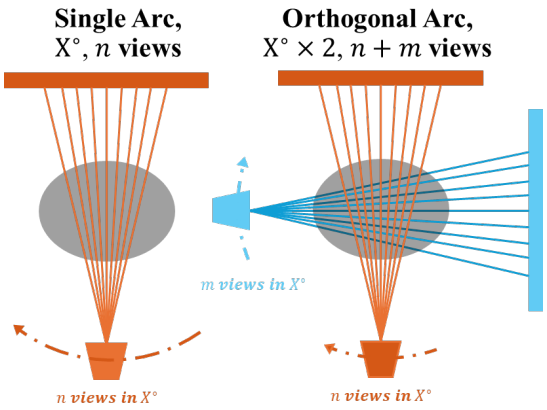


Fig. 1: Illustration of the acquisition geometries. Single Arc with n views in $X°$ (left): n views evenly distributed in $X°$. Orthogonal Arc with $n + m$ views in $X° \times 2$ (right): There are two arc orthogonal to each other while one arc covers $X°$ with m views, the other arc covers $X°$ with n views.

B. Model-based Reconstruction with Direction Total Variance

We first investigated a model-based reconstruction (MBIR) approach. The objective function consists of a Gaussian log-likelihood term and a directional total variation (DTV) regularization[12], $reg(\cdot)$, with regularization strength, λ:

$$\mathbf{x_v^*} = \arg\min_{\mathbf{x_v}} \|\mathbf{A}\mathbf{x_v} - \bar{\mathbf{l}}_\mathbf{v}\|^2_{\mathbf{\Sigma_v^{-1}}} + \lambda reg(\mathbf{x_v}). \quad (5)$$

The DTV regularization can effectively promote directionally sparse gradient representations, and therefore is well suited for angiographic reconstruction. It is written as:

$$\lambda reg(\mathbf{x_v}) = \lambda_x \|\nabla_x \mathbf{x_v}\|_1 + \lambda_y \|\nabla_y \mathbf{x_v}\|_1 + \lambda_z \|\nabla_z \mathbf{x_v}\|_1. \quad (6)$$

where ∇_x, ∇_y, and ∇_z represent finite-difference operators along the three spatial directions.

C. Diffusion Posterior Sampling with Advance Strategies

We applied the DPS reconstruction algorithm with several advanced acceleration strategies proposed in Jiang et al.[13] The data fidelity term is similar to that in Eq.5. The hand-crafted prior is replaced with a generative prior trained using a diffusion model (see Sec.II-D).

Instead of starting the reverse process from pure Gaussian noise at time step T, we applied the "jump-start" method[13]. The reverse process starts from time step t' and is initialized with a maximum likelihood reconstruction, $\mathbf{x_{ML}}$, with additive Gaussian noise:

$$\mathbf{x_{ML},t'} = \sqrt{\bar{\alpha}_{t'}}\mathbf{x_{ML}} + (1 - \bar{\alpha}_{t'})\epsilon_\mathbf{noise}. \quad (7)$$

Here, $\epsilon_\mathbf{noise} \sim N(0, \mathbf{I})$ represents independent Gaussian noise, and $\bar{\alpha}_t$ is a variance scheduler defined as a function of time step $t \in [0, T]$.

Furthermore, the likelihood term $p(\mathbf{y}|\mathbf{x_t})$ can be approximated by $p(\mathbf{y}|\hat{\mathbf{x}}_\mathbf{0})$ where $\hat{\mathbf{x}}_\mathbf{0}$ is Tweedie's approximation of $\mathbf{x_0}$ from $\mathbf{x_t}$:

$$p(\mathbf{y}|\mathbf{x_t}) \approx p(\mathbf{y}|\hat{\mathbf{x}}_\mathbf{0}) \text{ where}$$
$$\hat{\mathbf{x}}_\mathbf{0} := E[\mathbf{x_0}|\mathbf{x_t}] = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x_t} + (1 - \bar{\alpha}_t)\mathbf{s}_{\theta^*}(\mathbf{x_t}, t)), \quad (8)$$

We then conduct diffusion posterior sampling by alternating between a single diffusion update and multiple maximum likelihood updates. The update equations are:

$$\mathbf{x}'_{t-1} = \mathbf{x}_t - [\mathbf{f}(\mathbf{x_t}, t) + g^2(t)\mathbf{s}_{\theta^*}(\mathbf{x_t}, t)]\mathrm{d}t - g(t)\mathrm{d}\bar{\mathbf{w}}_\mathbf{t}, \quad (9)$$

$$\hat{\mathbf{x}}'_0 = \hat{\mathbf{x}}_0 - \eta_t \nabla_{\hat{\mathbf{x}}_0} \|\mathbf{A}\hat{\mathbf{x}}_\mathbf{0} - \bar{\mathbf{l}}_\mathbf{v}\|^2_{\mathbf{\Sigma_v^{-1}}} - \lambda_z \|\nabla_z \hat{\mathbf{x}}_\mathbf{0}\|_1, \quad (10)$$

$$\mathbf{x}_{t-1} = \mathbf{x}'_{t-1} - \hat{\mathbf{x}}_0 + \hat{\mathbf{x}}'_0, \quad (11)$$

where $\mathbf{x}'_{t-1}$ is intermediate sample from reverse diffusion sampling, $\hat{\mathbf{x}}_0$ is Tweedie's estimation from $\mathbf{x}'_{t-1}$, $\hat{\mathbf{x}}'_0$ is the updated estimation from maximum likelihood, and $\mathbf{x}_{t-1}$ is the final update. A TV regularization along the z direction is applied in the likelihood update in Eq.10.

D. Data Preparation and Training of the Diffusion Model

To train the generative prior for DPS, we generated 3822 3D vessel trees using the Hepatic Vessel Generation Tool [14]. Axial slices were extracted from the vasculature models as training data, i.e., the prior generates axial cuts through a vasculature tree. For training, we adopted a 2D U-Net [15] with an attention residual and time embedding module implemented in *PyTorch*.[16] For diffusion sampling, we predefined the scheduler $\bar{\alpha}_t = 1e^{-5t}$ with $t \in (0, T]$ where $T = 1$. We used the Adam optimizer with an initial learning rate of 0.001. The batch size was 8 and the numbers of epochs was set to 600 with an early stopping criterion.

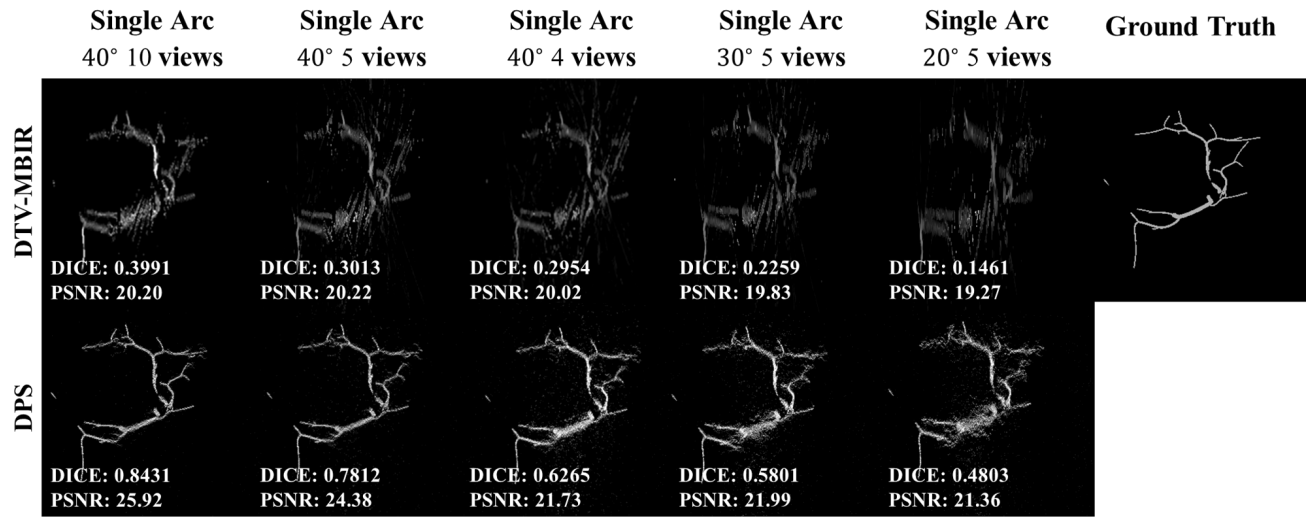


Fig. 2: Reconstructions of DTV-MBIR and DPS approach under single arc geometries with different angular extent and view counts. First three columns from the left optimize over view counts (10, 5, and 4 views) while last two columns optimize over angular extension (40○, 30°, 20°). All images share a common colormap with W/L = 2000/0 HU.

E. Simulation Data and Acquisition Geometry

We used the CT Lymph Node dataset [17] as the ground truth, pre-contrast anatomy. Vascular trees, scaled to attenuation coefficients corresponding to $20mg/mL$ of iodine, were digitally inserted into the liver to simulate post-contrast ground truth. Projection data were simulated using a monoenergetic beam at $80\ keV$ with a barebeam fluence of $I_0 = 10^6$ photons/detector element. The source-to-axis and source-to-detector distances are $800\ mm$ and $1500\ mm$, respectively. Separable footprint forward- and back-projectors from the *CTorch* [18] package were used during reconstructions.

We simulated different acquisition geometries to investigate their impact on imaging performance. For single-plane C-arms or standard tomosynthesis units, we simulated data acquired over a "single arc" of varying angular extent (10^o to 40^o) and different number of views (2-10) evenly distributed over the angle. Since bi-plane C-arms are often used for neurovascular interventions, we also performed data acquisition over two "orthogonal arcs", also as as function of different angular extent and number of views. Example geometries and notations are illustrated in Fig.1.

F. Evaluation

We compared the image quality of DPS and MBIR-DTV reconstruction under different acquisition geometries. Since the vascular tree is sparse in 3D, we computed the maximum intensity projection (MIP) along the z direction to visualize the results. The Dice similarity coefficient (DICE) and Peak Signal-to-Noise Ratio (PSNR) were computed as quantitative metrics for image quality. For DICE, we fist obtained vessel masks using a simple threshold equal to 50% of the linear attenuation coefficient of iodine.

We used DICE to optimize algorithmic parameters for MBIR-DTV and DPS. For MBIR-DTV, an exhaustive sweep of $(\nabla_x, \nabla_y, \nabla_z)$, each ranging from 5×10^{-8} to 5×10^{-4} was performed. For DPS, we optimized the jump start step t' (50-200), number of likelihood updates (2-80), and learning rate (0.0005-0.1).

III. Results

Figure 2 illustrates the MIP of the reconstruction results under the single-arc geometries. The first three columns correspond to a fixed angular extent at $40°$ with progressively fewer views (10, 5, and 4 views). With 10 views over $40°$, the DPS image is able to recover the vessel anatomy with high fidelity; the DTV-MBIR image, on the other hand, are blurrier in general and exhibit significant smearing along the central horizontal vessels - which corresponds to the directions of missing frequencies in the limited angle sampling scheme. As the number of views decreases, the performance of both algorithms suffers due to reduced sampling density along the acquisition trajectory. The last two columns of Figure 2 isolate the effect of reducing angular extent while keeping the number of views fixed at 5. Compared to the $40°$ configuration, reconstructions under $30°$ and $20°$ exhibit substantially increased blurring and streaking. Both DTV-MBIR and DPS show performance degradation, with the most severe case occurring at $20°$ and 5 views. These results suggest that, for single-arc acquisitions, angular extent is more important than number of views. The missing frequencies as a result of limited

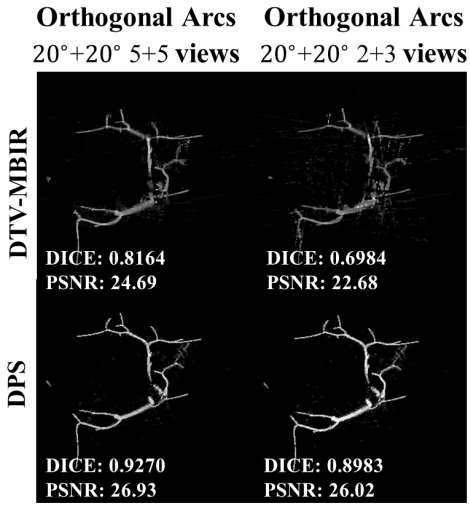


Fig. 3: Reconstructions under orthogonal arcs $20° + 20°$ geometries with $5 + 5$ and $2 + 3$ views.

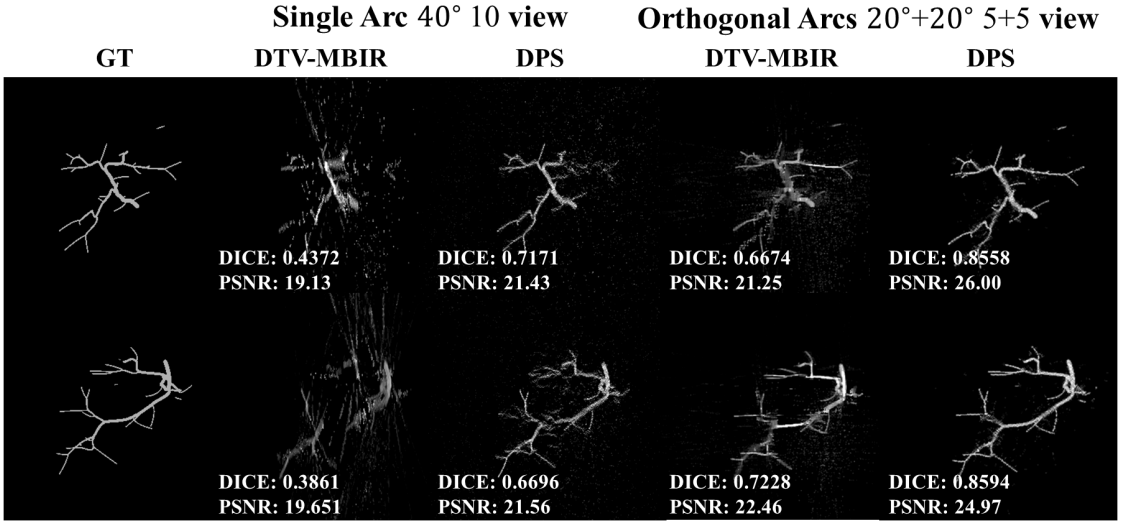


Fig. 4: Reconstructions under single-arc, comparing DTV-MBIR and DPS approach. Section on the left is under single arc 40° with 10 views while section on the right is orthogonal arcs 40° with $5+5$ views. Each row represent a realization. All images share a common colormap with W/L = 2000/0 HU.

angular coverage tend to be more challenging to recover using diffusion-based priors.

Figure 3 shows the reconstruction results using the orthogonal-arc geometries. For the same total angular coverage and number of views, orthogonal acquisitions yield substantially improved reconstruction quality for both methods, suggesting that a more distributed sampling strategy is advantageous for both model-based and DPS reconstructions. Reconstructions using the DPS algorithm consistently outperforms DTV-MBIR across all geometries.

Figure 4 presents reconstruction results for two additional vascular realizations using the best-performing single-arc and orthogonal-arc geometries. Performance comparisons between DTV-MBIR and DPS are similar to Figs.2 and 3. Interestingly, under the single-arc geometry, performance is noticeably worse for these additional realizations due to the presence of thinner branches along the horizontal direction. These observations suggest that geometry optimization and DPS parameter tuning should be evaluated across a broader range of vascular anatomies to ensure robustness. Under the orthogonal arc geometries, on the other hand, performance is fairly robust across all three vasculatures examined.

IV. Discussion and Conclusion

We proposed a diffusion posterior sampling (DPS) framework for limited-angle, sparse-view angiographic reconstruction that can be used as a single time point in a 4D-DSA sequence. The DPS algorithm outperforms DTV-MBIR in all geometries investigated and was able to recover high fidelity vascular anatomy under 40° using 10 views. A more distributed sampling strategy using orthogonal arcs is more beneficial and yields more robust performance across different vessel realizations. We are pursuing several areas of future work. First, we assumed negligible motion between the pre- and post-contrast acquisitions. Development of motion correction strategies and simulation of realistic patient motion are underway. Second, reconstructions in this work correspond to one time point in a 4D-DSA sequence. Temporal information across different time points can potentially provide a powerful form of regularization. Incorporation of temporal priors will be investigated in future work. Finally, we will evaluate image quality beyond simple metrics like DICE and PSNR. Specifically, we will develop metrics that reflect the extent of hallucinated vessels and evaluate the impact of reconstruction quality on 4D-DSA measurement including time attenuation curve, flow, etc. The DPS algorithm provides an effective approach to reconstruct 4D-DSA data acquired using limited angle and sparse view geometries.

Acknowledgment

Work was supported, in part, by NIH grants R01EB035908 and R01CA249538.

References

[1] Sandoval-Garcia, C *et al.*, *Journal of neurointerventional surgery*, vol. 8, no. 3, pp. 300–304, 2016.
[2] Davis, B *et al.*, *American Journal of Neuroradiology*, vol. 34, no. 10, pp. 1914–1921, 2013.
[3] Chen, K.-K *et al.*, *American Journal of Neuroradiology*, vol. 38, no. 4, pp. 740–746, 2017.
[4] Mistretta, C. A *et al.*, in *Medical Imaging 2010: Physics of Medical Imaging*, vol. 7622. SPIE, 2010, pp. 730–737.
[5] Buehler, M *et al.*, in *Medical Imaging 2017: Physics of Medical Imaging*, vol. 10132. SPIE, 2017, pp. 595–606.
[6] Oberstar, E. L *et al.*, in *Medical Imaging 2016: Physics of Medical Imaging*, vol. 9783. SPIE, 2016, p. 978302.
[7] Stephens, K, *AXIS Imaging News*, 2024.
[8] Sidky, E. Y *et al.*, *Journal of Medical Imaging*, vol. 12, no. S1, pp. S13 013–S13 013, 2025.
[9] Chung, H and Ye, J. C, in *Generative Machine Learning Models in Medical Image Computing*. Springer, 2025, pp. 129–148.
[10] Li, S *et al.*, *Journal of Medical Imaging*, vol. 11, no. 4, pp. 043 504–043 504, 2024.
[11] Fessler, J. A, in *Handbook of Medical Imaging, Volume 1: Physics and Psychophysics*. SPIE, 2000, updated PDF version dated April 20, 2016.
[12] Zhang, Z *et al.*, *Medical Image Analysis*, vol. 70, p. 102030, 2021.
[13] Jiang, X *et al.*, *ArXiv*, pp. arXiv–2407, 2024.
[14] Whitehead, J. F *et al.*, *Medical physics*, vol. 50, no. 9, pp. 5505–5517, 2023.
[15] Ronneberger, O *et al.*, in *MICCAI 2015*. Springer, 2015, pp. 234–241.
[16] Song, Y *et al.*, *arXiv preprint arXiv:2011.13456*, 2020.
[17] Roth, H. R *et al.*, in *International conference on medical image computing and computer-assisted intervention*. Springer, 2014, pp. 520–527.
[18] Jiang, X *et al.*, *arXiv preprint arXiv:2503.16741*, 2025.

Deep Learning Deblooming in Coronary CT Angiography with CatSim Training Data

Muhan Shao[1], Christopher Wiedeman[1], Lin Fu[1], Ge Wang[2], and Bruno De Man[1]

[1]GE HealthCare, Technology and Innovation Center, Niskayuna, NY, USA
[2]Department of Biomedical Engineering, School of Engineering, Rensselaer Polytechnic Institute, Troy, New York, USA

***Abstract*—Blooming artifacts caused by calcified plaques remain a major challenge in coronary CT angiography (CCTA), as they obscure vessel lumen and reduce diagnostic accuracy of coronary stenosis. This study proposes a deep learning (DL)-based framework to improve lumen visualization in CCTA images affected by calcium blooming. To address the lack of real ground-truth images without blooming artifacts, a simulation framework was designed to generate realistic paired data by inserting synthetic calcium into clinical CCTA volumes. A modified 2D U-Net++ was developed for image deblooming. A calcium mask was incorporated as an additional input channel to provide coarse localization of calcium regions. Quantitative evaluation on held-out test data demonstrated that the proposed model substantially improved deblooming performance. Mean RMSE within the calcification ROI decreased from 58.7 HU in the input to 46.6 HU with the proposed model. Dice scores increased by 4% and 15% across two threshold settings, indicating more accurate recovery of calcium shape and size. Overall, the proposed DL framework effectively reduces calcium blooming artifacts in CCTA images despite the absence of clinical paired training data. This approach has the potential to enhance lumen visualization and improve diagnostic accuracy in patients with heavy coronary calcification.**

***Keywords*—coronary CT angiography, cardiac CT, calcium deblooming, deep learning**

I. Introduction

Coronary CT angiography (CCTA) is an essential non-invasive imaging modality for evaluating coronary artery disease [1]. It provides detailed assessment of plaque burden [2], luminal narrowing, and overall coronary anatomy. However, the diagnostic accuracy of CCTA is often affected by blooming artifacts introduced by calcified plaques. Blooming artifacts arise from a combination of factors. The primary reason is partial volume averaging [3], which will cause high-attenuation calcium regions to appear larger than their true physical size. This artifactually enlarged calcium can obscure the lumen and lead to overestimation of coronary stenosis severity [4], [5], limiting the reliability of CCTA, particularly in patients with heavy coronary calcification. Therefore, it is of great importance to develop effective deblooming techniques to improve lumen visualization and stenosis assessment in these challenging cases.

Many approaches have been explored to mitigate calcium blooming, including hardware improvements to achieve higher spatial resolution, better image-reconstruction algorithms, and post-processing corrections [6]. More recently, data-driven approaches based on deep learning (DL) have shown promising performance across a range of CT image enhancement tasks, such as noise reduction, super-resolution, and artifact correction. Prior studies have developed DL algorithms to reduce blooming artifacts in CCTA. Ref [7] proposed a Dense-Unet architecture to remove calcifications and restore the underlying coronary lumen. Ref [8] employed an enhanced super-resolution generative adversarial network to suppress blooming artifacts and demonstrated improved diagnostic accuracy in CCTA. Ref [9] introduced a semantic-oriented network that uses the obtained semantic information to guide the network to have better deblooming performance.

In this study, we propose an end-to-end DL framework for CCTA deblooming. One of the major challenges in applying DL methods to CCTA deblooming application is the lack of suitable training data. High-quality CCTA images without blooming artifacts cannot be directly obtained in clinical practice, which has limited the development of supervised learning approaches. Inspired by our previous works [3], we built a simulation framework to generate realistic training data that models the blooming process. By inserting synthetic calcium into clinical cardiac images, this framework enables creation of large amounts of paired data with controllable blooming severity and noise characteristics.

The deblooming network is built upon a modified 2D U-Net++ [10] architecture which is optimized for fine-detail restoration through its dense skip-connection layers and multi-scale feature fusion design. A calcium mask is incorporated in the network, which serves as a spatial prior highlighting the locations of blooming artifacts. Because

the calcified plaques occupy only a small portion of the image, this additional input channel helps make the network focus on calcification regions while preserving anatomical structures elsewhere in the image.

Overall, we developed a robust DL-based deblooming algorithm for CCTA images despite the absence of real paired training data. The proposed method improves the estimation of the true size and shape of calcifications, thereby enhances lumen visualization and recovery in the presence of calcified plaques, which potentially supports more accurate diagnostic interpretation in CCTA.

II. Materials and Methods

A. Synthetic training data generation

A major challenge in CCTA deblooming is the lack of paired images with and without blooming artifacts, as ground-truth (GT) reference images are generally unavailable in clinical practice. To address this, we developed a simulation framework to generate paired training data by inserting synthetic calcium into clinical cardiac images.

Twelve CCTA patient images from the PERFECTION study [11] were used as cardiac background images for synthetic data generation. We manually segmented the coronary vessels and inserted calcifications along the vessel paths. Using these inputs, separate simulations of the calcification-only image and the background-only image were generated via CatSim [12], [13], modeling the acquisition geometry of a GE HealthCare Revolution scanner. Prior to simulation, the background images were slightly sharpened to ensure that the reprojected and reconstructed outputs matched the spatial-frequency characteristics of the original scans. For the calcification simulation, focal spot oversampling was set to 5 in both directions and view oversampling was set to 4 to more accurately capture partial volume effects. The calcification material was modeled as a mixture of calcium hydroxyapatite, arterial tissue, and cholesterol with adjustable contrast levels [14]. The calcifications lie within contrast-enhanced vessels, which are approximated as a 15 mg/mL iodine-water solution. Accordingly, this iodine component was subtracted from the calcification material to account for the pre-existing background signal. In both simulations, a mono-energetic source at 64keV was used, as beam-hardening artifacts have negligible contributions to blooming when correction algorithms are applied [3].

To generate the image volume with blooming artifacts, the simulated background and calcification data were combined and reconstructed using filtered back-projection. Images were reconstructed at the same voxel size as the original clinical background (0.3 to 0.4 mm in plane and 0.625 mm cross plane) and were subsequently upsampled to a common finer $0.2 \times 0.2 \times 0.3125$ mm voxel. The corresponding no-bloomed counterpart was generated by reconstructing and upsampling the background-only scan, followed by inserting the GT calcification map directly into the reconstructed image. This process provided paired bloomed and no-bloomed volumes with consistent spatial resolution for supervised learning.

Using the proposed simulation framework, we generated 285 pairs of synthetic CCTA image volumes with size of 128×128×64 from 12 patient datasets. Each image pair included 12 realizations, created by combining three calcium contrast levels (40%, 50%, and 60% calcium hydroxyapatite by mass) with four noise levels (50, 100, 300 mAs, and noiseless). Fig. 1 shows two examples of the synthetic CCTA image pairs.

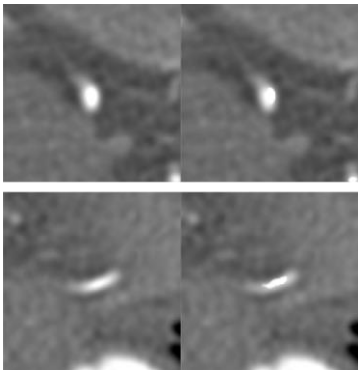


Fig. 1. Two examples of synthetic CCTA image pairs. Left: image with blooming artifacts. Right: image without blooming artifacts. WW/WL=1300/1150 HU (Air=0HU).

B. Deblooming network with spatial prior information

The proposed DL deblooming framework is illustrated in Fig. 2. To incorporate spatial prior information, we modified the U-Net++ network to include two input channels: (1) the original CCTA image containing blooming artifacts, and (2) an auxiliary calcium mask. Calcium masks are generated by intensity thresholding of the CCTA image to detect high-attenuation calcium. The calcium mask provides an estimate of the calcium region and may deviate from the true calcification boundary. By incorporating this auxiliary channel, the network is encouraged to focus its learning on regions with calcification while preserving anatomical structures in surrounding tissues.

The DL deblooming model is based on a modified 2D U-Net++ architecture. The network includes five resolution levels, with 32 feature channels at the first level and channel depth increasing with each down-sampling stage. The nested and dense skip-connections characteristic of U-Net++ enable multi-scale feature fusion and help capture fine anatomical details across the resolution hierarchy. The network predicts a debloomed image based on the two-channel input. It was trained end-to-end using the synthetic paired CCTA data.

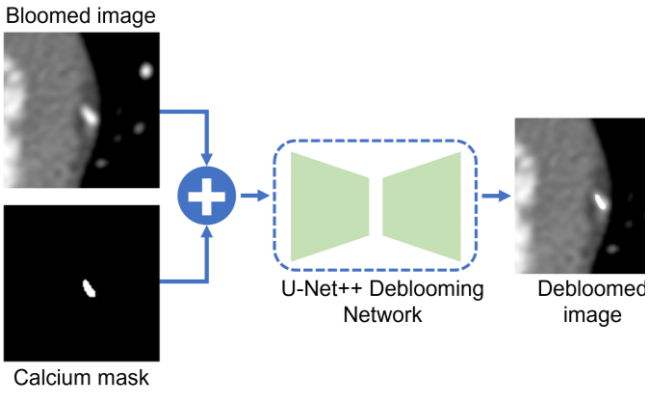


Fig. 2. Flowchart of the proposed deep learning deblooming framework.

The training loss function is a weighted combination of two L1 loss terms: an L1 loss computed over the entire image and an L1 loss restricted to the masked calcium region. The loss function is defined as:

$$L(Y_{pred}, Y_{gt}) = \|Y_{pred} - Y_{gt}\|_1 + \lambda\|M \odot (Y_{pred} - Y_{gt})\|_1, \quad (1)$$

where Y_{pred} and Y_{gt} denote the predicted and GT debloomed images, respectively. M is the calcium mask indicating the approximate calcium region, and $\odot$ denotes element-wise multiplication. The weighting factor λ assigns higher penalty to the error inside the calcium masked region, where the most significant blooming artifacts are expected to occur. This formulation encourages the network to predict accurate deblooming of calcium areas while preserving structures in surrounding areas.

C. Experiments

For model training, 2D slices of size 64×64 were extracted from the synthetic CCTA volumes. Synthetic images derived from 9 patients were used for training, 1 patient reserved for validation, and the remaining 2 patients held out for testing. Model parameters were optimized with the ADAM optimizer. The initial learning rate was set to 10^{-5} and exponentially decayed by a factor of 0.99 per epoch. Training was performed for 200 epochs. The final model was selected based on low validation loss while applying early stopping criteria to prevent overfitting.

To assess the contribution of the calcium mask, we performed an ablation study in which two models were trained under identical conditions except the input. The first model received only the CCTA image as input, whereas the second model used the two-channel input consisting of the CCTA image and the corresponding calcium mask. This comparison isolates the effect of the spatial prior provided by the calcium mask and quantifies its impact on deblooming performance.

D. Evaluation metrics

The goal of deblooming in CCTA is to improve the estimation of the true size and shape of calcified plaques. To evaluate the effectiveness of the proposed deblooming method, we employ the following quantitative metrics, all of which are applied in 3D volume:

- Dice score: The Dice similarity coefficient is computed between the calcium segmentations obtained from the GT and the DL-generated debloomed images. The calcium is segmented by applying an intensity threshold selected between the expected background level and max intensity within the calcification ROI. Higher Dice scores indicate more accurate recovery of the true calcium size and shape.
- Root mean square error (RMSE) within the ROI: RMSE is calculated between the GT and DL debloomed images within the calcification ROI to quantify the reconstruction accuracy of calcium shapes and intensities. Lower RMSE values indicate closer agreement with GT.
- RMSE over an annulus region: RMSE is further evaluated within an annulus region surrounding the calcium boundary to quantify reductions in blooming artifacts near the calcification edges. This metric specifically measures how well the model corrects the exaggerated spread of high-attenuation regions. Lower RMSE values indicate better deblooming performance.

III. Results

The proposed DL-based deblooming framework was evaluated on the held-out test dataset. A qualitative comparison between models trained with and without the calcium mask is presented in Fig. 3. The results show that the mask-guided DL model produces calcium boundaries that align more consistently with the GT.

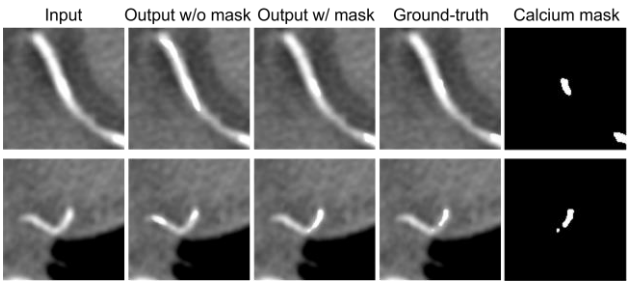


Fig. 3. Visual comparison of deblooming results generated by deep learning models trained without and with the calcium mask. WW/WL=1300/1150 HU (Air=0HU).

Quantitative evaluations are summarized in Fig. 4 and Fig. 5. The mean RMSE within the calcification ROI for input, DL output without calcium mask, and DL output

with calcium mask were 58.7±2.2, 58.7±17.0, and 46.6±9.3HU, respectively. The mask-guided DL model shows an improvement of 20.6% relative to both the input and the model without calcium mask guidance. For the Dice score evaluation, two intensity thresholds (0.1 and 0.45) relative to the maximum calcium intensity were selected to assess robustness across different segmentation sensitivities. In both threshold settings, the mask-guided model achieved higher mean Dice scores (0.1-threshold: .71±.15, .70±.14, .74±.12; 0.45-threshold: .59±.10, .59±.09, .68±.09 for input, without and with calcium mask, respectively), indicating more accurate recovery of calcium shape and size. The RMSE measured over the annulus surrounding the calcium boundary further illustrates the improvements of the mask-guided model. Specifically, the RMSE-annulus increased from 46.3±3.5 HU in the input images to 62.8±38.9 HU for the DL output without a guiding mask, whereas it decreased to 35.1±17.3 HU with guiding mask usage. Compared with the input and the DL output without a calcium mask, the mean RMSE-annulus values of the mask-guided DL output were reduced by 24.3% and 44.1%, respectively. Overall, the quantitative metrics consistently show that incorporating the calcium mask enhances the accuracy and spatial consistency of deblooming.

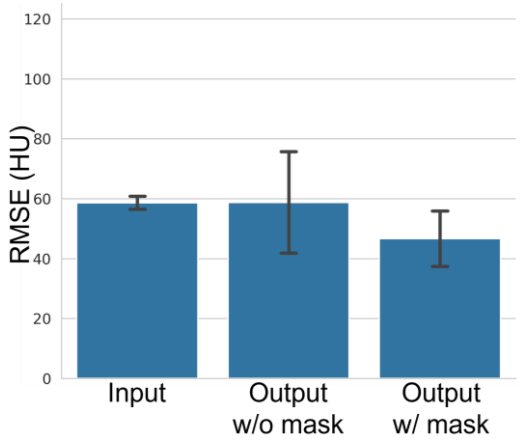


Fig. 4. RMSE within the calcification ROI. From left to right, the bars show RMSE between the GT and: (1) the input bloomed image, (2) the debloomed image from the model trained without calcium masks, and (3) the debloomed image from the model trained with calcium masks.

IV. Discussion and Conclusion

The proposed simulation framework plays a key role in enabling supervised learning in deblooming task, where artifact-free CCTA data is unavailable. By inserting synthetic calcium into clinical background images and separately simulating blooming and no-blooming images, we are able to generate realistic paired datasets with anatomical variations. This synthetic strategy provides a controllable framework for supervised-learning model development and can be extended for future applications, such as varying scanner geometries or introducing pathological conditions.

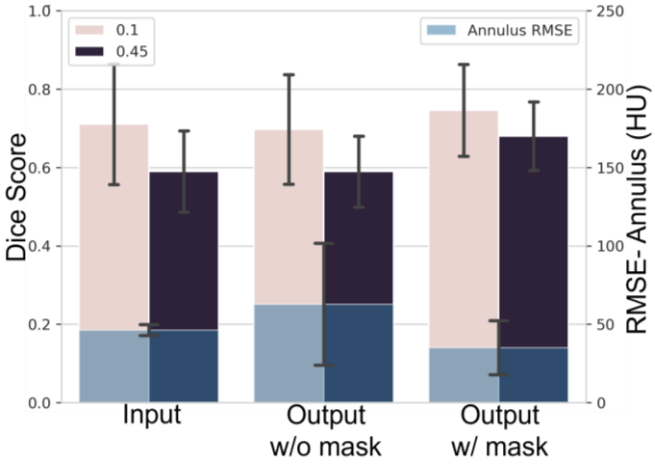


Fig. 5. Dice score and RMSE over the annulus region. The two Dice bars represent results obtained using two threshold levels (0.1 and 0.45) defined relative to the maximum calcium intensity.

Our results demonstrate the effectiveness of incorporating spatial prior information for DL-based deblooming in CCTA. The calcium-mask-based network produces more accurate debloomed results, particularly along calcium edges where blooming artifacts have the greatest impact on lumen interpretation.

In this work, we developed a DL-based deblooming framework that integrates a calcium mask and a modified U-Net++ architecture to improve lumen recovery in CCTA. Both qualitative and quantitative evaluations demonstrate that the mask-guided DL model provides more accurate and consistent deblooming. This approach has the potential to enhance lumen visualization and improve diagnostic confidence in patients with heavy coronary calcification. Future work will focus on expanding the simulation pipeline and validating the model performance on clinical datasets.

Acknowledgment

Research reported in this publication was supported by the NIH/NHLBI grant# R01HL151561. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

We thank Drs. Ying Fan, Jed Pack, and Jiayong Zhang for their assistance in designing the digital calcification phantom used in this study. We also appreciate Drs. Jhimli Mitra and Mingye Wu for their valuable technical discussions.

REFERENCES

[1] J. K. Min, L. J. Shaw, and D. S. Berman, "The Present State of Coronary Computed Tomography Angiography," *J. Am. Coll. Cardiol.*, vol. 55, no. 10, pp. 957–965, Mar. 2010, doi: 10.1016/j.jacc.2009.08.087.

[2] N. Vatsa, C. Faaborg-Andersen, T. Dong, M. J. Blaha, L. J. Shaw, and R. A. Quintana, "Coronary Atherosclerotic Plaque Burden Assessment by Computed Tomography and Its Clinical Implications," *Circ. Cardiovasc. Imaging*, vol. 17, no. 8, Aug. 2024, doi: 10.1161/CIRCIMAGING.123.016443.

[3] J. D. Pack, M. Xu, G. Wang, L. Baskaran, J. Min, and B. De Man, "Cardiac CT blooming artifacts: clinical significance, root causes and potential solutions," *Vis. Comput. Ind. Biomed. Art*, vol. 5, no. 1, p. 29, Dec. 2022, doi: 10.1186/s42492-022-00125-0.

[4] R. T. Yan *et al.*, "Predictors of Inaccurate Coronary Arterial Stenosis Assessment by CT Angiography," *JACC Cardiovasc. Imaging*, vol. 6, no. 9, pp. 963–972, Sep. 2013, doi: 10.1016/j.jcmg.2013.02.011.

[5] S. Zhang, D. C. Levin, E. J. Halpern, D. Fischman, M. Savage, and P. Walinsky, "Accuracy of MDCT in Assessing the Degree of Stenosis Caused by Calcified Coronary Artery Plaques," *Am. J. Roentgenol.*, vol. 191, no. 6, pp. 1676–1683, Dec. 2008, doi: 10.2214/AJR.07.4026.

[6] J. R. Weir-McCall *et al.*, "Effect of a calcium deblooming algorithm on accuracy of coronary computed tomography angiography," *J. Cardiovasc. Comput. Tomogr.*, vol. 14, no. 2, pp. 131–136, Mar. 2020, doi: 10.1016/j.jcct.2019.07.007.

[7] S. Yan *et al.*, "Calcium removal from cardiac ct images using deep convolutional neural network," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, Washington, DC: IEEE, Apr. 2018, pp. 466–469. doi: 10.1109/ISBI.2018.8363617.

[8] Z. Sun and C. K. C. Ng, "Artificial Intelligence (Enhanced Super-Resolution Generative Adversarial Network) for Calcium Deblooming in Coronary Computed Tomography Angiography: A Feasibility Study," *Diagnostics*, vol. 12, no. 4, p. 991, Apr. 2022, doi: 10.3390/diagnostics12040991.

[9] H. Zhao *et al.*, "Calcium deblooming in coronary computed tomography angiography via semantic-oriented generative adversarial network," *Comput. Med. Imaging Graph.*, vol. 122, p. 102515, Jun. 2025, doi: 10.1016/j.compmedimag.2025.102515.

[10] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, in Lecture Notes in Computer Science, vol. 11045. Cham: Springer International Publishing, 2018, pp. 3–11. doi: 10.1007/978-3-030-00889-5_1.

[11] G. Pontone *et al.*, "Stress Computed Tomography Perfusion Versus Fractional Flow Reserve CT Derived in Suspected Coronary Artery Disease," *JACC Cardiovasc. Imaging*, vol. 12, no. 8, pp. 1487–1497, Aug. 2019, doi: 10.1016/j.jcmg.2018.08.023.

[12] B. De Man *et al.*, "CatSim: a new computer assisted tomography simulation environment," presented at the Medical Imaging, J. Hsieh and M. J. Flynn, Eds., San Diego, CA, Mar. 2007, p. 65102G. doi: 10.1117/12.710713.

[13] M. Wu *et al.*, "XCIST—an open access x-ray/CT simulation toolkit," *Phys. Med. Biol.*, vol. 67, no. 19, p. 194002, Oct. 2022, doi: 10.1088/1361-6560/ac9174.

[14] T. J. Römer *et al.*, "Histopathology of Human Coronary Atherosclerosis by Quantifying Its Chemical Composition With Raman Spectroscopy," *Circulation*, vol. 97, no. 9, pp. 878–885, Mar. 1998, doi: 10.1161/01.CIR.97.9.878.

Cardiac Coronary CT Angiography Image Quality Enhancement with Multi-Vendor Generalizability using Deep Learning-based High-resolution and Multi-energy Imaging

Shaojie Chang, Maryam S. Sadeghian, Benjamin D. Wilson, Eric E. Williamson, Thomas A. Foley, Cynthia H. McCollough, Shuai Leng

***Abstract*— We propose CARNELIAN (CARdiac coroNary CT angiography image quality Enhancement using deep Learning-based hIgh-resolution and multi-energy imAgiNg), an AI-based framework designed to reduce noise, enhance spatial resolution, and mitigate blooming artifacts in coronary CT angiography (cCTA) across various conventional energy-integrating detector CT (EID-CT) systems. The framework comprises two cascaded deep learning subnetworks: one for spatial resolution enhancement and one for multi-energy image generation. CARNELIAN was trained using ultra-high-resolution, multi-energy photon-counting detector CT (PCD-CT) datasets acquired from eight patients on a NAEOTOM Alpha scanner (Siemens Healthineers). To ensure multi-vendor generalizability, protocol information from EID-CT systems—such as reconstruction kernels and tube voltage—was used to guide the selection of matching PCD-CT training data. This strategy enabled the training inputs to reflect the spatial resolution and effective energy characteristics of each vendor's EID-CT protocol. The high-resolution subnetwork was trained to map standard EID-equivalent reconstructions (e.g., Bv40) to ultra-high-resolution PCD-CT targets (e.g., Bv72), while the multi-energy subnetwork converted 70 keV inputs, approximating 120 kVp EID-CT acquisitions, to 100 keV virtual monoenergetic images (VMIs), thereby mitigating calcification-induced blooming artifacts. The trained CARNELIAN framework was applied to clinical cCTA datasets from three EID-CT scanners: SOMATOM Force and Definition Flash (Siemens Healthineers), and Aquilion ONE (Canon). Qualitative analysis demonstrated consistent improvements in vessel sharpness and artifact suppression. Quantitatively, mean CT number in the ascending aorta decreased (e.g., from 459 to 283 HU on Force) across scanners—reflecting the effective generation of 100 keV VMIs. These findings highlight CARNELIAN's generalizability and its potential to enhance imaging performance on existing EID-CT systems without hardware upgrades.**

***Index Terms*— Coronary CT Angiography, Photon-Counting Detector CT, Artificial Intelligence, Deep Learning, Blooming Artifact, Multi-Vendor Generalizability.**

I. Introduction

CORONARY CT angiography (cCTA) is a cornerstone in the non-invasive evaluation of coronary artery diseases, offering high sensitivity for detecting clinically significant stenosis [1]. However, image quality on conventional energy-integrating detector CT (EID-CT) systems is inherently limited by moderate spatial resolution, high image noise, and blooming artifacts from calcified plaques and metallic stents. These factors can hinder accurate lumen assessment and may reduce diagnostic confidence in clinical practice.

Photon-counting detector CT (PCD-CT) has recently emerged as a promising technology that addresses many of these challenges [2]. By directly converting x-ray photons into electrical signals and enabling energy binning, PCD-CT offers two key advantages over EID-CT: (1) ultra-high spatial resolution (UHR), achieved through smaller detector elements, which reduces partial volume effects and improves visualization of fine anatomical structures such as small coronary vessels [3]; and (2) intrinsic spectral imaging, in which each photon is individually counted and categorized by energy thresholds, enabling the reconstruction of virtual monoenergetic images (VMIs). High-energy VMIs (e.g., 100 keV) have been shown to reduce blooming artifacts from calcifications, improving lumen visualization and measurement accuracy [4, 5].

Despite these advantages, widespread clinical adoption of PCD-CT remains limited due to high cost and restricted availability. Most cCTA exams are performed using energy-integrating detectors, single-energy CT (SECT) acquisitions. While multi-energy capabilities enable VMI generation on EID-CT, these involve substantial cost and workflow

Corresponding author: Shuai Leng (E-mail: Leng.Shuai@mayo.edu)

Shaojie Chang, Maryam S. Sadeghian, Benjamin D. Wilson, Eric E. Williamson, Thomas A. Foley, Cynthia H. McCollough, and Shuai Leng are with the Department of Radiology, Mayo Clinic, Rochester, MN, 55905, USA.

This project was supported in part by the National Institutes of Health R01 EB028590, EB038746 and the American Heart Association Postdoctoral Fellowship (https://doi.org/10.58275/AHA.25POST1372972.pc.gr.227349).

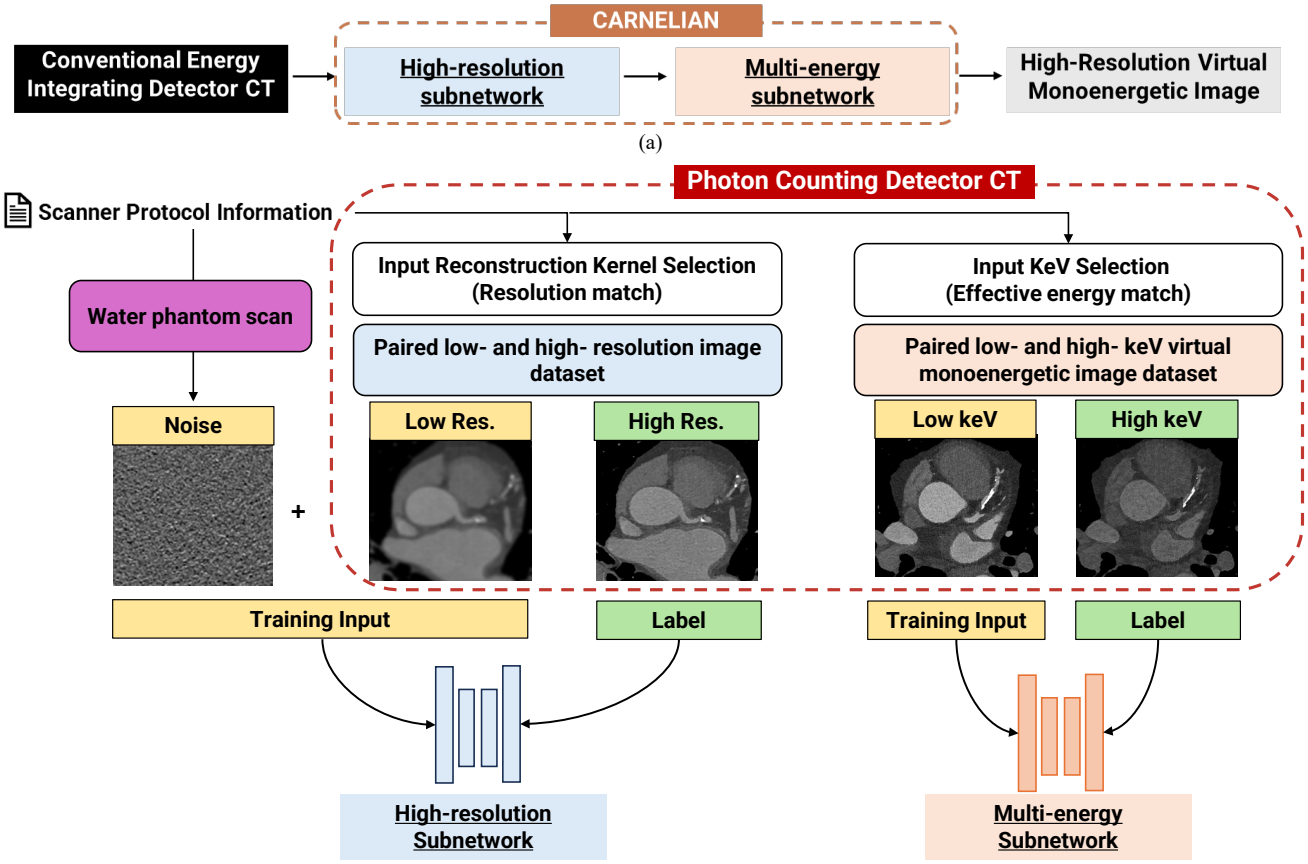


Figure 1. Overview of the CARNELIAN framework and training strategy. (a) The framework consists of two cascaded deep learning subnetworks: a high-resolution subnetwork that enhances spatial resolution and a multi-energy subnetwork that synthesizes higher-energy virtual monoenergetic images (VMIs) to reduce blooming artifacts. (b) Protocol-aware training strategy: EID-CT scanner parameters (e.g., reconstruction kernel and tube voltage) are used to select matched training data from photon-counting detector CT (PCD-CT). For the high-resolution subnetwork, input images reconstructed with EID-equivalent kernels (e.g., Bv40) are paired with higher-resolution labels (e.g., Bv72). A water phantom scan acquired on EID-CT is also added to the low-resolution inputs account for differences in noise texture between EID-CT and PCD-CT. For the multi-energy subnetwork, 70 keV PCD-CT VMIs (simulating 120 kVp EID-CT) serve as inputs, while 100 keV VMIs are used as labels for blooming artifact suppression.

complexity. Additionally, some multi-energy scanner designs sacrifice temporal or spatial resolution, which is critical in cCTA to minimize motion artifacts and preserve diagnostic image quality.

To bridge this gap, we introduce CARNELIAN (CARdiac coroNary CT angiography image quality Enhancement using deep Learning-based hIgh-resolution and multi-energy imAgiNg), a deep learning–based framework designed to improve image quality on conventional EID-CT systems. CARNELIAN consists of two cascaded neural networks for spatial resolution enhancement and multi-energy image synthesis. A key feature of this framework is its protocol-aware, multi-vendor generalizability: during training, protocol-specific parameters—such as reconstruction kernels and tube voltage—from various EID-CT systems are used to guide the selection of matched training data from PCD-CT. This enables CARNELIAN to learn scanner-specific characteristics and deliver consistent image quality enhancement across a broad range of commercial CT systems.

II. Methods and Materials

A. Overview of the CARNELIAN Workflow

Figure 1(a) illustrates the overall workflow of CARNELIAN, which consists of two cascaded convolutional neural networks (CNNs) designed to enhance cCTA images acquired on conventional EID-CT systems. The two subnetworks address distinct but complementary goals: (1) a high-resolution subnetwork that improves spatial resolution, and (2) a multi-energy subnetwork that generates higher-energy VMIs to reduce blooming artifacts caused by calcifications.

As shown in Figure 1(b), CARNELIAN incorporates scanner protocol information—specifically reconstruction kernel and x-ray tube voltage—to guide training data selection, enabling multi-vendor generalizability **using only PCD-CT datasets**, without the need to collect patient data for different scanner models and vendors. This protocol-aware strategy allows for the selection of training inputs and labels that replicate the spatial and spectral characteristics of different EID-CT systems.

B. Protocol-aware Data Selection Strategy

To achieve generalizability across multiple CT vendors using only PCD-CT data, we developed a protocol-aware data selection strategy that aligns the spatial resolution and spectral characteristics of different EID-CT systems with matched PCD-CT reconstructions. This strategy ensures that CARNELIAN learns scanner-specific image characteristics from a one-single dataset without requiring EID-CT images during training.

1) Reconstruction Kernel Matching: To match the spatial resolution characteristics of EID-CT images, reconstruction kernel sharpness was aligned between EID-CT and PCD-CT systems using a modulation transfer function (MTF)-based approach. A lookup table providing MTF curves for a wide range of CT scanners and reconstruction kernels across multiple vendors was used to identify kernels with comparable spatial frequency responses.

For each EID-CT scanner (e.g., Siemens Force, Siemens Flash, Canon Aquilion ONE), we selected its corresponding clinical cCTA kernel (e.g., Bv40 for Siemens, FC03 for Canon) and matched them to the most similar PCD-CT kernels based on MTF response. For example, the Bv40 kernel on Siemens EID-CT was matched to the Bv40 kernel on PCD-CT and used as the input, while the Bv72 kernel (a sharper reconstruction) served as the target for training the high-resolution subnetwork.

To further account for domain mismatch between PCD-CT and EID-CT, especially in noise characteristics, we acquired a water phantom scan using the same protocol on the target EID-CT system [6]. After subtracting the mean CT value from each input patch, noise patches extracted from the phantom scan were randomly added to the low-resolution PCD-CT inputs. This helped simulate EID-CT noise texture and improve generalization during inference.

2) Effective Energy Matching: Since conventional EID-CT acquisitions are performed at specific tube potentials (kVp), a calibration step is required to map each kVp to its corresponding effective energy keV level (Eff_E). This mapping can be derived from half-value layer measurements, phantom studies, scanner metadata, or standardized lookup tables [7]. Once the effective energy (Eff_E) keV is determined, PCD-CT–derived virtual monoenergetic images (VMIs) at that energy level can be used to simulate the EID-CT appearance.

In this study, most cCTA protocols used 120 kVp on EID-CT, corresponding to an effective energy of approximately 70 keV. Thus, 70 keV VMIs from PCD-CT were selected as inputs to the multi-energy subnetwork. The target outputs were 100 keV VMIs, which are known to reduce blooming artifacts while maintaining vascular contrast.

C. Network Training Details

Training was conducted using PCD-CT datasets from eight patients, acquired in ultra-high-resolution multi-energy mode on a NAEOTOM Alpha scanner (Siemens Healthineers) using our clinical cCTA protocol. All training data were selected using the protocol-aware strategy described in Section II.B to represent the spatial and spectral properties of multiple EID-CT systems.

For the high-resolution subnetwork, input patches of 128×128 pixels were reconstructed using EID-CT–matched kernels (e.g., Bv40), and targets used sharper PCD-CT kernels (e.g., Bv72). Acquired EID-CT noise from water phantom, as described in Section II.B, was added to input patches. For the multi-energy subnetwork, 70 keV VMIs, approximating 120 kVp EID-CT acquisitions, served as inputs, and 100 keV VMIs were used as targets to reduce blooming artifacts.

Both subnetworks employed a simplified 2D U-Net architecture with nine modules, each comprising convolution, batch normalization, and exponential linear unit (ELU) activation. Max pooling and transposed convolutions were used for down sampling and up sampling, with skip-connection concatenation to preserve structural fidelity.

The high-resolution subnetwork was trained with a weighted loss function: 0.4×MSE + 0.5×SSIM + 0.1×gradient loss. The multi-energy subnetwork used MSE loss. Both were trained for 100 epochs using the Adam optimizer, with an initial learning rate of 1×10^{-3}, decayed to 1×10^{-5} over time.

D. Experimental Evaluation

CARNELIAN was evaluated on clinical cCTA datasets acquired from three EID-CT systems:

- SOMATOM Force (Siemens Healthineers)
- SOMATOM Definition Flash (Siemens Healthineers)
- Aquilion ONE (Canon Medical Systems)

Test cases were selected to include calcified coronary artery segments, with routine cCTA protocols: 120 kVp tube voltage and vendor-specific reconstruction kernels (e.g., Bv40 for Siemens, FC03 for Canon). CARNELIAN was applied to the conventional EID-CT reconstructions to generate enhanced images.

Evaluation comprised both qualitative and quantitative assessments. Qualitative analysis included overall image quality, vessel sharpness, blooming artifact reduction, and line profile analysis. Quantitative measurements focused on CT number and noise level at matched anatomical locations before and after enhancement.

III. Results

A. Qualitative Image Enhancement

CARNELIAN demonstrated consistent visual improvements in cCTA images across all three EID-CT systems. Figure 2(a) shows representative examples from 3 different scanner models, including both axial images and zoomed-in views of calcified coronary segments. Compared to conventional EID-CT reconstructions, CARNELIAN-enhanced images exhibited

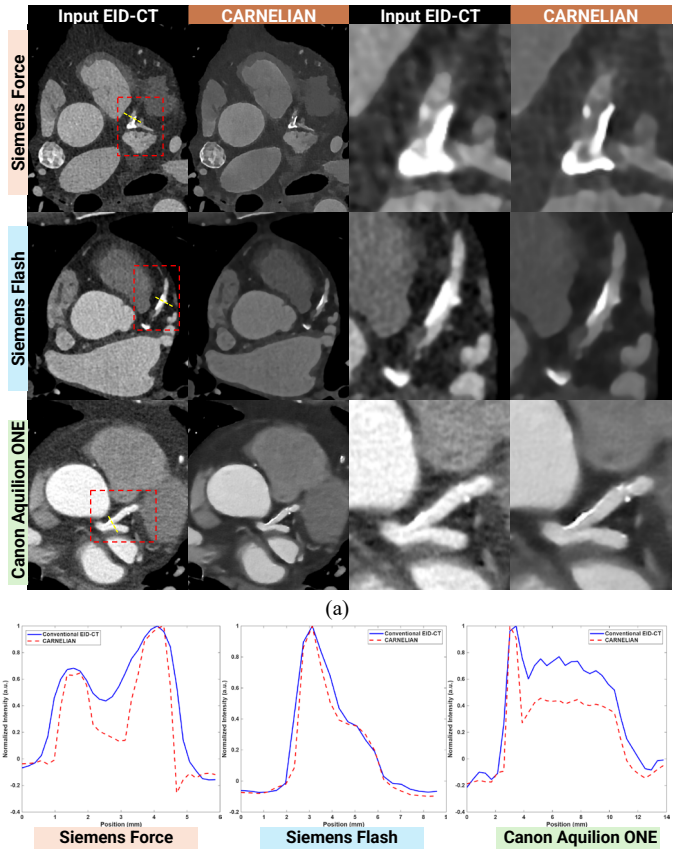


Figure 2. Qualitative comparison of cCTA images from three commercial EID-CT scanners before and after CARNELIAN enhancement. (a) The left two columns show conventional EID-CT images and corresponding CARNELIAN-enhanced outputs. The right two columns display magnified views of the red-boxed regions, highlighting calcified coronary segments. From top to bottom: SOMATOM Force, Definition Flash (both Siemens Healthineers), and Aquilion ONE (Canon Medical Systems). CARNELIAN consistently improves vessel wall sharpness and reduces blooming artifacts across all scanner platforms. All images are displayed using a window level/window width (WL/WW) of 300/1200 HU. (b) Normalized line profiles through calcified regions, corresponding to the yellow dashed lines in (a). Profiles are plotted for both conventional EID-CT (blue solid line) and CARNELIAN-enhanced images (red dashed line). CARNELIAN shows sharper edge transitions and reduced blooming, indicated by narrower peak widths and improved background separation.

sharper vessel wall delineation, reduced blooming from calcifications and improved visualization of the coronary lumen, particularly in heavily calcified areas.

To further illustrate these qualitative improvements, line profile analysis was performed across calcified coronary segments (yellow dashed lines) and are shown in Figure 2(b). Profiles extracted perpendicular to the vessel axis revealed narrower intensity distributions and steeper edge gradients in CARNELIAN outputs, consistent with improved spatial resolution and artifact reduction.

B. Quantitative Evaluation

To quantitatively assess CARNELIAN's impact on image quality, we measured two key metrics within the ascending aorta on all three EID-CT scanners: mean CT number (in Hounsfield units, HU) and image noise, defined as the standard deviation (SD) of pixel values within a uniform region of interest.

Table 1 summarizes the quantitative results. After applying CARNELIAN, the mean CT number consistently decreased across all scanners—reflecting the effect of generating 100 keV VMIs, which exhibit lower attenuation values than standard 120 kVp EID-CT images. Concurrently, image noise was substantially reduced, owing to the noise-aware training strategy described in Section II. For example, on the SOMATOM Force system, the mean CT value in the aorta decreased from 459.2 HU to 283.5 HU, while the standard deviation decreased from 47.3 to 21.6 HU. Similar trends were observed on the SOMATOM Definition Flash and the Canon Aquilion ONE.

Table I Major parameters for patient scanning and image reconstruction

Scanner	Mean HU (EID-CT)	Mean HU (CARNELIAN)	Noise (EID-CT)	Noise (CARNELIAN)
Siemens Force	459.19	283.45	47.28	21.58
Siemens Flash	501.10	242.13	35.67	13.10
Canon	361.15	263.82	22.61	8.75

IV. Conclusion

We presented CARNELIAN, a deep learning–based framework designed to enhance cCTA on conventional EID-CT systems by leveraging high-resolution, multi-energy imaging from PCD-CT. Through a protocol-aware training strategy, CARNELIAN generalizes across multiple vendors by using only PCD-CT data during training. Qualitative and quantitative evaluations demonstrated that CARNELIAN consistently improves vessel wall sharpness, reduces blooming artifacts, and lowers image noise across three different EID-CT systems. These results highlight CARNELIAN's potential to enhance image quality to existing EID-CT platforms without requiring hardware upgrades.

References

[1] M. Andrew and H. John, "The challenge of coronary calcium on coronary computed tomographic angiography (CCTA) scans: effect on interpretation and possible solutions," (in English), *Int J Cardiovasc Imaging,* vol. 31 Suppl 2, no. Suppl 2, pp. 145-57, Dec 2015, doi: 10.1007/s10554-015-0773-0.

[2] K. Rajendran *et al.*, "First Clinical Photon-counting Detector CT System: Technical Evaluation," (in English), *Radiology,* vol. 303, no. 1, pp. 130-138, Apr 2022, doi: 10.1148/radiol.212579.

[3] G. Fahrni *et al.*, "Quantification of Coronary Artery Stenosis in Very-High-Risk Patients Using Ultra-High Resolution Spectral Photon-Counting CT," *Invest Radiol,* vol. 60, no. 2, pp. 114-122, Feb 1 2025, doi: 10.1097/RLI.0000000000001109.

[4] S. Chang, E. K. Koons, H. Gong, J. E. Thorne, C. H. McCollough, and S. Leng, "Improving Stenosis

Assessment in Energy Integrating Detector CT via Learned Monoenergetic Imaging Capability," (in eng), *Proc SPIE Int Soc Opt Eng,* vol. 12925, Feb 2024, doi: 10.1117/12.3006468.

[5] S. Chang *et al.*, "Deep Learning–based Monoenergetic Imaging for Calcified Coronary Stenosis Assessment at Energy-integrating Detector CT," *Radiology: Cardiothoracic Imaging,* vol. 8, no. 2, p. e250230, 2026.

[6] E. K. Koons *et al.*, "Learned high resolution energy-integrating detector CT angiography: Harnessing the power of ultra-high-resolution photon counting detector CT," *Med Phys,* May 8 2025, doi: 10.1002/mp.17874.

[7] S. Chang *et al.*, "CT data harmonization via learned virtual monoenergetic imaging for cross-kV scan translation and radiomics reproducibility," *Medical physics,* vol. 53, no. 4, p. e70403, 2026.

An Accurate and Real-Time 3D/2D Coronary Artery Registration Method for Depth-Guided Coronary Intervention Navigation

Ruifeng Chen, Zhongwei Sun, Lei Song, Mengxuan Yuan, Jun Xiang, Yang Chen, Dong Liang, Yinsheng Li

***Abstract*—X-ray coronary angiographic (XCA) imaging using C-arm cone-beam CT systems is widely used for coronary intervention navigation. To address the two-dimensional (2D) nature of XCA limits, using 3D/2D coronary artery registration, centerlines with depth information derived from three-dimensional (3D) coronary computed tomography angiography (CCTA) can provide the necessary depth guidance and discriminate vessel overlapping. Existing tree-aligning feature-based methods suffer from error accumulation or time consuming challenges. To address these challenges, we propose an novel 3D/2D coronary artery registration method that eliminates the need for image-to-graph transformation during intraoperative pipeline. Numerical simulation studies and experimental human subject studies demonstrate the clinical feasibility of our approach, which achieves accurate and real-time 3D/2D registration at over 15 FPS with potential for depth-guided coronary intervention navigation.**

***Index Terms*—3D/2D Coronary Artery Registration, Depth-Guided Coronary Intervention Navigation, Deep Learning.**

I. Introduction

X-ray coronary angiographic (XCA) imaging using C-arm [1] is widely used for coronary intervention navigation. To address the two-dimensional (2D) nature of XCA limits, by utilizing 3D/2D coronary artery registration technique as shown in Fig. 1, centerlines with depth information derived from three-dimensional (3D) coronary computed tomography angiography (CCTA) can provide the necessary depth guidance and discriminate vessel overlapping.

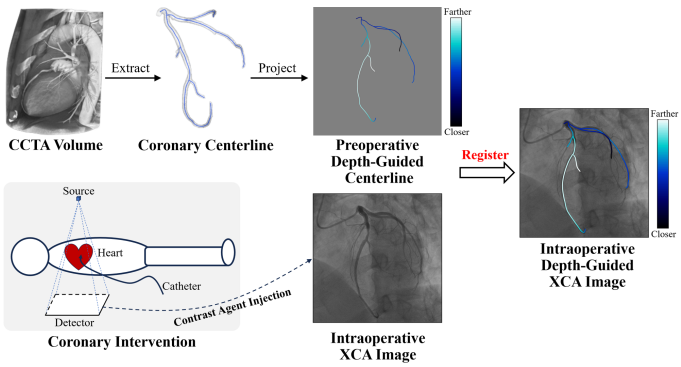


Fig. 1. The general workflow of depth-guided coronary intervention navigation. The 3D/2D coronary artery registration (in red color) is technical challenging.

However, 3D/2D coronary artery registration technique faces accuracy and speed challenges. Existing methods for 3D/2D coronary artery registration can be categorized into intensity-based and feature-based methods. Intensity-based methods [2] perform registration using image intensity, which are constrained by computationally inefficient and background interference. To address these limitations, feature-based methods use coronary centerline coordinates for registration. Among these, methods based on the 3D/2D coronary tree (a type of graph) alignment have demonstrated better registration accuracy [3]–[5], as vascular topology remains consistent across modalities and dimensions, making it a robust feature for 3D/2D registration.

They typically include preoperative and intraoperative pipelines. The preoperative pipeline is nearly error-free due to ample time for manual correction, while the intraoperative pipeline must perform complex frame-by-frame processing within milliseconds without compromising accuracy. Its difficulties includes 2D-tree cycle removal, 3D/2D tree inconsistencies and many-to-many 3D/2D node correspondences. To address these difficulties, Liu et al. [3] used a connectedness clustering method [6] to remove cycles, but the iterative optimization in centerline registration is computationally expensive. Wu et al. [4] proposed an end-to-end CAR-Net to simplify this process. However, generating topologically aligned centerline pairs requires a pre-alignment pipeline involving vessel mask skeletonization, cycle removal, root selection, tree construction, and topology matching. These hand-crafted steps introduce uncertainty and error accumulation. To mitigate this, Zhu et al. [5] introduced a Monte Carlo tree search (MCTS) framework that jointly handles tree construction and 3D/2D alignment, but it still contain a complex iterative optimization and cannot meet millisecond-level demands for intraoperative use. Tree-aligning feature-based methods contain image-to-graph transformation in the intraoperative pipeline, involving high information compression and requiring complex heuristic algorithms, thus hindering the ability to achieve both accuracy and speed simultaneously. In this work, we eliminates the image-to-graph modality transformation by achieving intensity-based 3D/2D tree alignment in centerline heatmaps in the intraoperative pipeline. Besides, we use graph-regularized convolutional framework to regularize coronary displacement fields. Through the proposed framework, the accurate and real-time 3D/2D coronary artery registration is realized for the depth-Guided coronary intervention navigation.

II. Methodology

Our framework's workflow (as shown in Fig. 2) consists of five steps: (i) Generate an XCA vessel mask from the XCA image. (ii) Project the preoperative CCTA-derived 3D vessel mask and tree into a DRR vessel mask and a depth-attached projected tree; split the tree into linear paths and convert them into centerline heatmaps. (iii) Feed the DRR vessel mask,

XCA vessel mask, and DRR centerline heatmaps into AlignNet to estimate XCA centerline heatmaps (Sec. II-A). (iv) Feed the DRR/XCA vessel masks, their centerline heatmaps, and linear path control points into RegisNet to estimate displacement fields for those control points (Sec. II-B). (v) Deform the original linear paths using the estimated displacements, recombine them into a tree, and color-map its depth information onto the XCA image.

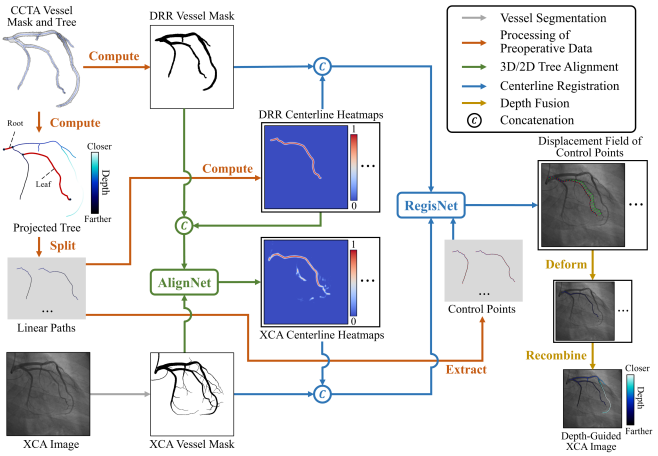


Fig. 2. The general workflow of our framework.

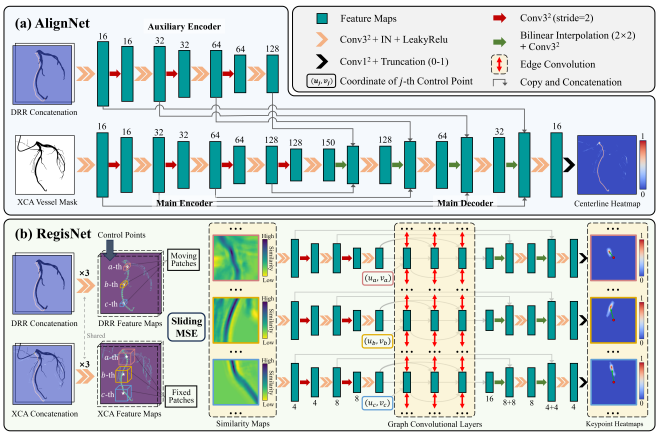


Fig. 3. The architecture of AlignNet and RegisNet. (a) AlignNet is designed for intensity-based 3D/2D tree alignment. It consists of an auxiliary encoder, a main encoder, and a main decoder. The inputs are the XCA vessel mask and the concatenation of the DRR vessel mask and the DRR centerline heatmap, and the output is the XCA heatmap. (b) RegisNet is designed for centerline registration with graph regularization. The inputs are control points, XCA concatenation, and DRR concatenation, and the outputs are keypoint heatmaps of control points. The displacement vectors are calculated from the center point to the weighted mean position of the heatmap intensities. The graph convolution layers play a role in graph regularization

A. AlignNet for Intensity-Based 3D/2D Tree Alignment

AlignNet is designed using a dual-encoder UNet architecture with an auxiliary encoder, a main encoder, and a main decoder. The auxiliary encoder extracts multi-resolution features from the DRR concatenation, while the main encoder does so from the XCA vessel mask. The main decoder predicts the centerline heatmap on the XCA vessel mask, progressively fusing features from both encoders during upsampling. The auxiliary encoder provides supervisory information, and the main encoder refines spatial details. Features span five resolutions, with spatial dimensions halved and channels increasing by $16 \rightarrow 32 \rightarrow 64 \rightarrow 128 \rightarrow 150$. Finally, the output centerline heatmap is obtained through the 1×1 convolution kernels ("Conv1^2") followed by the truncation of pixel values within range of 0 to 1.

Defining $\boldsymbol{\Theta}$ as the parameters of AlignNet $\mathcal{F}_{\boldsymbol{\Theta}}(\cdot)$, the optimal $\boldsymbol{\Theta}^*$ is obtained by solving the mean square error (MSE) optimization problem.

B. RegisNet for Centerline Registration with Graph Regularization

In the centerline registration step, to realize regularization of coronary displacement fields, RegisNet is designed as shown in Fig. 3 (b). It is modified from a graph-regularized convolutional framework [7] by redesigning three aspects:

(i) Topology definition. Rather than directly utilizing the tree topology, we split the coronary tree into a series of linear paths from root to leaves, and employ line topology as inputs.

(ii) Vertex location sampling. As the 3D coronary tree is projected and split into multiple linear paths, we sampled vertices (i.e., control points) started from the root node along the line at equal intervals.

(iii) Vertex feature extraction. A shared learnable CNN is introduced to extract feature maps from the DRR and XCA concatenation of the vessel mask and the centerline heatmap.

After modification, the graph regularization for coronary displacement fields is realized in graph convolution layers:

$$
\begin{aligned}
\mathbf{Y}_j &= \mathbf{X}_j \circ (u_j, v_j) \qquad (1)\\
\mathbf{X}_j^{(1)} &= \frac{1}{2} \sum_{n\in\{-1,1\}} \mathrm{CNN}\left(\left(\mathbf{Y}_{j+n} - \mathbf{Y}_j\right), \mathbf{Y}_j\right)\\
\mathbf{Y}_j^{(1)} &= \mathbf{X}_j^{(1)} \circ \mathbf{X}_j \circ (u_j, v_j)\\
\mathbf{X}_j^{(2)} &= \frac{1}{2} \sum_{n\in\{-1,1\}} \mathrm{CNN}\left(\left(\mathbf{Y}_{j+n}^{(1)} - \mathbf{Y}_j^{(1)}\right), \mathbf{Y}_j^{(1)}\right)\\
\mathbf{Y}_j^{(2)} &= \mathbf{X}_j^{(2)} \circ \mathbf{X}_j^{(1)} \circ \mathbf{X}_j \circ (u_j, v_j)\\
\mathbf{X}_j^{(3)} &= \frac{1}{2} \sum_{n\in\{-1,1\}} \mathrm{CNN}\left(\left(\mathbf{Y}_{j+n}^{(2)} - \mathbf{Y}_j^{(2)}\right), \mathbf{Y}_j^{(2)}\right),
\end{aligned}
$$

where the $\mathbf{X}_j$ denotes the j-th encoded similarity map, the $\circ$ denotes the concatenation operation, the (u_j, v_j) denotes the coordinate of the j-th control point, the CNN$(\cdot)$ denotes the convolution neural network, and the $\mathbf{X}_j^{(3)}$ denotes the output and then is decoded into a keypoint heatmap indicating the displacement vector from the center point to the weighted mean position of the heatmap intensities.

Defining $\boldsymbol{\Phi}$ as the parameters of RegisNet $\mathcal{G}_{\boldsymbol{\Phi}}(\cdot)$, the optimal $\boldsymbol{\Phi}^*$ is obtained by solving the MSE optimization problem.

III. Materials and Evaluation Methods

A. Numerical Simulation Studies with Ground Truth

7,200 paired CCTA-XCA samples, containing a total of 636,800 control points, categorized into 30,800 overlapped points and 606,000 non-overlapped points, were simulated from the ASOCA dataset using the simulated coronary deformations. Five-fold cross-validation was conducted for ablation studies of AlignNet and RegisNet.

B. Experimental CCTA-XCA Subjects Studies

CCTA volumes and XCA images from six patients were acquired from the Yangzhou Jiangdu People's Hospital. The CCTA volumes are acquired from the Philips iCT 256. The XCA images acquired from the Philips Allura-Xper system acquired at a frame rate of 15 frame per second (FPS). XCA images contain LCA and RCA sequences from different view angle. For each sequence of XCA images, the contrast agent-filled frames which exhibit complete coronary structures are selected. Approximately 8-16 images are selected from each sequence.

C. Metrics for Quantitative Performance Assessment

The mean target registration error (mTRE) was utilized to quantify the registration accuracy in millimeter (mm):

$$\text{mTRE} = \frac{1}{N_1} \sum_{j=0}^{N_1-1} w_j \cdot \Delta_j \cdot ||\hat{c}_j^{2D} - c_j^{2D}||_2, \tag{2}$$

where N_1 is the total number of points; $\hat{c}_j^{2D}$ and c_j^{2D} are the predicted and ground-truth 2D coordinates (in pixels) of the j-th control point; $w_j = \text{SOD}_j/\text{SDD}_j$ weights the error to normalize it across varying C-arm geometries; Δ_j is the pixel size (mm/pixel); and $\|\cdot\|_2$ denotes the L2 norm.

The mean frames per second (mFPS) was utilized to quantify the registration speed:

$$\text{mFPS} = \frac{N_2}{\sum_{s=0}^{N_2-1} t_s}, \tag{3}$$

where the N_2 denotes the total number of frames, and the t_s denotes the processing time of the s-th frame in unit of second.

IV. Results

A. Numerical simulation studies

1) Ablation studies on AlignNet: AlignNet performs 3D/2D tree alignment using centerline heatmaps (CHMs) as image-based centerline representations to eliminate intraoperative image-to-graph transformation. Ablation studies assess CHMs' role in alignment and AlignNet's ability to predict them accurately. As shown in Fig. 4: (1) Comparing the first, second, and fourth rows shows that without CHMs, displacement vectors fail to align with red centerlines, as high-similarity regions in similarity maps do not lie on topologically correct centerlines, demonstrating CHMs' effectiveness; (2) Comparing the second, third, and fourth rows shows that "The Proposed Technique" produces displacement fields similar to "w. True CHMs", confirming AlignNet's accurate CHM prediction.

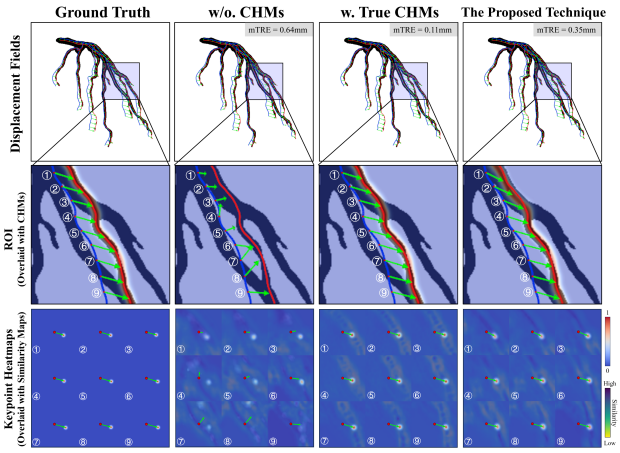


Fig. 4. Results generated from ablation studies on AlignNet include: without CHMs ("w/o. CHMs"), with ideal intensity-based alignment using ground truth CHMs ("w. True CHMs"), with AlignNet-predicted CHMs ("The Proposed Technique"), and ground truth. The blue centerlines refer to initial state, the red centerlines refer to ground truth centerlines, and the green arrows refer to estimated displacement vectors.

2) Ablation studies on RegisNet: RegisNet is designed for centerline registration, using graph convolutional layers (GCLs) to regularize coronary displacement fields. This study evaluates the regularization effect of GCLs. As shown in Fig. 5, without GCLs, displacement vectors show irregular patterns despite generally pointing toward the red centerlines. When combined with similarity maps, the keypoint heatmaps in "The Proposed Technique" are more focused than those in "w/o. GCLs".

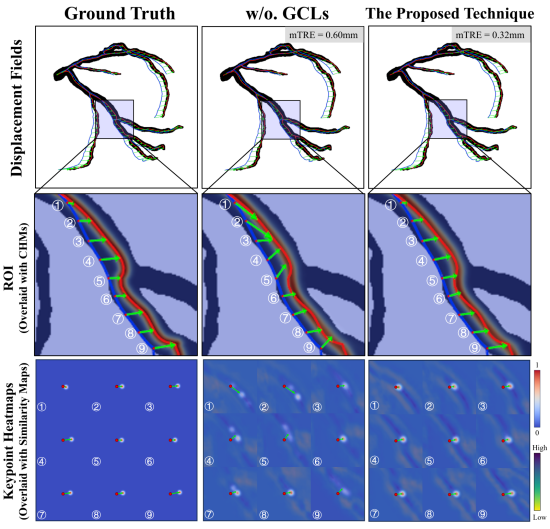


Fig. 5. Results generated from ablation studies on RegisNet including without GCLs ("w/o. GCLs"), with GCLs ("The Proposed Technique"), and ground truth.

3) Quantitative accuracy assessments and comparison: Fig. 6 shows TRE-based registration error comparisons. Wilcoxon signed-rank tests revealed significant improvements with our framework versus "w/o. CHMs" ($p < 0.05$) and "w/o. GCLs" ($p < 0.05$) for the "Total" category. The fact that removing GCLs increased errors for both "Overlapped"

and "Non-overlapped" points, confirms that GCLs regularize displacements across all control points. The fact that removing CHMs increased errors only for "Overlapped" points and leaving "Non-overlapped" errors unchanged, indicates that CHMs, as image-based centerline representations, aid vessel overlap identification in intensity-based 3D/2D tree alignment.

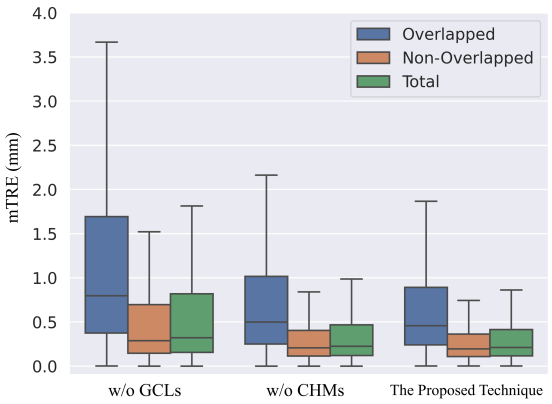


Fig. 6. The comparison of registration error quantified by mTRE using "w/o. GCLs", "w/o. CHMs" and "The Proposed Technique". The median [25th, 75th percentiles] of TRE of "w/o. GCLs", "w/o. CHMs" and "The Proposed Technique" results were measured for categories "Overlapped" , "Non-Overlapped" and "Total" respectively. A total of 636,800 testing control points were included in the statistics, categorized into 30,800 overlapped points and 606,000 non-overlapped points.

B. Experimental CCTA-XCA subjects studies

This study evaluates our framework's registration accuracy and speed against the baseline MCTSR method [5] on experimental subjects. As shown in Fig. 7, our framework achieves higher qualitative accuracy than MCTSR. Table I reports inference speed (mFPS) on four-view XCA sequences, which demonstrates that our framework processes at over 15 FPS (meeting clinical real-time requirements) and significantly outperforms MCTSR.

TABLE I
THE REGISTRATION SPEED (IN THE UNIT OF FRAMES-PER-SECOND) OF OUR FRAMEWORK COMPARED WITH THE BASELINE METHOD MCTSR. THE XCA SEQUENCES FROM FOUR VIEW ANGLES ARE EVALUATED USING THE METRIC MFPS.

	View1	View2	View3	View4
MCTSR	0.210	0.204	0.219	0.205
The Proposed Technique	15.45	15.45	17.42	16.84

V. CONCLUSION

In this work, we propose an accurate and real-time 3D/2D coronary artery registration method, which eliminates the image-to-graph modality transformation in the traditional intraoperative pipeline of tree-aligning feature-based methods. Specifically, AlignNet is designed for the intensity-based 3D/2D tree alignment, and RegisNet is designed for the centerline registration with graph regularization. Numerical simulation containing qualitative and quantitative analysis of large-scale CCTA-XCA samples has demonstrated that the proposed technique achieves high accuracy (<0.35mm). Validation of clinical feasibility using CCTA-XCA testing samples from six patients has demonstrated that the proposed technique has the potential to achieves real-time registration (>15 FPS) with high qualitative accuracy.

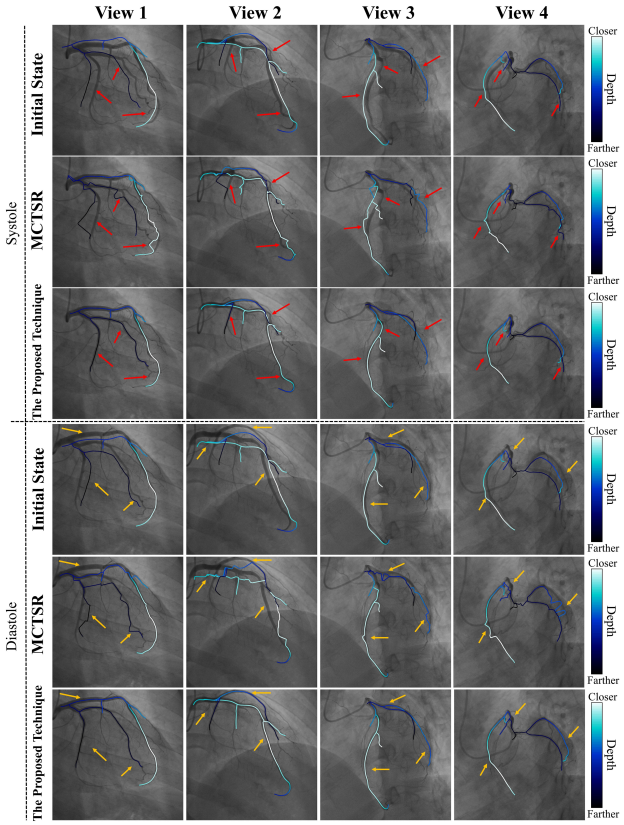


Fig. 7. Results from experimental CCTA-XCA subjects studies to evaluate the registration accuracy compare with the baseline method MCTSR [5]. The red and yellow arrows highlight the prone-error locations in registration. Dynamic "The Proposed Technique" results are available on https://figshare.com/s/b7a4903b17358e560270

REFERENCES

[1] T. Chua, D. Sarkar, E. Leatham, and M. Webb-Peploe, "Coronary angiography using a portable super-c-arm imaging system in a district general hospital," *International journal of cardiology*, vol. 94, no. 2-3, pp. 173–176, 2004.

[2] C. Gatta, S. Balocco, V. Martin-Yuste, R. Leta, and P. Radeva, "Non-rigid multi-modal registration of coronary arteries using siftflow," in *Pattern Recognition and Image Analysis: 5th Iberian Conference, IbPRIA 2011, Las Palmas de Gran Canaria, Spain, June 8-10, 2011. Proceedings 5*. Springer, 2011, pp. 159–166.

[3] S. Liu, P. Liu, Z. Li, Y. Zhang, W. Li, and X. Tang, "A 3d/2d registration of the coronary arteries based on tree topology consistency matching," *Biomedical Signal Processing and Control*, vol. 38, pp. 191–199, 2017.

[4] W. Wu, J. Zhang, W. Peng, H. Xie, S. Zhang, and L. Gu, "Car-net: a deep learning-based deformation model for 3d/2d coronary artery registration," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2715–2727, 2022.

[5] J. Zhu, C. Wang, Y. Zhang, M. Zhan, W. Zhao, S. Teng, L. Lu, and G.-J. Teng, "3d/2d vessel registration based on monte carlo tree search and manifold regularization," *IEEE Transactions on Medical Imaging*, 2023.

[6] Z. Li, Y. Zhang, G. Liu, H. Shao, W. Li, and X. Tang, "A robust coronary artery identification and centerline extraction method in angiographies," *Biomedical Signal Processing and Control*, vol. 16, pp. 1–8, 2015.

[7] L. Hansen and M. P. Heinrich, "Graphregnet: Deep graph regularisation networks on sparse keypoints for dense registration of 3d lung cts," *IEEE Transactions on Medical Imaging*, vol. 40, no. 9, pp. 2246–2257, 2021.

Diffusion Model-Based Motion Compensation for Coronary CT Angiography in Dual Wide-Detector CT

Zhiyang Fu, Jiao Tian, Xi Zhang, Yifu Mao, Wenjing Cao, Guotao Quan, and Wenying Wang

***Abstract*—Motion artifacts remain a primary challenge in coronary computed tomography angiography (CCTA). Conventional multi-phase acquisition for cardiac motion estimation and compensation increases radiation dose and limits reconstruction temporal resolution. While single-phase motion compensation methods have been proposed, their performances are constrained by the tradeoff between temporal and spatial fidelity. Specifically, improving the reconstruction temporal resolution by using narrower data windows introduces severe limited-angle artifacts that often render the images unsuitable for reliable motion estimation. In this work, we present an advanced single-phase CCTA reconstruction method that leverages the orthogonal geometry of dual wide-coverage CT combined with a conditioned generative diffusion model. Single-phase data are partitioned into multiple subphases to achieve high temporal resolution. Subphasic images reconstructed from the dual wide-coverage data yield more manageable artifact patterns than single-source systems. The conditional diffusion model maps these subphase images to motion-compensated single-phase images. During inference, the model iteratively generates anatomically consistent, non-motion-blurred images for robust motion field estimation. The final motion-compensated image is then produced by warping the subphase images using these derived fields. Clinical evaluation demonstrated that this framework effectively suppressed the limited-angle smearing and sharpens coronary artery boundaries, leading to superior motion artifact reduction compared to conventional multi-phase motion compensation methods. This synergy between dual wide-coverage geometry and generative artificial intelligence enables high-fidelity cardiac motion blurring reduction, offering a promising path for reliable single-phase CCTA that secures accurate diagnosis.**

***Index Terms*—Motion Compensation, Coronary Computed Tomography Angiography (CCTA), Diffusion model**

INTRODUCTION

In cardiac computed tomography (CT), the primary challenge lies in imaging small coronary arteries that are moving rapidly. The lumen diameter of a normal coronary artery ranges from 1.9 mm to 4.5 mm [1] and has a mean velocity of 35-75 mm/s [2]. Achieving diagnostic image quality places extreme demands on the spatial and temporal resolution of the CT system and the reconstruction process. Conventional Filtered Back-Projection (FBP) methods require at least a short scan (180° plus fan angle) to obtain artifact-free images, yielding a temporal resolution roughly equal to half the gantry rotation time—often over 100 microseconds. Within this short-scan window, coronary artery can still move several millimeters. This inherent movement introduces motion artifacts including blurring, blooming, shading, and distortion. Such degradations can all reduce diagnostic confidence in assessing luminal stenosis level and plaque characterizations.

To mitigate cardiac motion, clinical coronary CT angiography (CCTA) scans often utilize *retrospective* acquisition with ECG gating, allowing for the selection of the optimal cardiac phase from a continuous dataset [3], albeit at a higher radiation dose. In view of dose optimization, *prospective* ECG gating is used to trigger acquisition only during specific windows; however, this can be challenging in patients with high or irregular heart rates. Further improved temporal resolution is often highly demanded and favored. Hardware advancements have fundamentally pushed these limits: *dual-source* CT systems effectively halve acquisition time with two X-ray sources and detectors mounted approximately 90° to each other, while *wide-coverage detectors* (up to 16 cm) enable volumetric imaging of the entire heart in a single rotation. Traditional dual-source systems are limited by detector coverage, often necessitating helical scanning and therefore still relying heavily on retrospective ECG-gating for the whole heart coverage. On the other hand, traditional wide-coverage CT faces a limitation in temporal resolution, which can compromise image quality in patients with high heartrates. By combining dual-source geometry with wide-coverage hardware, a dual wide-coverage system can achieve superior temporal resolution while collecting the data of the whole cardiac region of interest within a single heartbeat.

In addition to new hardware design, advanced reconstruction algorithms can also improve temporal resolution by reducing the required angular coverage to approximately 120°, much smaller than a short scan. Iterative model-based algorithms with properly designed regularizations based on histograms [4] or prior images [5] have been proposed to suppress the limited-angle artifacts associated with the ill-posed problem. Cardiac motion can also be modeled via motion vector fields (MVF) [6], which are estimated from and applied to image sequences acquired at consecutive phases. These techniques, collectively known as motion estimation and compensation (ME/MC), often utilize partial angle reconstructions (PARs) to estimate the MVF. PAR-based methods [7], [8], [9] rely on the data consistency condition that two conjugate PARs separated by 180° should be consistent after motion compensation. As a consequence, while PAR-based methods can capture motion orthogonal to the central view direction, they fail to detect motion along the parallel axis. Dual-source CT systems may

Z. Fu and W. Wang are with the CT R&D Department, United Imaging Healthcare North America, Houston, TX 77054 USA. (e-mail: zhiyang.fu@united-imaging.com, wenying.wang@united-imaging.com). J. Tian, X. Zhang, Y. Mao, W. Cao, and G. Quan are with the CT RPA Department, Shanghai United Imaging Healthcare, Shanghai, China.

alleviate the parallel motion issue, as there are two pairs of conjugate PARs capturing motion components in orthogonal directions.

Recently, deep-learning-based approaches have been proposed to overcome the computational and accuracy limits of traditional ME/MC methods. Models like Deep PAMoCo [3] replace the lengthy iterative optimization of MVFs with convolutional neural networks [10], [11] and further improved motion compensation for CCTA with severe motion artifacts. Other models like diffusion models [12] directly reduce motion artifacts in the image domain by learning from pairs of motion corrupted and motion-free simulated data [13].

In this work, we investigate the potential of a diffusion model to further improve the temporal resolution of a dual wide-coverage CT system, where we aim to use deep-learning-based method to prepare PAR images for subsequent MVF estimation and, finally, motion compensation. The proposed method was evaluated in clinical data.

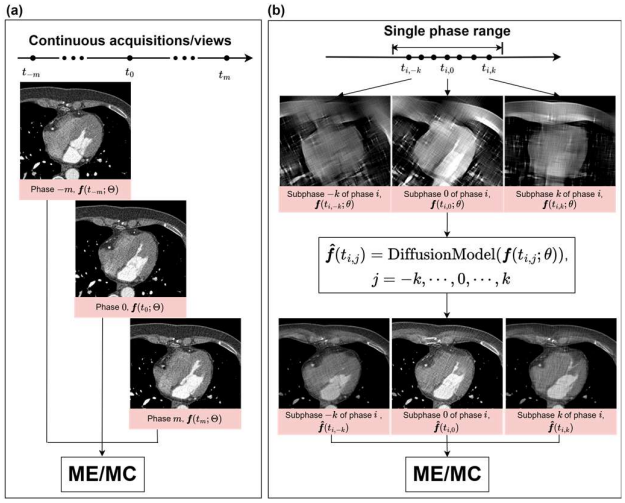


Figure 1. Comparison of cardiac CT motion estimation and compensation (ME/MC) workflows. (a) Conventional multi-phase workflow. In this illustration, $2m+1$ evenly distributed and overlapping cardiac phases are reconstructed, each requiring a full short-scan data range. These multi-phase images are used to estimate motion vector field (MVF) relative to a target central phase. (b) The proposed *single-phase* workflow. A single-phase dataset that typically covers a short-scan range is first partitioned into $2k+1$ subphases. Each sub-phase data are reconstructed using weighted FBP. These reconstructed sub-phase images are processed by a trained diffusion model before undergoing ME/MC.

METHODS

A. Motion Compensation using Single-phase Data

Conventional cardiac motion compensation typically requires acquisition of multi-phase data, resulting in increased radiation dose. As shown in Fig. 1(a), a continuously acquired cardiac CT dataset is partitioned into $2m+1$ phases, where central phase 0 is the target phase, and $t_{i=-m,\cdots,m}$ denotes the central time point of the i^{th} phase. Each image, $\boldsymbol{f}(t_i;\Theta)$, is reconstructed using FBP over a short-scan range Θ and subsequently used for ME/MC. In contrast, our proposed motion compensation scheme in Fig. 1(b) utilizes only a single-phase dataset with short-scan angular coverage. The single-phase data is evenly divided into $2k+1$ subphases centered at $t_{i,j=-k,\cdots,k}$, and $\boldsymbol{f}(t_{i,j};\theta)$ denotes the j-th subphase image of phase i obtained via weighted FBP. Each subphase data yields angular coverage smaller than the short-scan range ($\theta < \Theta$), resulting in limited-angle artifacts (smearing, shading, non-homogeneity) but shorter absolute data acquisition time. Note that our approach leverages the orthogonal geometry of dual wide-coverage systems. This configuration mitigates the severe smearing and information loss typical of single-source limited-angle reconstructions. An image-to-image translation diffusion model maps the subphase image $\boldsymbol{f}(t_{i,j};\theta)$ to the corresponding motion-free, artifact-free image with matched center time point, $t_{i,j}$. The outputs of diffusion model, $\hat{\boldsymbol{f}}(t_{i,j})$, undergo ME/MC to estimate the motion vector fields (MVF) $\boldsymbol{v}(t_{i,j})$ towards the center subphase, $\hat{\boldsymbol{f}}(t_{i,0})$. Suppose $\boldsymbol{r}$ denotes the spatial coordinate, the final motion-corrected image, $\boldsymbol{g}(\boldsymbol{r},t_i)$, can be obtained as follows,

$$\boldsymbol{g}(\boldsymbol{r},t_i)=\sum_{j=-k}^{k}\boldsymbol{f}(\boldsymbol{r}+\boldsymbol{v}(\boldsymbol{r},t_{i,j}),t_{i,j};\theta).$$

B. Diffusion Model Training and Inference

The diffusion model is designed to process the sub-phase images for reliable subsequent motion estimation, including limited-angle artifact reduction, coronary artery extraction, and MVF estimation. We thus pair each sub-phase image with its *simultaneous* label image, ideally with both motion and image artifacts suppressed. A reasonable training label can be motion-compensated single-phase image. Fig. 2(a) summarizes the training process of a noise-prediction diffusion model. The label image $\boldsymbol{x}$ and Gaussian noise image $\boldsymbol{\epsilon}_n$ (subscript n is used to distinguish diffusion step from the time variable t) are weighted to form a noisy image $\boldsymbol{x}_n$, that is,

$$\boldsymbol{x}_n=\alpha_n\boldsymbol{x}+\sigma_n\boldsymbol{\epsilon}_n,\boldsymbol{\epsilon}_n\in\mathcal{N}(\boldsymbol{0},\boldsymbol{1}),$$

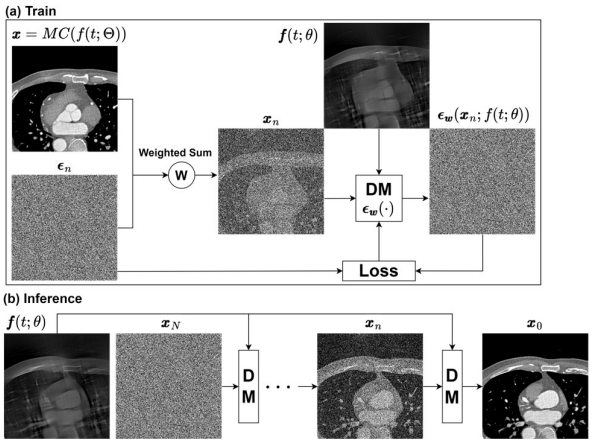


Figure 2. (a) Training and (b) inference of the proposed conditional diffusion model. The model $\boldsymbol{\epsilon}_{\boldsymbol{w}}(\cdot)$ is trained to predict the noise $\boldsymbol{\epsilon}_n$ added to the target label $\boldsymbol{x}$, where $\boldsymbol{x}$ represents the motion compensated reconstruction from short-scan data (angular coverage Θ). The model is conditioned on the subphase image denoted as $\boldsymbol{f}(t;\theta),\theta\ll\Theta$. The parameters of the model, $\boldsymbol{w}$, are optimized by minimizing the loss between the randomly sampled noise $\boldsymbol{\epsilon}_n$ and the predicted noise $\boldsymbol{\epsilon}_{\boldsymbol{w}}(\boldsymbol{x}_n;\boldsymbol{f}(t;\theta))$. During inference, the trained diffusion model iteratively denoises the initial Gaussian noise image $\boldsymbol{x}_N$ and generates images conditioned on the subphase image $\boldsymbol{f}(t;\theta)$ at each step.

where (α_n, σ_n) are a pre-defined noise schedule [12]. The model $\boldsymbol{\epsilon_w}(\cdot)$ also takes the subphase image $\boldsymbol{f}(t;\theta)$ as a condition to predict the added noise. The loss between the predicted noise and the added noise is calculated to optimize the weights of the model. Fig. 2 (b) shows the iterative sampling process of our diffusion model. The model gradually denoises initial Gaussian noise over N steps, conditioned on the subspace image $\boldsymbol{f}(t;\theta)$ at each step, to generate the final cardiac image.

C. *Experimental Setup*

The study utilized five clinical cardiac multi-phase datasets acquired on the uCT SiriuX dual wide-coverage CT scanner (United Imaging Healthcare, Shanghai, China), with institutional review board approval for research use. Three datasets were dedicated to training, and the remaining two were reserved for inference and performance evaluation. Each of the three datasets were reconstructed at every 5% cardiac cycle, yielding 21 phases per dataset. This augments the size of the training dataset, diversifies the training data with varying start angles of the gantry, and captures the full motion states of coronary arteries in a cardiac cycle. The single phase was further divided into multiple overlapping subphases, where each subphase image was effectively reconstructed from 75° of data (37.5° per source) using weighted FBP [8].

Our diffusion model was implemented in PyTorch using a 2D UNet architecture with attention modules at the bottom [14], [15]. The depth, input feature size, and total parameters of the deployed UNet are 7, 64, and 34 million, respectively. An initial learning rate of 10^{-4} was used with Adam optimizer. The learning rate decayed by 0.33 for every 66 epoch until 300 epochs were reached. The inputs and targets of network were normalized to a range of [-1,1] assuming the original dynamic range was [-1000, 3000] HU. The network was trained on an NVIDIA H20 (144 GB memory) GPU with a batch size of 40. The network inference required 16 steps and took about 0.23 seconds per image.

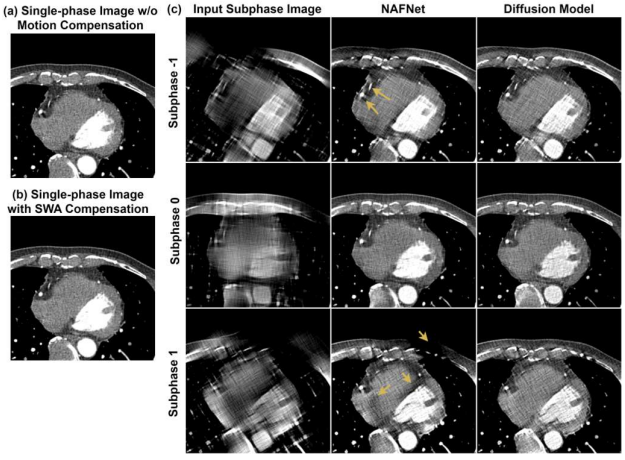


Figure 3. Comparison of subphase image between the proposed diffusion model and a state-of-the-art convolutional neural network (NAFNet) . (a) Single-phase reconstruction displaying half-moon motion artifacts in the RCA. (b) Motion artifacts suppressed using the multi-phase motion compensation algorithm named SWA. (c) Three consecutive subphase images of the single phase. Each subphase image, severely suffering from limited-angle artifacts, underwent the trained NAFNet and diffusion model, respectively. The proposed diffusion model better suppresses these artifacts. The display window levels and widths are -450 HU /700 HU for input subphase images and 100 HU/700 HU for all other images.

For the performance evaluation of limited-angle artifact suppression, the model was compared against a non-linear activation free network (NAFNet [16]), which has demonstrated competitive performance in limited-angle CT reconstruction [17]. We followed the same network configurations as [16] except that we dropped the residual learning since the dynamical ranges of the inputs and the targets are not in the same scale. We adopted an input feature size of 32, resulting in 29 million network parameters that are comparable to that of the proposed diffusion model. The NAFNet and diffusion model were visually evaluated in a clinical test case. To evaluate the impact of our diffusion model on motion compensation, we include a multi-phase algorithm named "Subphasic Warp and Add" (SWA) [18], which estimates MVF from multi-phase images and applies it to subphasic image. In contrast, we used subphase images processed with the diffusion model, replacing the conventional multi-phase inputs for MVF estimation. The rest were kept the same: The subphase images were warped by the estimated three-dimensional MVF and summed up to yield the final motion-compensated reconstruction. The performance of the SWA approach without and with diffusion model were assessed in the 5th clinical case.

RESULTS

Figure 3 shows single-phase images reconstructed from short-scan data (a) without and (b) with the multi-phase motion

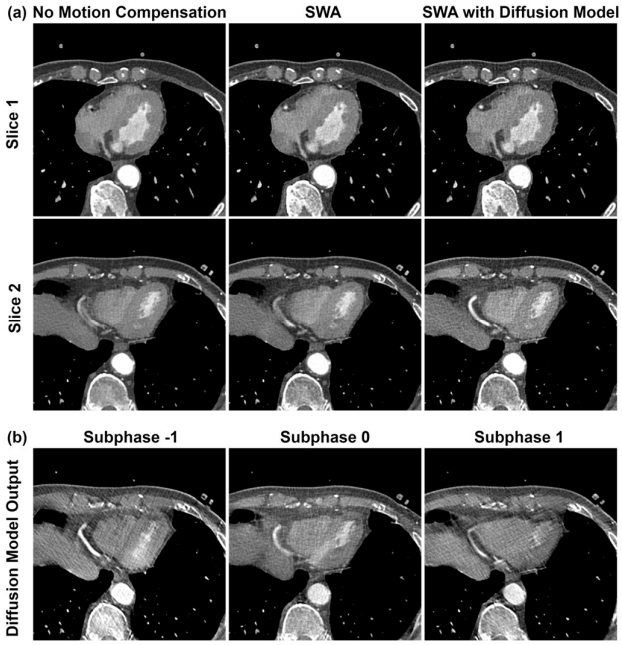


Figure 4. (a) Motion compensation performance without and with diffusion model on a test case at 60% phase cycle. Diffusion model clearly helps resolve the severe motion artifacts associated with the right coronary artery (RCA) in slice 2. (b) We show the three consecutive subphase images of slice 2 processed with the trained diffusion model. These images visibly depict the rapid in-plane deformations and through-plane motions of RCA through a short time interval. The display window level and width are 100 HU/700 HU.

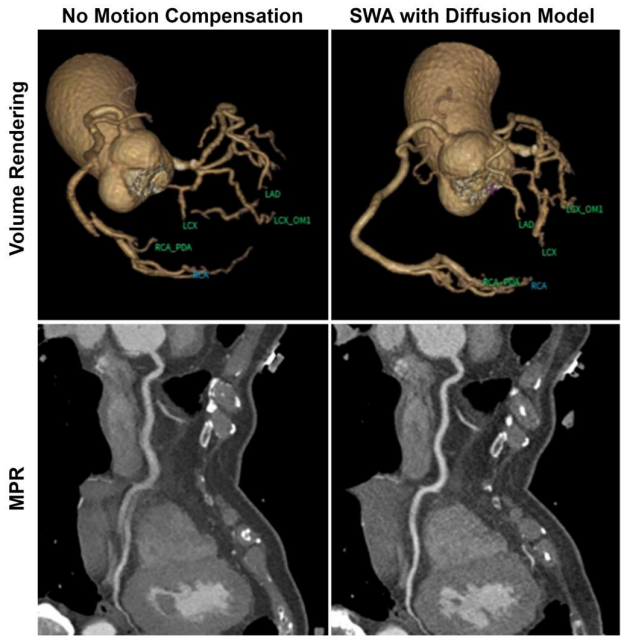


Figure 5. Follow-up image analysis of the test case in Fig. 4. Both the volume rendering and the curved MPR illustrate the disconnected RCA affected by motion was fully motion compensated with the help of diffusion model.

compensation algorithm, SWA [18]. Fig. 3(c) illustrates three consecutive subphases of this single phase. Because these subphase images are reconstructed from only 37.5° of data per source, they are severely degraded by limited-angle artifacts, appearing as cross shaped, shading, and blooming. While NAFNet and the proposed diffusion model both can faithfully recover the central subphase (subphase 0) with mild underlying motions, NAFNet fails to suppress the cross-shaped artifact around the right coronary artery (RCA) in subphase -1 and the shading artifacts in subphase 1 (indicated by yellow arrows). The proposed diffusion model better suppresses these limited-angle artifacts in each subphase.

Figure 4 shows the subsequent cardiac motion compensation results on another clinical dual wide-coverage data. In Slice 1, the mild motion artifacts associated with the coronary artery are successfully suppressed by SWA without or with diffusion model. In Slice 2, the severe motion blurring of the RCA is only *fully* resolved by SWA with diffusion model, showing sharp boundary delineation and high contrast of the RCA. This is a direct result of the inherent high temporal resolution provided by the subphase images, as further evidenced by Fig. 4(b). The three subphase images processed with the diffusion model clearly capture the three transient states of the RCA and reveal strong in-plane deformation and through-plane motion within this single phase.

A follow-up image analysis of the cardiac case in Figure 5 presents the motion affected RCA through volume rendering and curved multi-planar reconstruction (MPR). SWA combined with the diffusion model resolves the motion blurring and provides high-fidelity visualization of the RCA.

Conclusion

This work has demonstrated that integrating an image-to-image diffusion model into a dual wide-coverage CT workflow can improve motion compensation performance in clinical cardiac imaging. By enhancing limited-angle artifacts corrupted yet high temporal resolution subphase reconstructions, the diffusion model provides a higher-fidelity basis for motion field estimation compared to convolutional neural networks.

References

[1] J. T. Dodge, B. G. Brown, E. L. Bolson, and H. T. Dodge, "Lumen diameter of normal human coronary arteries. Influence of age, sex, anatomic variation, and left ventricular hypertrophy or dilation.," *Circulation*, vol. 86, no. 1, pp. 232–246, Jul. 1992.

[2] G. Shechter, J. R. Resar, and E. R. McVeigh, "Displacement and velocity of the coronary arteries: cardiac and respiratory motion," *IEEE Trans. Med. Imaging*, vol. 25, no. 3, pp. 369–375, Mar. 2006.

[3] J. Maier *et al.*, "Deep learning-based coronary artery motion estimation and compensation for short-scan cardiac CT," *Med. Phys.*, vol. 48, no. 7, pp. 3559–3571, Jul. 2021.

[4] H. Schöndube, T. Allmendinger, K. Stierstorfer, H. Bruder, and T. Flohr, "Evaluation of a novel CT image reconstruction algorithm with enhanced temporal resolution," presented at the SPIE Medical Imaging, Florida, Mar. 2011, p. 79611N.

[5] J. Tang, J. Hsieh, and G. Chen, "Temporal resolution improvement in cardiac CT using PICCS (TRI-PICCS): Performance studies," *Med. Phys.*, vol. 37, no. 8, pp. 4377–4388, Aug. 2010.

[6] C. Rohkohl, H. Bruder, K. Stierstorfer, and T. Flohr, "Improving best-phase image quality in cardiac CT by motion correction with MAM optimization," *Med. Phys.*, vol. 40, no. 3, p. 031901, Mar. 2013.

[7] S. Kim, Y. Chang, and J. B. Ra, "Cardiac motion correction based on partial angle reconstructed images in x-ray CT," *Med. Phys.*, vol. 42, no. 5, pp. 2560–2571, May 2015.

[8] J. Hahn *et al.*, "Motion compensation in the region of the coronary arteries based on partial angle reconstructions from short-scan CT data," *Med. Phys.*, vol. 44, no. 11, pp. 5795–5813, Nov. 2017.

[9] S. Kim, Y. Chang, and J. B. Ra, "Cardiac Image Reconstruction via Nonlinear Motion Correction Based on Partial Angle Reconstructed Images," *IEEE Trans. Med. Imaging*, vol. 36, no. 5, pp. 1151–1161, May 2017.

[10] T. Lossau (Née Elss) *et al.*, "Motion estimation and correction in cardiac CT angiography images using convolutional neural networks," *Comput. Med. Imaging Graph.*, vol. 76, p. 101640, Sep. 2019.

[11] D. Yao *et al.*, "Deep Learning-Based Cardiac CT Coronary Motion Correction Method with Temporal Weight Adjustment: Clinical Data Evaluation," *J. Imaging Inform. Med.*, Sep. 2025.

[12] Z. Fu *et al.*, "Motion Correction for Coronary Artery in Short-scan Cardiac CT using a Conditional Diffusion Model," in *Proc. 18th Int. Mtg. Fully 3D Image Reconstruction Radiol. Nucl. Med.*, Shanghai, China, May 2025, pp. 69–72.

[13] S. Jung, S. Lee, B. Jeon, Y. Jang, and H.-J. Chang, "Deep Learning Cross-Phase Style Transfer for Motion Artifact Correction in Coronary Computed Tomography Angiography," *IEEE Access*, vol. 8, pp. 81849–81863, 2020

[14] P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2021, pp. 8780–8794.

[15] T. Salimans and J. Ho, "Progressive Distillation for Fast Sampling of Diffusion Models," in *International Conference on Learning Representations*, 2022.

[16] L. Chen, X. Chu, X. Zhang, and J. Sun, "Simple Baselines for Image Restoration," in *Computer Vision – ECCV 2022*, vol. 13667, in Lecture Notes in Computer Science, Cham: Springer Nature Switzerland, 2022, pp. 17–33.

[17] J. Guo *et al.*, "RN-SDEs: Limited-Angle CT Reconstruction with Residual Null-Space Diffusion Stochastic Differential Equations," Nov. 22, 2025, *arXiv*: arXiv:2409.13930.

[18] R. Bhagalia, J. D. Pack, J. V. Miller, and M. Iatrou, "Nonrigid registration-based coronary artery motion correction for cardiac computed tomography," *Med. Phys.*, vol. 39, no. 7Part1, pp. 4245–4254, Jun. 2012.

NERVE: A Near-Real-Time Three-Dimensional Reconstruction Approach for Interventional Devices and Surrounding Vessels Using a Monoplane C-arm Cone-Beam CT

Yang Liu, Zexi Sun, Bingshuai Zhao, Jun Xiang, Yikun Zhang, Yang Chen, Dong Liang, Yinsheng Li

***Abstract*—Real-time three-dimensional (3D) reconstruction of interventional devices, such as stents and embolization coils, together with surrounding vessels, using a monoplane C-arm cone-beam CT (CBCT) has emerged as a critical capability in modern neurovascular interventions. Due to the slow gantry rotation speed of current CBCT systems, real-time 3D reconstruction of interventional devices and surrounding vessels has never been accomplished. In this work, we propose a framework for near-real-time 3D reconstruction of interventional devices and vasculature, termed NERVE, which integrates deep structure extraction (DSE) model and deep backprojection filtration (DBF) model together to accomplish 3D reconstruction of interventional devices and vasculature using data acquired at three projection views over a 60° angular range. DSE is trained to extract targets from acquired projection data by removing irrelevant anatomical background. DBF is trained to transform direction backprojection image volumes to final 3D image volumes accordingly. DBF is designed to learn prior regularization that captures the distribution-level characteristics of ideal interventional devices and vasculature to eliminating limited view angle artifacts and high-attenuation-induced artifacts. Extensive numerical simulations (1000 samples for testing) validate the feasibility and effectiveness of NERVE. NERVE explicitly leverages subject-specific measured data, ensuring near-real-time 3D reconstruction with high fidelity, potentially facilitating image-guided interventional navigation for patients with neurovascular diseases.**

***Index Terms*—Interventional Devices; Real-Time Three-Dimensional Reconstruction; Deep Learning**

I. Introduction

Real-time three-dimensional (3D) reconstruction of interventional devices, such as stents and embolization coils, together with surrounding vascular structures, using C-arm cone-beam computed tomography (CBCT) has emerged as a critical capability in modern neurovascular interventions. Accurate 3D reconstruction of device morphology, deployment status, spatial relationship and interactions to surrounding vascular structures during the procedure enables interventionists to move beyond limited two-dimensional fluoroscopic projections, which often suffer from vessel overlap, foreshortening, and ambiguous depth discrimination. Importantly, the ability to obtain such real-time 3D information intraprocedurally integrates seamlessly into the interventional workflow without the need for patient transfer, enabling immediate feedback, decision-making, and corrective actions.

Despite its strong clinical potential, real-time 3D reconstruction of interventional devices and surrounding vascular structures using C-arm CBCT has never been accomplished. A fundamental limitation preventing real-time reconstruction arises from the slow gantry rotation speed of current C-arm CBCT systems, typically requiring several seconds to complete one rotation. Moreover, interventional devices are composed of high-attenuation metallic materials, which exacerbate beam hardening, photon starvation, and streak artifacts in filtered backprojection (FBP)-based reconstructions.

Improving 3D reconstruction performance without increasing the gantry rotation speed, one needs to reconstruct interventional devices and surrounding vascular structures using data acquired over a very limited angular range, spanning only a small fraction of a full rotation. By drastically reducing the angular acquisition window, the temporal window of each reconstruction can be shortened to sub-second timescales, thereby approaching real-time 3D reconstruction. However, such limited view angle reconstruction severely violates the data sufficiency conditions assumed by conventional CBCT reconstruction algorithms, rendering most existing techniques inadequate due to pronounced artifacts and loss of structural fidelity.

Although substantial efforts have been made to limited view angle reconstruction, existing approaches either fail to meet real-time requirements due to the computational burden of iterative optimization in model-based iterative reconstruction (MBIR) framework or self-supervision-based deep learning framework, or do not adequately exploit the distinctive characteristics of interventional devices and the neurovascular system. Consequently, no prior work has yet demonstrated real-time three-dimensional reconstruction of *both interventional devices and the surrounding vascular structures*, although several approaches have been proposed to reconstruct interventional devices only using data acquired at several projection view angles [1], [2].

In this work, we propose a deep learning-based approach for near-real-time 3D reconstruction of interventional devices and vasculature, termed NERVE, which integrates deep structure extraction (DSE) model and deep backprojection filtration (DBF) model together to accomplish 3D reconstruction of interventional devices and vasculature using data acquired at three projection views over a 60° angular range. DSE is trained to extract interventional devices and vasculature from acquired projection data by removing irrelevant anatomical background, facilitating the following reconstruction task. Direction backprojection (DBP) image volumes are then formed via unfiltered backprojection algorithm using three projection views over a 60° angular range. DBF is trained to transform DBP image volumes to final 3D image volumes of interventional devices and vasculature accordingly. Inspired by the classical backprojection and filtration algorithm [3] and recent generalizations leveraging deep learning framework [4],

DBF is designed to learn prior regularization that captures the distribution-level characteristics of ideal interventional devices and vasculature to eliminating limited view angle artifacts and high-attenuation-induced artifacts. Moreover, NERVE explicitly leverages subject-specific measured data, ensuring near-real-time 3D reconstruction with high fidelity, therefore facilitating image-guided interventional navigation for patients with neurovascular diseases.

II. NERVE: A Near-Real-Time Three-Dimensional Reconstruction Approach for Interventional Devices and Surrounding Vessels Using Deep Learning

A. General Pipeline

The input consists of three sets of stent and vessel projections sampled over a 60° range. In practice, the stent is first imaged after deployment, followed by vessel projection acquisition using a high-attenuation contrast agent. As illustrated in Fig. 1, the proposed pipeline decouples the stent and vessel processing.

Within the DSE module, input projections are initially segmented via DSE-SEG component to generate binary masks, which are then multiplied by the original images to isolate target features. These features are processed by DSE-MAP component to predict corresponding depth maps. Then, a direct back-projection operator reconstruct these maps. Subsequently, the Deep Backprojection Filtration (DBF) module employs a neural network to refine the preliminary images. Finally, the high-quality filtered results for both stent and vessel are integrated for joint 3D rendering.

B. Deep Structure Extraction (DSE)

The Deep Structure Extraction (DSE) module is designed to extract the target components from the projections. The DSE module consists of two primary components: DSE-SEG, which is responsible for segmenting the target mask from the projections, and DSE-MAP, which handles the post-processing of the target mask.

1) DSE-SEG: The DSE-SEG component extracts the target masks from 2D projection images by modeling the task as a binary classification problem:

$$f_\theta : \mathbb{R}^{H\times W\times C} \to \{0,1\}^{H\times W\times K}. \tag{1}$$

where f_θ denotes the DSE-SEG network, H, W, C, K denotes image width, height, channels, classes respectively.

The architecture of DSE-SEG follows a 2D UNet-like structure comprising 6 layers. Each layer contains two identical blocks consisting of a 3×3 convolution, an instance normalization, and a LeakyReLU activation function to introduce nonlinearity. Downsampling is achieved via strided convolutions, while upsampling is performed using transposed convolutions. The total number of feature map channels is defined as $\min(512, 32\times 2^n)$ for $n = 0, 1, \ldots, 6$.

During training, an SGD optimizer with a momentum of 0.99 is employed. The initial learning rate is 0.01 and follows a poly decay strategy defined as $\text{lr} = (1 - \frac{epoch}{max_epoch})^{0.9}$, with a total of 1000 epochs, a patch size of 128×160, and a batch size of 65. The total loss function is a linear combination of Dice loss and Cross-Entropy loss:

$$L_{total} = \lambda_1 L_{dice} + \lambda_2 L_{CE}. \tag{2}$$

where λ_1 and λ_2 are both set to 1.

2) DSE-MAP: The DSE-MAP component transfers the binary mask output by DSE-SEG into depth map. Following the acquisition of the clean masks via DSE-SEG, it is noted that these masks lack spatial depth information. Directly inputting these masks into the DBF module for back-projection would likely result in significant deviations from the ground truth.

Inspired by [2], the DSE-MAP component is designed to transform original projection values into the intersection length between the X-ray and the region of object. The intersection length represents the path length of the X-ray through the interior of target (stent and vessel). Specifically, defining the indicator function $\mathbf{1}_O(\mathbf{x})$ as:

$$\mathbf{1}_O(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in O \\ 0, & \mathbf{x} \notin O \end{cases} \tag{3}$$

where O is the volume occupied by the target in 3D voxel space. Thus the intersection length for a ray $r(t)$ is expressed as:

$$L = \int \mathbf{1}_O(r(t))dt. \tag{4}$$

However, it is worth noting that projection pixels not containing the target do not contribute to the intersection length. Therefore, it is of necessity to isolate the target regions and set the remaining pixels to zero, which is done by DSE-SEG component.

The network of DSE-MAP is similar to the U-Net design in [1], trained using Mean Absolute Error (MAE) loss.

C. Deep Backprojection Filtration (DBF)

The Deep Backprojection Filtration (DBF) module aims to transform the reconstructed image generated by the back-projection of three DSE-processed projections into clean masks (voxel-wise reconstruction).

In this study, a Region of Interest (ROI) imaging strategy is adopted, thus the projection data are inherently truncated. Reconstruction using FDK will introduces significant truncation artifacts during the filtering stage. Therefore, the filtering process is skipped and only back-projection is performed to generate the reconstructed image I_{recon}:

$$I_{recon} = \mathcal{B}\{P_{DSE}\} \tag{5}$$

where P_{DSE} represents the clean projection output from the DSE module and $\mathcal{B}$ denotes the back-projection operator.

Structurally, the DBF module adopts a 3D U-Net architecture analogous to DSE-SEG, comprising five downsampling stages. The total number of feature map channels is defined as $\min(320, 32\times 2^n)$ for $n = 0, 1, \ldots, 6$. Each layer consists of blocks comprising a $3\times3\times3$ convolution layer, an instance normalization layer, and a LeakyReLU activation function. The upsampling and downsampling mechanisms

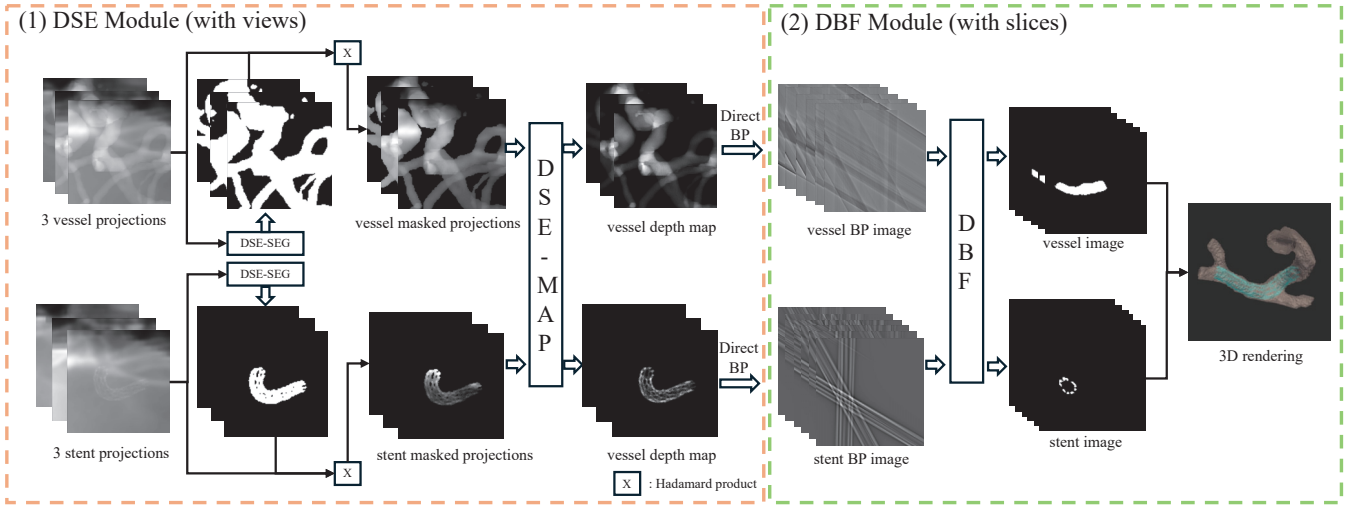


Fig. 1. Overview of the proposed pipeline. Input projections are processed through Deep Structure Extraction (**DSE**) and Deep Backprojection Filtration (**DBF**) modules to generate precise 3D masks, followed by joint spatial rendering for final visualization.

remain consistent with those of DSE-SEG. The model is trained on 128^3 ROI patches using a batch size of 3, while other hyperparameters and the loss function remain identical to DSE-SEG.

D. Simulations

1) Vessel Data Set: Due to the scarcity of public CTA datasets, MRA data were used for training, as vascular morphology and localization remain modality-independent. Intensity mapping was applied to align MRA voxel values with CT attenuation ranges. The COSTA [5] dataset was used for training and testing, with TopBrain [6] and SMILE-UHURA [7] as external test sets. All images were resampled to 0.2 mm isotropic resolution. A synthetic skull layer was added to the COSTA data to improve realism. Spherical ROIs of 128 voxels in diameter were randomly sampled within vascular labels.

2) Stent Data Set: Ten categories of stent phantoms were generated using analytical formulas and 3D model conversion, and divided into training and testing sets at an 8:2 ratio. Stent deformation was simulated via synchronous bending and radial scaling based on vascular centerlines and local diameters. The stents were inserted into phantom images according to vascular labels, with intensities randomly assigned from a predefined metal attenuation range.

3) Training Data Set for DSE: For each phantom, 20 CBCT projections were acquired along a circular trajectory. Ground truth projections were generated by binarized projection of 3D stent and vessel labels. Attenuation coefficients were randomly perturbed, and correlated Poisson noise and scattering effects were added in the projection domain.

4) Training Data Set for DBF: 3D images reconstructed by direct BP from three DSE projections were used as training inputs, with original 3D vessel and stent volumes as ground truth. Three independent projection sequences were extracted from each ROI under a half-scan trajectory.

E. Implementation Details

The DSE-SEG module is implemented using a 2D nnU-Net[8], while the DSE-MAP component adopts a 2D U-Net architecture [2]. The DBF module is implemented using a 3D nnU-Net. All experiments were conducted on an NVIDIA RTX A6000 with Python 3.10, PyTorch 2.3.1, and CUDA 12.2.

III. Result

Performance was evaluated under a unified DSE framework with different DBF module algorithms. Quantitative comparisons with U-Net and U-Net++ for stent and vessel segmentation are summarized in Table I. The test set was augmented by extracting ten ROIs per image, yielding 460, 240, 200, and 100 samples from COSTA, TopBrain-MRA, TopBrain-CT, and SMILE-UHURA, respectively, with all stent data combined. The proposed method consistently outperforms U-Net and U-Net++ in Dice and HD95 on both vessel and stent datasets, achieving notably lower HD95 and improved overlap. These results indicate strong robustness and generalizability across datasets and anatomical structures.

Figure 2 presents segmentation results for three representative stent cases. The predictions show stable and coherent stent structures across cases, with only minor voxel-level variations that do not affect the overall spatial configuration, as illustrated by the volume renderings. Given the clinical importance of spatial localization and 3D topology, the proposed method shows high qualitative consistency with the ground truth, as evidenced by the volume renderings.

Figure 3 provides a qualitative evaluation of vessel reconstruction. Compared with FDK, which suffers from severe streak artifacts, the proposed method more faithfully recovers vascular topology. The 3D renderings and error maps show accurate reconstruction of complex branches, with only minor errors in distal microvessels, indicating robust 3D spatial recovery.

In clinical diagnosis, accurate stent–vessel spatial topology is essential. Fig. 4 shows 3D visualizations of three repre-

TABLE I
COMPARISON OF RECONSTRUCTION PERFORMANCE (MEAN VALUE ± STANDARD DEVIATION). VESSEL RESULTS ARE REPORTED PER DATA SET, WHILE STENT RESULTS REPRESENT THE AVERAGE ACROSS ALL STENT-BEARING CASES. A TOTAL OF 1000 TESTING SAMPLES HAVE BEEN USED IN THE EVALUATION.

Method	Metrics	Vessel (Dataset-specific)				Stent
		Costa	TopBrain_MR	TopBrain_CT	SMILE-UHURA	all datasets
UNet	Dice ↑	0.661±0.261	0.623±0.259	0.664±0.265	0.210±0.257	0.547±0.250
	HD95 ↓	19.80±21.46	22.47±18.62	22.37±19.54	52.46±25.44	9.039±16.43
UNet++	Dice ↑	0.691±0.259	0.643±0.282	0.651±0.333	0.168±0.260	0.587±0.280
	HD95 ↓	15.62±16.13	20.06±16.47	20.07±19.78	42.58±20.40	15.27±29.53
NERVE	Dice ↑	**0.934±0.021**	**0.942±0.022**	**0.944±0.016**	**0.872±0.097**	**0.822±0.040**
	HD95 ↓	**1.813±1.036**	**1.709±1.506**	**1.649±2.159**	**5.029±5.864**	**1.032±0.130**

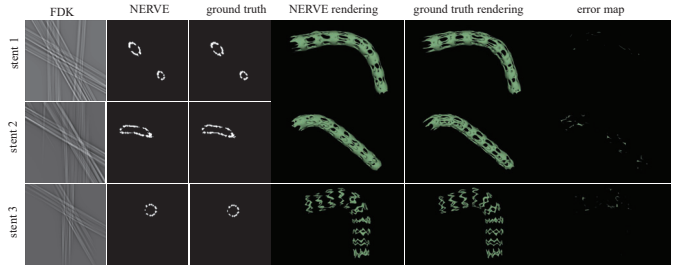


Fig. 2. Comparative stent reconstruction results. Left to right: FDK reconstruction; NERVE predictions alongside ground truth; and 3D rendered visualizations of the NERVE predicted versus ground truth.

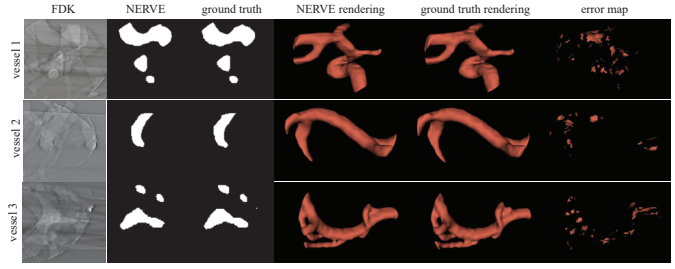


Fig. 3. Vessel reconstruction results. (Left to right) FDK reconstruction; NERVE prediction alongside ground truth; 3D volume renderings; and the spatial error distribution map between NERVE prediction and ground truth.

sentative cases, where the predicted stent and vessel surfaces exhibit strong spatial alignment across all axes, with minimal stent voxels outside the vessel wall. Despite minor voxel-level discrepancies, the predictions maintain high geometric consistency, indicating accurate localization and 3D morphology.

IV. CONCLUSION

In this work, we propose a framework for near-real-time 3D reconstruction of interventional devices and vasculature, termed NERVE, to accomplish 3D reconstruction of both interventional devices and vasculature using data acquired at three projection views over a 60° angular range. NERVE ensures near-real-time 3D reconstruction with high fidelity, potentially facilitating image-guided interventional navigation for patients with neurovascular diseases.

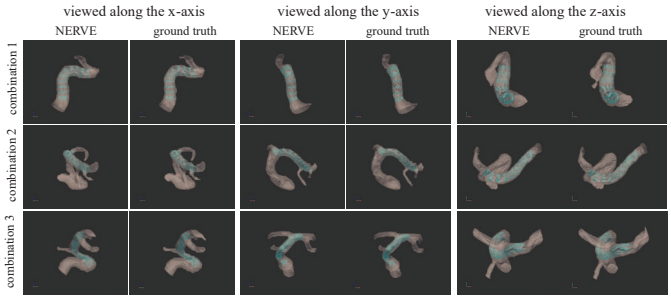


Fig. 4. 3D spatial alignment of stents and vessels. Representative samples (1–3) are displayed across three orthogonal axes (X, Y, Z) to compare model predictions with ground truth. Translucent grey surfaces represent the vessel lumen, while cyan meshes denote the implanted stents.

REFERENCES

[1] T. Vöth, T. König, E. Eulig, M. Knaup, V. Wiesmann, K. Hörndler, and M. Kachelrieß, "3d reconstruction of stents and guidewires in an anthropomorphic phantom from three x-ray projections," in *7th International Conference on Image Formation in X-Ray Computed Tomography*, vol. 12304. SPIE, 2022, pp. 223–229.

[2] T. Vöth, T. Koenig, E. Eulig, M. Knaup, V. Wiesmann, K. Hörndler, and M. Kachelrieß, "Real-time 3d reconstruction of guidewires and stents using two update x-ray projections in a rotating imaging setup," *Medical Physics*, vol. 50, no. 9, pp. 5312–5330, 2023.

[3] R. Clackdoyle and M. Defrise, "Tomographic reconstruction in the 21st century," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 60–80, 2010.

[4] C. Zhang and G.-H. Chen, "Deep-interior: A new pathway to interior tomographic image reconstruction via a weighted backprojection and deep learning," *Medical physics*, vol. 51, no. 2, pp. 946–963, 2024.

[5] L. Mou, Q. Yan, J. Lin, Y. Zhao, Y. Liu, S. Ma, J. Zhang, W. Lv, T. Zhou, A. F. Frangi, and Y. Zhao, "Costa: A multi-center tof-mra dataset and a style self-consistency network for cerebrovascular segmentation," *IEEE Transactions on Medical Imaging*, pp. 1–1, 2024.

[6] K. Yang, F. Musio, Y. Ma, N. Juchler, J. C. Paetzold, R. Al-Maskari, L. Höher, H. B. Li, I. E. Hamamci, A. Sekuboyina *et al.*, "Benchmarking the cow with the topcow challenge: Topology-aware anatomical segmentation of the circle of willis for cta and mra," *ArXiv*, pp. arXiv–2312, 2025.

[7] S. Chatterjee, H. Mattern, M. Dörner, A. Sciarra, F. Dubost, H. Schnurre, R. Khatun, C.-C. Yu, T.-L. Hsieh, Y.-S. Tsai *et al.*, "Smile-uhura challenge–small vessel segmentation at mesoscopic scale from ultra-high resolution 7t magnetic resonance angiograms," *arXiv preprint arXiv:2411.09593*, 2024.

[8] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021.

Coronary Mask–Guided Registration for 4D Cardiac CT Dataset Construction

Yuang Wang, Shuo Wang, Changyu Chen, Le Shen, Yunqiang An, Yang Gao, Bin Lu, Dongrui Dai, Muge Du, Li Zhang, Zhiqiang Chen

***Abstract*—Motion artifacts in clinical multiphase cardiac CT reconstructions degrade visualization of the right coronary artery (RCA), particularly outside the end-diastolic (ED) and end-systolic (ES) phases, limiting their use for deep learning–based CT reconstruction. We propose a coronary mask–guided registration framework for 4D cardiac CT dataset construction from clinical multiphase images. A phase near ED is selected as the fixed reference, and deformation vector fields (DVFs) are estimated to warp the ED volume to the remaining phases. For each phase, an RCA mask is derived to guide the registration process, enabling the DVFs to capture RCA motion while preserving its geometric integrity. Experimental results demonstrate that the proposed approach improves RCA alignment compared with several representative registration methods, while maintaining competitive performance in modeling global cardiac motion.**

***Index Terms*—4D cardiac CT, coronary artery, registration,**

I. Introduction

IMAGING the heart with computed tomography (CT) remains challenging due to rapid and complex cardiac motion and the limited temporal resolution of current CT systems. These factors often lead to pronounced motion artifacts in clinical multiphase reconstructions, particularly affecting visualization of the right coronary artery (RCA). In practice, only one or two phases near end-diastole (ED) or end-systole (ES) typically exhibit relatively low level of motion artifacts, while the remaining phases are severely degraded. As a result, clinical multiphase CT images cannot be directly used as reliable ground truth for training or evaluating deep learning–based cardiac CT reconstruction methods.

To circumvent this issue, several strategies have been explored in previous studies. Some approaches rely on digital phantoms, such as XCAT [1], to provide artifact-free dynamic cardiac data, but the phantoms may not adequately capture the anatomical variability. Other methods [2] extract motion information from phantom-based datasets and apply it to static, low-artifact clinical volumes to synthesize dynamic cardiac sequences. However, these approaches may fail to represent the complex motion patterns present in real cardiac anatomy. Additionally, some studies [3] generate paired training data by applying synthetic motion fields to warp the RCA in static, low-artifact volumes to simulate motion artifacts. Such methods, however, focus primarily on localized coronary motion and ignore the concurrent motion of other cardiac structures.

In this work, we propose a coronary mask–guided registration framework to construct 4D cardiac CT dataset from clinical multiphase reconstructions. A phase close to ED is selected as the fixed reference, and all remaining phases are treated as moving images. For each moving phase, a plausible RCA mask is derived. Guided by this mask, we estimate a deformation vector field (DVF) that captures RCA motion and preserves its geometric integrity. This RCA-focused DVF is then composed with an additional DVF that captures the motion of the remaining cardiac structures, yielding a unified DVF for the entire heart. When applied to the fixed reference, the resulting warped volumes preserve overall cardiac anatomy and substantially reduce motion artifacts relative to the corresponding clinical reconstructions.

We compare our method with several representative registration approaches. The results demonstrate that our method achieves substantially improved RCA alignment and shape preservation, and provides competitive performance in modeling the global cardiac motion.

II. Methods

We consider a clinical multiphase cardiac CT sequence in which a phase close to ED, exhibiting minimal motion artifacts, is selected as the fixed reference, and all remaining phases are treated as moving images. For each moving phase, our objective is to estimate a DVF that describes the underlying cardiac motion relative to the reference phase. Rather than aiming to maximize image similarity between the warped reference and the motion-corrupted moving image, which may lead to overfitting to motion artifacts, the estimated DVF is designed to satisfy three criteria: (1) to warp the RCA to a spatial configuration that is plausible given the observed motion-corrupted appearance in the moving phase, (2) to preserve the geometric integrity of the RCA, and (3) to model the motion of the remaining cardiac structures. The framework for formulating the DVF is illustrated in Fig. 1.

A. RCA Mask Extraction and Refinement

The first stage of the proposed framework aims to obtain a plausible RCA mask for each moving phase. We first apply TotalSegmentator [4] to all cardiac phases to obtain initial coronary artery segmentations. For the reference phase, only the RCA mask is retained. The endpoints of the proximal RCA

Yuang Wang, Shuo Wang, Changyu Chen, Le Shen, Dongrui Dai, Muge Du, Li Zhang and Zhiqiang Chen are (were) with the Department of Engineering Physics, Tsinghua University, Beijing 100084, China.

YunQiang An, Yang Gao, Bin Lu are with the Department of Radiology, Fuwai Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China.

Send correspondence to Li Zhang (zli@mail.tsinghua.edu.cn) and Zhiqiang Chen (czq@mail.tsinghua.edu.cn)

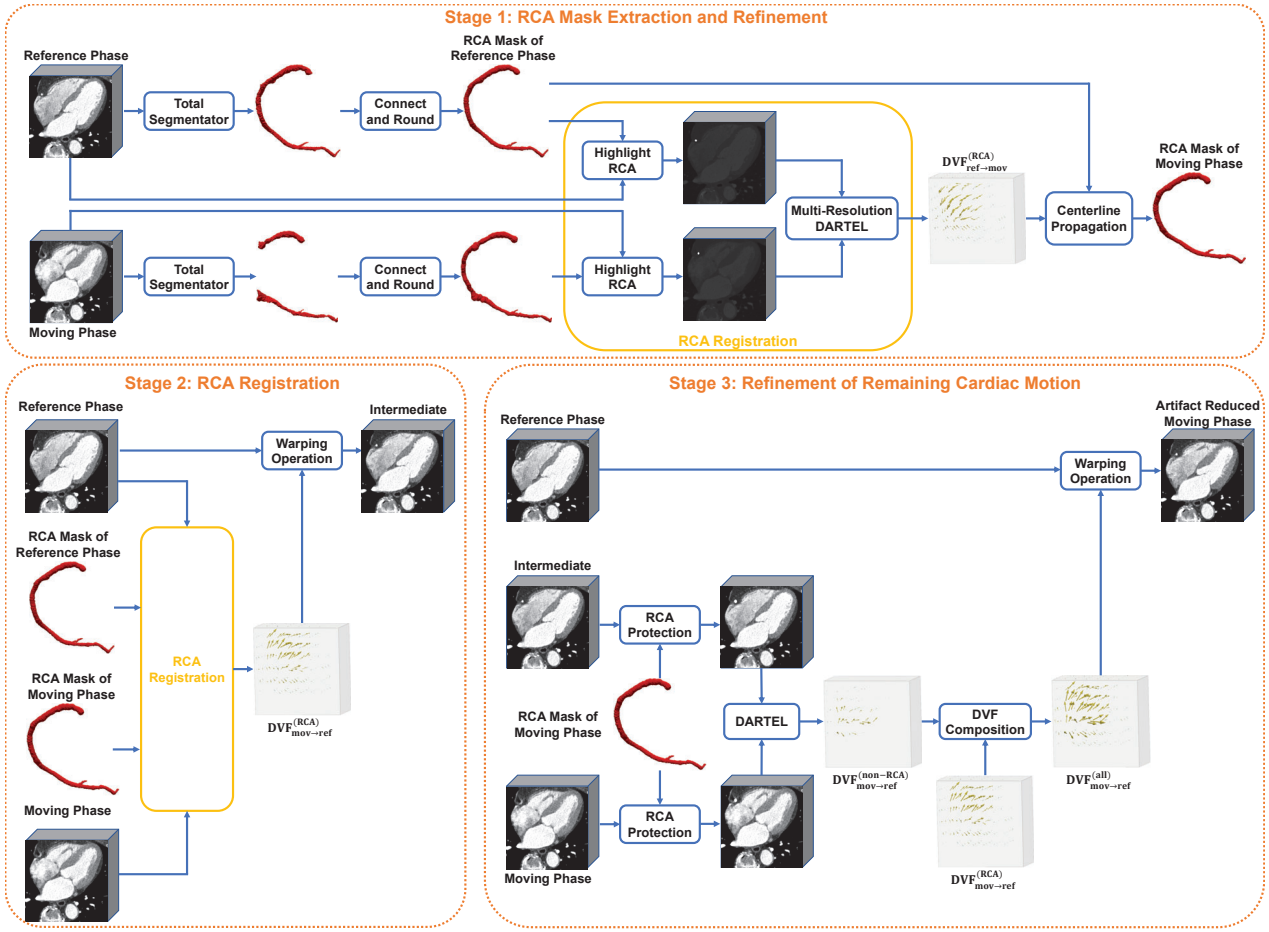


Fig. 1. Framework of coronary mask-guided registration.

and the posterior descending artery (PDA) are identified and used as anatomical reference points for subsequent processing.

For the moving phases, motion artifacts substantially degrade segmentation quality. The initially segmented coronary structures often suffer from multiple issues, including vessel discontinuity, distorted coronary geometry, over-deformed centerlines, and missing or spurious branches.

To restore RCA connectivity, we first identify the segmented coronary fragments corresponding to the proximal RCA and PDA as those closest to the two reference endpoints. If these fragments are not connected, we determine the pair of boundary points with the minimum Euclidean distance. The shortest path between the two boundary points is then computed on the negative log-probability map produced by TotalSegmentator, which serves as a cost function favoring anatomically plausible vessel trajectories. A tubular region with a radius of 1.5 mm centered along this path is merged with the identified fragments, yielding a connected RCA mask.

To regularize distorted coronary geometry, we transform the RCA mask into a circular tubular representation. Specifically, the RCA centerline is extracted using a 3D skeletonization algorithm [5]. For each centerline point p, a local vessel radius $r(p)$ is estimated as

$$r(p) = \mathbb{E}_{b \in S(p)} \|p - b\|_2, \tag{1}$$

where

$$S(p) = \left\{ b \mid b \in B(M),\ \arg\min_{p_0 \in C} \|p_0 - b\|_2 = p \right\}. \tag{2}$$

Here, C denotes the set of centerline points, M the set of voxels within the RCA mask, $B(M)$ the boundary voxels of M. The centerline points together with their associated radii are then oversampled, and the union of the resulting spherical regions is taken as the RCA mask.

At this stage, the RCA mask for each moving phase is ensured to have a continuous and circular main trunk. To address the remaining issues, we perform registration to estimate DVF between the RCA mask in the reference phase and that in each moving phase. The registration volume is constructed by enhancing the region inside the RCA mask and downweighting regions outside it. This design encourages the estimated DVF to prioritize alignment of the RCA while remaining compatible with the surrounding cardiac anatomy. Registration is carried out using DARTEL [6]. To capture the large-amplitude motion of the RCA, a multi-scale registration strategy is adopted. The registration is initialized at a substantially downsampled resolution, and the velocity field estimated at each scale is used to initialize the registration at the next finer resolution. Downsampling is implemented using max pooling; because the RCA is assigned the highest intensity, it remains clearly

identifiable at all resolution levels, allowing it to guide the registration throughout the multi-scale process.

Given the estimated DVF, together with the reference-phase RCA centerline and its associated radii, we update the RCA mask for each moving phase via centerline propagation. To avoid overfitting to potentially over-deformed centerlines induced by motion artifacts, we enforce displacement smoothness along the centerline by solving the optimization problem:

$$\mathbf{U}^\star = \arg\min_{\mathbf{U}} \left(\sum_i \|\mathbf{U}_i - \mathbf{D}_i\|_2^2 + k \sum_{(i,j)\in E} \|\mathbf{U}_i - \mathbf{U}_j\|_2^2 \right), \tag{3}$$

where i and j index points along the reference centerline, $\mathbf{D}$ denotes the displacement given by the DVF, and E represents the set of adjacent centerline point pairs. The optimized displacement $\mathbf{U}^\star$ is applied to the reference centerline to obtain its corresponding trajectory in the moving phase. The vessel radius associated with each centerline point is assumed to remain unchanged. The union of the resulting balls is taken as the RCA mask. This propagation strategy ensures consistent RCA topology and radii across all cardiac phases.

B. RCA Registration

Using the RCA masks obtained in Stage 1, we perform RCA registration between the reference phase and each moving phase in Stage 2. The registration follows the same strategy described in Stage 1, with the RCA serving as the primary structures to be aligned. Applying the resulting DVF to the reference-phase image produces an intermediate warped volume, in which the RCA is mapped to a plausible position in the moving phase and preserves its geometric integrity. This intermediate result is used as the input for subsequent refinement of the remaining cardiac structures.

C. Refinement of Remaining Cardiac Motion

In Stage 3, we refine the motion of the remaining cardiac structures while preserving the RCA alignment achieved in the previous stages. To this end, a RCA protection strategy is introduced to limit further deformation of the RCA during global registration. Specifically, the RCA mask of the moving phase obtained in Stage 1 is isotropically dilated by 0.9 mm to define an expanded RCA region. This region is highlighted with elevated intensity in both the intermediate warped volume from Stage 2 and the original moving-phase image. Consequently, the similarity loss used by DARTEL enforces strong correspondence within the protected region, constraining the RCA from undergoing additional deformation.

Using these protected volumes, DARTEL is applied to estimate a DVF that captures the motion of non-RCA cardiac structures. This DVF is then composed with the RCA-focused DVF obtained in Stage 2 to produce the final DVF. Applying the composed DVF to the reference-phase image yields the final artifact-reduced moving-phase volume.

III. Results

We compare our method with several representative registration approaches, including nonrigid registration [7], DARTEL [6], and UniGradIcon [8]. Experiments are conducted on five clinical cardiac CT cases collected from Fuwai Hospital and the Chinese People's Liberation Army General Hospital. The use of these data was approved by the IRBs.

TABLE I
QUANTITATIVE COMPARISON OF DIFFERENT REGISTRATION METHODS. HD AND MSD ARE IN MM. RMSE IS IN HU. **BOLD**: BEST; UNDERLINE: SECOND BEST.

Method	Dice↑	HD↓	MSD↓	RMSE↓	SSIM↑	LPIPS↓	NC↑	FOR↑
Nonrigid	0.176	13.7	4.46	50.8	0.920	0.184	0.613	0.757
DARTEL	0.317	13.4	3.77	50.6	0.920	0.183	0.617	0.769
UniGradICON	0.636	13.7	2.61	**46.1**	**0.924**	0.179	0.581	0.738
Ours	**0.821**	**7.31**	**0.32**	47.7	0.922	**0.178**	**0.658**	**0.773**

Quantitative evaluation is performed from two complementary perspectives. First, three cases contain a secondary cardiac phase with high image quality close to ES. For these cases, the ED phase is used as the fixed reference, and the ES phase serves as both the moving image and the ground truth. To assess RCA alignment, we compute Dice, Hausdorff distance (HD), and mean surface distance (MSD) between the RCA from the warped volume and the ground-truth ES RCA. To evaluate the capture of global cardiac motion, we report RMSE, SSIM, and LPIPS. Second, for all cardiac phases from all cases, we computed normalized circularity (NC) [9] and fold overlap ratio (FOR) [2] to quantify the circularity and symmetry of through-plane RCA vessel cross-sections.

Table I summarizes the quantitative results. Overall, the proposed method achieves the best or second-best performance across all evaluated metrics. For RCA alignment accuracy, our method significantly outperforms all comparison approaches. It achieves the highest Dice, indicating substantially improved overlap between the warped RCA and the ground-truth ES RCA. In addition, our method yields the lowest HD and MSD, demonstrating markedly improved RCA boundary alignment. Regarding global cardiac motion capture, UniGradICON attains the lowest RMSE (46.1 HU) and the highest SSIM (0.924). Our method achieves the second-best RMSE (47.7 HU) and SSIM (0.922), together with the lowest LPIPS (0.178). These results indicate that our approach demonstrates competitive performance in modeling global cardiac motion, while prioritizing accurate RCA alignment. For vessel cross-section regularity, our method achieves the highest NC and FOR. This suggests improved circularity and symmetry of through-plane RCA vessel cross-sections across phases.

The superior performance of our method is further supported by the visualization results in Fig. 2. In the original clinical images, the RCA is clearly visualized at 40% and 70%, corresponding to the ES and ED phases, respectively, but it is substantially degraded by motion artifacts in the remaining phases. For the comparison methods, in certain phases such as 30% and 90%, the RCA is located outside the blue boxes that indicate the plausible RCA regions identified from the original images, suggesting inaccurate RCA localization. In addition, noticeable RCA distortions can be observed at some phases, particularly at 0% and 100%. In contrast, our method consistently preserves the RCA within the plausible region across all phases and maintains a relatively stable and circular vessel appearance throughout the cardiac cycle.

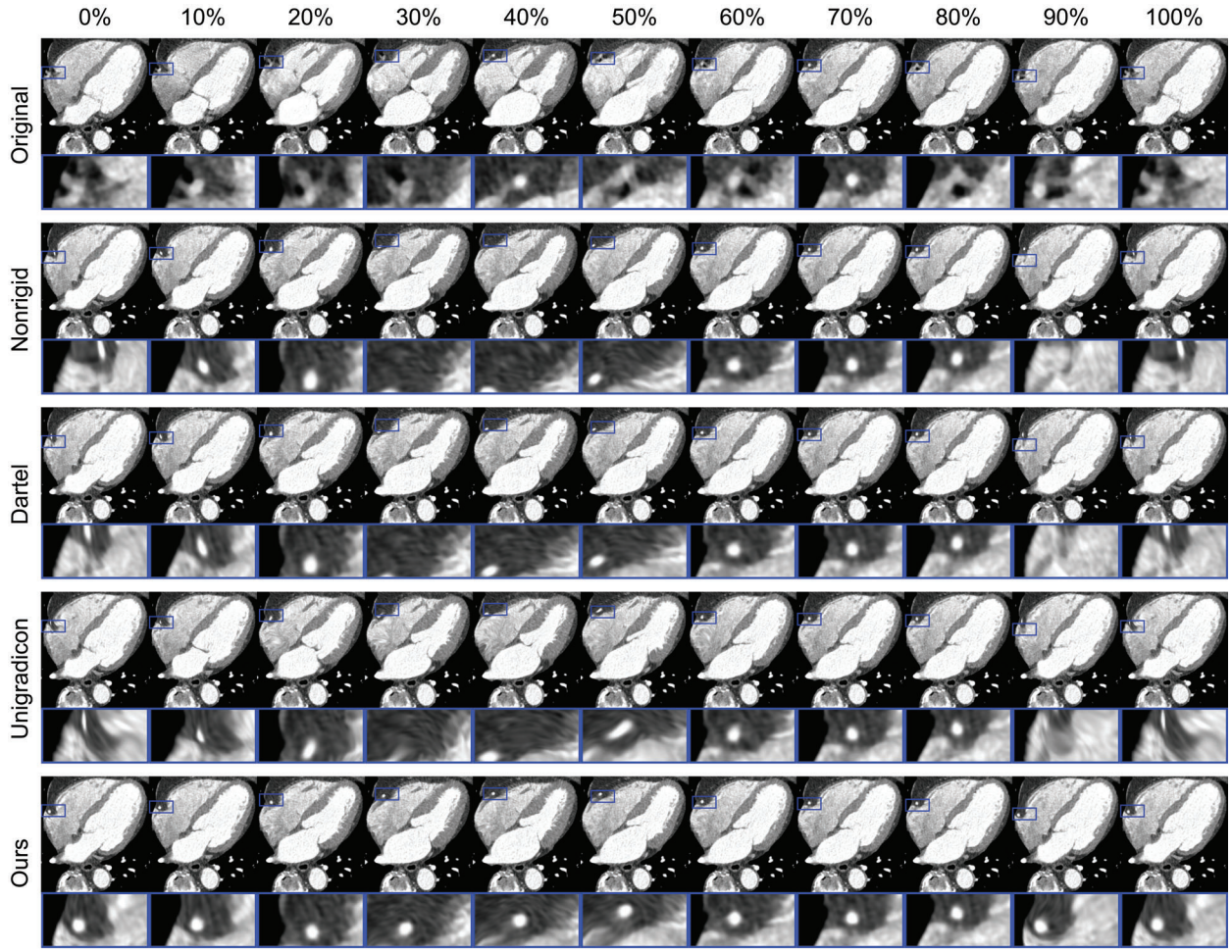


Fig. 2. Representative visualization results over the entire cardiac cycle. Blue boxes indicate the plausible RCA regions identified from the original clinical images. For each cardiac phase, the same spatial region is enlarged for all compared methods. The display window is set to [-250, 450] HU.

IV. Conclusion

We propose a coronary mask–guided registration framework to construct 4D cardiac CT dataset from clinical multiphase reconstructions. The proposed method demonstrates superior performance in capturing RCA motion while preserving its geometric integrity, and simultaneously maintains global cardiac motion. Importantly, the framework relies only on clinical reconstructions and requires neither external training data nor raw projection data. Consequently, it offers a practical means to produce large amounts of realistic, artifact-reduced 4D cardiac CT data, potentially facilitating development of deep learning–based cardiac CT reconstruction.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 12327809 and No. 62031020) and by China Postdoctoral Science Foundation (No. 2025M780861).

References

[1] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4d xcat phantom for multimodality imaging research," *Medical physics*, vol. 37, no. 9, pp. 4902–4915, 2010.

[2] Z. Deng, W. Zhang, K. Chen, Y. Zhou, J. Tian, G. Quan, and J. Zhao, "Tt u-net: Temporal transformer u-net for motion artifact reduction using pad (pseudo all-phase clinical-dataset) in cardiac ct," *IEEE Transactions on Medical Imaging*, vol. 42, no. 12, pp. 3805–3816, 2023.

[3] T. Lossau, H. Nickisch, T. Wissel, R. Bippus, H. Schmitt, M. Morlock, and M. Grass, "Motion estimation and correction in cardiac ct angiography images using convolutional neural networks," *Computerized medical imaging and graphics*, vol. 76, p. 101640, 2019.

[4] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang *et al.*, "Totalsegmentator: robust segmentation of 104 anatomic structures in ct images," *Radiology: Artificial Intelligence*, vol. 5, no. 5, p. e230024, 2023.

[5] T.-C. Lee, R. L. Kashyap, and C.-N. Chu, "Building skeleton models via 3-d medial surface axis thinning algorithms," *CVGIP: graphical models and image processing*, vol. 56, no. 6, pp. 462–478, 1994.

[6] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811907005848

[7] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: application to breast mr images," *IEEE transactions on medical imaging*, vol. 18, no. 8, pp. 712–721, 2002.

[8] L. Tian, H. Greer, R. Kwitt, F.-X. Vialard, R. San José Estépar, S. Bouix, R. Rushmore, and M. Niethammer, "unigradicon: A foundation model for medical image registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2024, pp. 749–760.

[9] H. Ma, E. Gros, A. Szabo, S. G. Baginski, Z. R. Laste, N. M. Kulkarni, D. Okerlund, and T. G. Schmidt, "Evaluation of motion artifact metrics for coronary ct angiography," *Medical physics*, vol. 45, no. 2, pp. 687–702, 2018.

PC-FLOW: A Physics-Constrained Conditional Flow Matching Model For High Pitch Helical Three Dimensional Digital Subtraction Angiography

Yi Liu, Ruifeng Chen, Zexi Sun, Canzhen Ma, Jun Xiang, Dong Liang, Yinsheng Li

***Abstract*—Three-dimensional digital subtraction angiography (3D-DSA) based on C-arm cone-beam CT (CBCT) has become an indispensable imaging modality in the interventional suite. The combined imaging of intracranial, carotid, and aortic vasculature across these interconnected vascular territories is critical. Compared with multi-segment axial scanning (step-and-shoot) using clinical C-arm CBCT systems with limited longitudinal coverage, helical 3D-DSA has attracted increasing interest due to its continuous scanning trajectory and sustained contrast agent injection, which together offer clear advantages in reducing acquisition time, lowering radiation dose, and maintaining a more uniform contrast agent concentration. To address data insufficiency-induced technical challenges associated with high-pitch helical 3D-DSA, this work proposes to introduce a conditional flow matching model as a learned regularization prior, tightly coupled with the geometric characteristics and physical constraints of helical 3D-DSA imaging. Extensive validation and evaluation using numerical simulation studies on 3,485 samples and in vivo animal subject studies have demonstrated that the proposed technique not only alleviates artifacts induced by limited-view-angle and undersampled data at pitches up to 6 (at the isocenter collimation of 3 cm), but also significantly improves the fidelity of structural recovery, enabling more accurate and reliable image reconstruction.**

***Index Terms*—Three-Dimensional Digital Subtraction Angiography, Flow Matching, Separable Paraboloidal Surrogate, High Pitch Helical Acquisition**

I. Introduction

THREE-dimensional digital subtraction angiography (3D-DSA) based on C-arm cone-beam CT (CBCT) has become an indispensable imaging modality in the interventional suite, providing high-resolution volumetric visualization of the vascular tree, depicting vessel morphology, branching patterns, and pathological features [1]. Importantly, many cerebrovascular diseases, such as acute ischemic stroke, intracranial aneurysms, and carotid artery stenosis, are intrinsically linked to vascular territories spanning multiple anatomical regions, including the intracranial circulation, extracranial carotid arteries, and the aortic arch. In the interventional setting, combined imaging of intracranial, carotid, and aortic vasculature across these interconnected vascular territories is critical for accurate assessment of lesion origin, access route planning, embolic source identification, and safe catheter navigation.

As most clinical C-arm CBCT systems equipped with typical flat-panel detectors have limited longitudinal coverage, the combined imaging of intracranial, carotid, and aortic vasculature cannot be achieved via conventional single-segment axial scanning mode. Currently, the most widely adopted approach is multi-segment axial scanning (step-and-shoot protocol), in which three vascular regions are scanned sequentially to achieve complete anatomical coverage [2]. Although practical, this strategy suffers from several major limitations. First, the overall acquisition time is relatively long, increasing the burden of patient cooperation. For patients in acute conditions, prolonged examination time may delay optimal intervention. Second, repeated scans expose the patient to a comparatively higher cumulative radiation dose. Third, the need for multiple acquisitions and repeated contrast injections can lead to inconsistency in contrast agent concentration across three vascular regions, which may adversely affect diagnostic accuracy and inter-study consistency. Finally, due to the constraints of this scanning strategy, contrast agent injection must be performed separately for each scan to maintain adequate vascular opacification, resulting in increased total contrast dose and a potentially elevated risk of contrast-induced renal injury.

To address the aforementioned challenges, helical 3D-DSA has attracted increasing interest due to its continuous helical scanning trajectory and sustained contrast agent injection, which together offer clear advantages in reducing acquisition time, lowering radiation dose, and maintaining a more uniform contrast agent concentration [3]. Furthermore, when the scanning protocol is optimized by increasing the helical pitch, high-pitch helical scanning can significantly shorten scan duration, reduce radiation exposure, achieve more consistent contrast enhancement, and decrease contrast agent volume, thereby improving procedural efficiency and safety in clinical workflows [4].

However, high-pitch helical scanning inevitably results in sparse angular sampling and limited angular ranges of acquired projection data, making the corresponding image reconstruction a severely ill-posed inverse problem. Consequently, reconstructed images often suffer from pronounced artifacts and structural distortions, leading to substantial degradation in image quality.

Motivated by these limitations, this work proposes to introduce a conditional flow matching model as a learned regularization prior, tightly coupled with the geometric characteristics and physical constraints of helical 3D-DSA imaging. A conditional flow matching model pre-trained on large-scale datasets is leveraged to generate candidates through posterior sampling, while subject-specific measured data are used as constraints to guide the reconstruction toward data-consistent solutions via implementing the separable paraboloidal surrogate algorithm. Extensive validation and evaluation using numerical simulation studies and in vivo animal subject studies have demonstrated that the proposed technique not only alleviates artifacts induced by limited-view-angle and undersampled data at high pitches, but also improves the fidelity of structural recovery, enabling more accurate and reliable image reconstruction.

II. Methods

A. Forward Modeling of The Data Acquisition Process

Under a monoenergetic approximation, the object is characterized by the linear attenuation coefficient vector $\boldsymbol{\mu} = \{\mu_j\}$ after discretizing the imaging domain into N voxels. For the i-th ray, the expected photon count $\bar{N}_i$ is a function of $\boldsymbol{\mu}$ governed by the Beer–Lambert law:

$$\bar{N}_i = N_{0,i} \exp\left(-\sum_j a_{ij}\mu_j\right), \tag{1}$$

where $N_{0,i}$ denotes the incident photon count, and a_{ij} is an element of the system matrix, $\mathbf{A} \in \mathbb{R}^{M\times N}$, where M denote the total number of measurements.

B. Helical Image Reconstruction via Maximum a Posterior Estimation

Accounting for the stochastic nature of photon detection, the observed photon count at the i-th ray is denoted by N_i. Given the attenuation distribution $\boldsymbol{\mu}$, the measurements $\{N_i\}_{i=1}^M$ are modeled as mutually independent Poisson random variables, whose means $\bar{\mathbf{N}}$ are determined by the forward model in Eq. (1). Accordingly, the joint conditional probability mass function is expressed as:

$$P(\mathbf{N} \mid \bar{\mathbf{N}}) = \prod_{i=1}^{M} P(N_i \mid \bar{N}_i) = \prod_{i=1}^{M} \frac{\bar{N}_i^{N_i}\exp(-\bar{N}_i)}{N_i!}. \tag{2}$$

Using Bayesian's formula, $P(\bar{\mathbf{N}} \mid \mathbf{N}) \propto P(\mathbf{N} \mid \bar{\mathbf{N}})\,P(\bar{\mathbf{N}})$. Since $\bar{\mathbf{N}}$ is determined by $\boldsymbol{\mu}$, both $P(\mathbf{N} \mid \bar{\mathbf{N}})$ and $P(\bar{\mathbf{N}})$ can be viewed as functions of $\boldsymbol{\mu}$. Consequently, the estimation of $\boldsymbol{\mu}$ is framed as a maximum a posteriori (MAP) problem. By taking the negative log-likelihood and omitting terms constant with respect to $\boldsymbol{\mu}$, maximizing the posterior probability is equivalent to minimizing the following objective functional:

$$\begin{aligned}
\boldsymbol{\mu}^* &= \arg\max_{\boldsymbol{\mu}} \log P\big(\bar{\mathbf{N}} \mid \mathbf{N}\big) \\
&= \arg\max_{\boldsymbol{\mu}} \big\{\log P\big(\mathbf{N} \mid \bar{\mathbf{N}}\big) + \log P\big(\bar{\mathbf{N}}\big)\big\} \\
&= \arg\min_{\boldsymbol{\mu}} \Big\{\Big(\sum_i \bar{N}_i - N_i \log \bar{N}_i\Big) - \log P(\bar{\mathbf{N}})\Big\} \\
&:= \arg\min_{\boldsymbol{\mu}} \{L(\boldsymbol{\mu}) - \log P(\boldsymbol{\mu})\},
\end{aligned} \tag{3}$$

Solving the above MAP problem is challenging for two reasons. First, the data-consistency term induced by the Beer–Lambert law is nonlinear, often leading to a non-convex objective function that complicates optimization. Second, it is non-trivial to specify a computationally tractable prior that can faithfully capture the complex structures of the target attenuation map.

C. Separable Paraboloidal Surrogate of $L(\boldsymbol{\mu})$ via Optimization Transfer Principle

To efficiently minimize the Poisson data-consistency term $L(\boldsymbol{\mu})$, we construct a separable paraboloidal surrogate (SPS) based on the optimization transfer principle [5]. The original coupled objective is replaced by a quadratic surrogate function $\mathbf{Q}(\boldsymbol{\mu}; \boldsymbol{\mu}_n)$ that is separable with respect to the voxel-wise variables. At each iteration, minimizing $\mathbf{Q}(\boldsymbol{\mu}; \boldsymbol{\mu}_n)$ ensures a monotonic decrease of the original objective L. According to the optimization transfer principle, the surrogate function $\mathbf{Q}(\boldsymbol{\mu}; \boldsymbol{\mu}_n)$ can be formulated as:

$$\mathbf{Q}(\boldsymbol{\mu}; \boldsymbol{\mu}_n) = \sum_i \sum_j \alpha_{ij} q_i \left(\frac{a_{ij}}{\alpha_{ij}}(\mu_j - \mu_{j,n}) + \sum_k a_{ik}\mu_{k,n}\right), \tag{4}$$

where $\alpha_{ij} > 0$ are the weights satisfying $\sum_j \alpha_{ij} = 1$, and $q_i(\cdot)$ denotes the univariate quadratic surrogate for the i-th measurement, which typically incorporates the curvature and measurement data.

D. Learning a Conditional Rectified Flow Prior for Helical Image Reconstruction

To tackle the second challenge, the prior is modeled as a conditional distribution $P(\boldsymbol{\mu}|\mathbf{y})$ via the Rectified Flow framework [6]. To incorporate volumetric spatial continuity, the conditioning variable $\mathbf{y}$ is formed by channel-wise concatenation of the initial reconstructions from adjacent slices. This formulation explicitly encourages the model to distinguish the vascular structures from helical artifacts.

Unlike standard diffusion models, Rectified Flow establishes a deterministic transport map between a noise distribution $\pi_0 = \mathcal{N}(\mathbf{0}, \mathbf{I})$ and the target data distribution π_1. Specifically, given a sample pair $(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ drawn from a joint coupling of π_0 and π_1, a linear interpolation path is defined as:

$$\boldsymbol{\mu}_t = t\boldsymbol{\mu}_1 + (1-t)\boldsymbol{\mu}_0, \quad t \in [0,1], \tag{5}$$

which implies a constant ideal velocity field $\mathbf{v}^* = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_0$. Consequently, the generative process is governed by the ordinary differential equation (ODE):

$$\frac{\mathrm{d}\boldsymbol{\mu}_t}{\mathrm{d}t} = \mathbf{v}_\theta(\boldsymbol{\mu}_t, t, \mathbf{y}), \tag{6}$$

where $\mathbf{v}_\theta$ is a velocity field network parameterized by θ.

As illustrated in Fig. 1, $\mathbf{v}_\theta$ employs a U-Net architecture that takes the noisy state $\boldsymbol{\mu}_t$, the condition $\mathbf{y}$, and the time step t as inputs to predict the flow velocity. The network is trained by minimizing the conditional flow-matching objective:

$$\mathcal{L}(\theta) = \mathbb{E}_{t,\boldsymbol{\mu}_1,\boldsymbol{\mu}_0}\left[\|\mathbf{v}_\theta(\boldsymbol{\mu}_t, t, \mathbf{y}) - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)\|_2^2\right]. \tag{7}$$

During inference, the artifact-free image is recovered by numerically integrating Eq. (7) from $t = 0$ to $t = 1$ using an Euler solver.

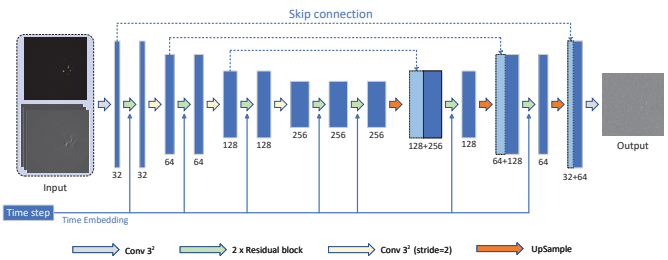


Fig. 1. The architecture of the conditional velocity field network $\mathbf{v}_\theta$

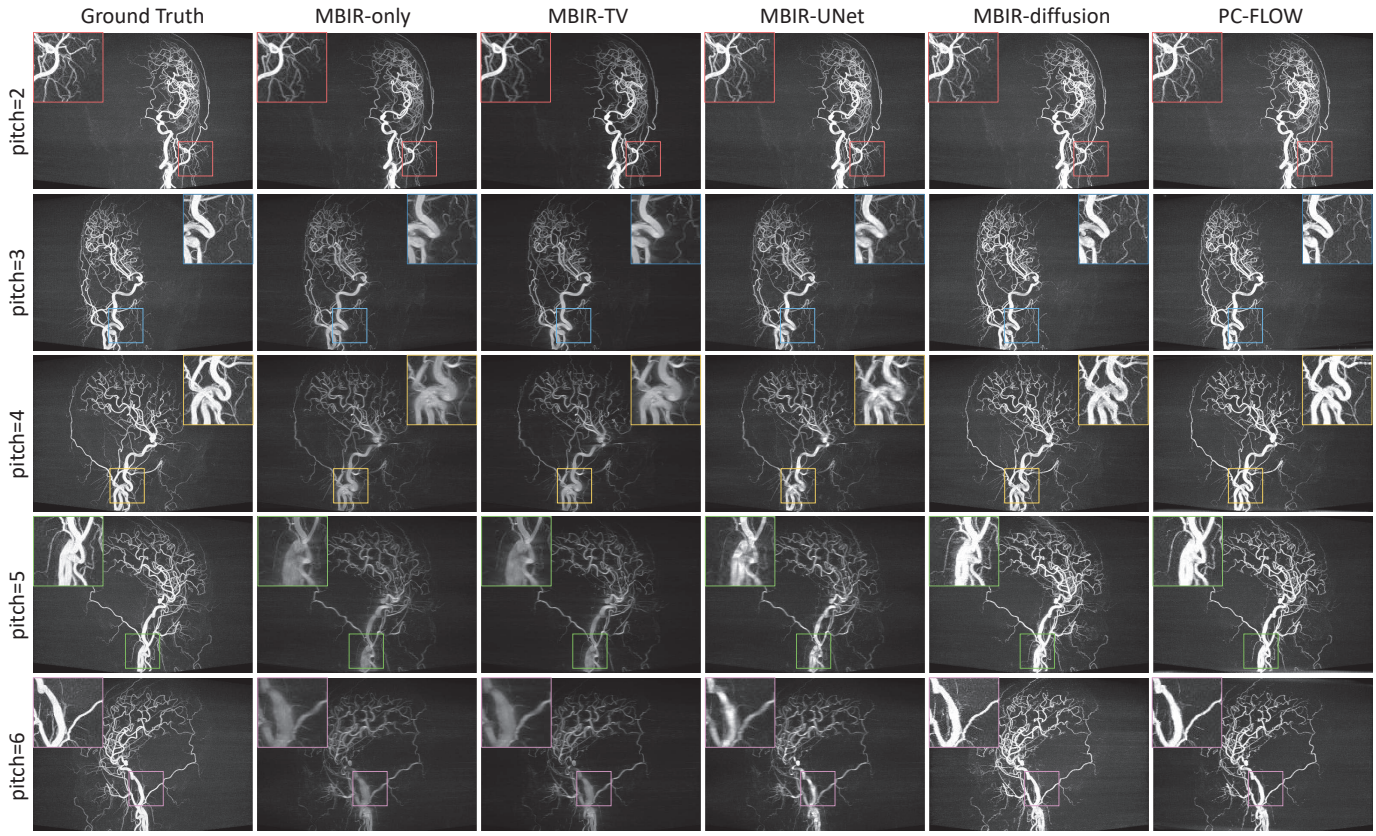


Fig. 2. **Qualitative comparison of MIP results from the numerical simulation study across helical pitches** $p \in \{2, 3, 4, 5, 6\}$**.** The images are reconstructed via conventional MBIR, MBIR-TV, MBIR-UNet, diffusion-regularized MBIR (MBIR-diffusion) and PC-FLOW. Note that PC-FLOW preserves vascular continuity even at high pitches.

E. Numerical Optimization Algorithm

The MAP optimization problem in Eq. (3) is solved via an alternating minimization strategy, splitting the variable update into a data-consistency step and a prior-refinement step:

$$\boldsymbol{\mu}_{t+\frac{\Delta t}{2}} = \arg\min_{\boldsymbol{\mu}} Q(\boldsymbol{\mu}; \boldsymbol{\mu}_t), \tag{8}$$

$$\boldsymbol{\mu}_{t+\Delta t} = \arg\min_{\boldsymbol{\mu}} \left\{ \frac{1}{2} \left\| \boldsymbol{\mu} - \boldsymbol{\mu}_{t+\frac{\Delta t}{2}} \right\|_F^2 - \log P(\boldsymbol{\mu} \mid \mathbf{y}) \right\}, \tag{9}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

The first sub-problem Eq. (8) regarding data consistency can be solved using Newton's Method:

$$\begin{aligned} \mu_{j,(t+\frac{\Delta t}{2})} &= \mu_{j,t} - \frac{\left[\frac{\partial Q(\boldsymbol{\mu};\boldsymbol{\mu}_t)}{\partial \mu_j}\right]_{\boldsymbol{\mu}=\boldsymbol{\mu}_t}}{\left[\frac{\partial^2 Q(\boldsymbol{\mu};\boldsymbol{\mu}_t)}{\partial \mu_j^2}\right]_{\boldsymbol{\mu}=\boldsymbol{\mu}_t}} \\ &:= \mu_{j,t} - [\mathcal{H}(\boldsymbol{\mu}_t)]_j, \end{aligned} \tag{10}$$

where $\frac{\partial Q(\boldsymbol{\mu};\boldsymbol{\mu}_t)}{\partial \mu_j}$ and $\frac{\partial^2 Q(\boldsymbol{\mu};\boldsymbol{\mu}_t)}{\partial \mu_j^2}$ correspond to the gradient and curvature of the SPS surrogate, respectively. The second sub-problem Eq. (9) is solved by the following updating sequence based on the learned Rectified Flow prior:

$$\boldsymbol{\mu}_{t+\Delta t} = \boldsymbol{\mu}_{t+\frac{\Delta t}{2}} + \Delta t \cdot \mathbf{v}_\theta(\boldsymbol{\mu}_{t+\frac{\Delta t}{2}}, t, \mathbf{y}), \tag{11}$$

where $\mathbf{v}_\theta$ is the predicted velocity field conditioned on $\mathbf{y}$, and $\Delta t = 1/N$ denotes the uniform discretization step size, with N representing the total number of sampling steps over the time interval $t \in [0, 1]$. The whole iterative reconstruction process is shown in Algorithm 1.

Algorithm 1 PC-FLOW

Require: k, Δt, $\mathcal{H}(\cdot)$, $\mathbf{v}_\theta(\cdot)$, $\mathbf{y}$
$\boldsymbol{\mu}_0 \leftarrow \mathcal{N}(\mathbf{0}, \mathbf{I})$
for $t = 0 \to 1$ **do**
  for $i = 1, \ldots, k$ **do**
    $\boldsymbol{\mu}_t \leftarrow \boldsymbol{\mu}_t - \mathcal{H}(\boldsymbol{\mu}_t)$
  end for
  $\boldsymbol{\mu}_{t+\frac{\Delta t}{2}} \leftarrow \boldsymbol{\mu}_t$
  $\boldsymbol{\mu}_{t+\Delta t} \leftarrow \boldsymbol{\mu}_{t+\frac{\Delta t}{2}} + \Delta t \cdot \mathbf{v}_\theta(\boldsymbol{\mu}_{t+\frac{\Delta t}{2}}, t, \mathbf{y})$
end for
return $\boldsymbol{\mu}_1$

F. Experimental Setup

1) Numerical Simulations: Numerical experiments were conducted using 3D cerebrovascular digital phantoms to qualitatively and quantitatively assess reconstruction quality. Helical cone-beam projections were synthesized using a clinically realistic C-arm geometry with an isocenter collimation of 3 cm across a spectrum of helical pitches $p \in \{2, 3, 4, 5, 6\}$. To establish the conditional prior, the dataset was partitioned into 6,750 slices for training and 3,485 slices for testing.

Quantitative performance is evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) against the ground truth.

2) Preclinical Validation: To validate clinical feasibility, an in-vivo canine helical DSA scan was acquired on a commercial C-arm system with an isocenter collimation of approximately 24 cm. Two distinct acquisition protocols were investigated: (1) a standard clinical routine protocol utilizing the full detector array (1294 rows) at a pitch of $p = 1.7$; and (2) a challenging high-pitch scenario ($p = 3.4$) emulated by retrospectively subsampling the detector to 647 rows while maintaining the table translation speed. Standard vendor-provided geometric calibration and pre-processing were applied prior to reconstruction. Due to the absence of ground truth in vivo, performance is evaluated qualitatively based on artifact suppression capability and the preservation of distal vascular details.

III. Results

A. Qualitative Image Quality Assessments and Comparisons

Fig. 2 illustrates the Maximum Intensity Projections (MIPs) of the volumes reconstructed via conventional MBIR, total-variation regularized MBIR (MBIR-TV), supervised U-Net post-processing (MBIR-UNet), diffusion-regularized MBIR (MBIR-diffusion) and PC-FLOW. The PC-FLOW images demonstrate significantly suppressed artifacts and longitudinal stretching, preserved distal vascular continuity that closely resembles the ground truth, and superior structural integrity compared to the over-smoothed or artifact-corrupted comparison methods.

Qualitative results for in-vivo canine helical DSA ($p = 1.7, 3.4$) are shown in Fig. 3. At $p = 3.4$, comparison methods exhibit significant vascular warping while PC-FLOW and MBIR-diffusion appear visually competitive, both successfully mitigating the high-pitch artifacts. PC-FLOW provides clear structural continuity and natural vessel morphology, demonstrating a significant improvement over the over-smoothed or distorted results observed in MBIR-UNet and MBIR-TV.

B. Quantitative Image Quality Assessments and Comparisons

Quantitative comparisons in Table I show that PC-FLOW achieves state-of-the-art PSNR and SSIM, matching the performance of the computation-heavy MBIR-diffusion. However, PC-FLOW is drastically faster (11.4 min vs. 182.3 min), making it a far more practical solution for clinical workflows.

IV. Conclusion

In this work, it has been demonstrated that the utilization of a conditional flow matching prior can effectively address the artifacts associated with high-pitch helical 3D-DSA. Specifically, numerical simulations and preclinical in-vivo studies indicate that PC-FLOW significantly enhances vascular conspicuity and suppresses data-incompleteness artifacts. Furthermore, PC-FLOW enables high-quality reconstruction with high computational efficiency, demonstrating great potential for rapid neurointerventional workflows.

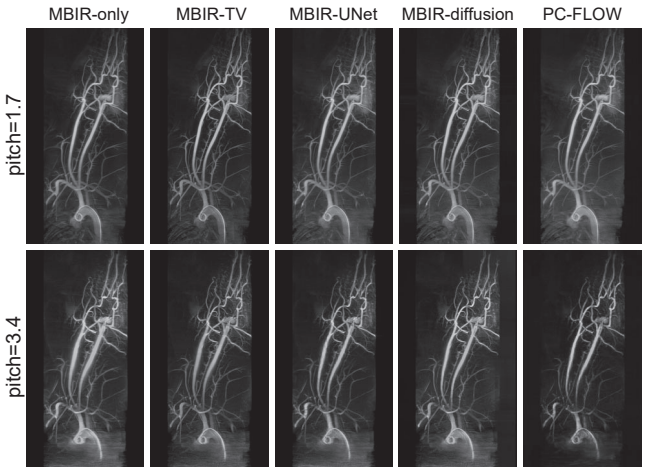


Fig. 3. **Qualitative comparison of MIP results from the in-vivo canine study.** The top row displays results from the standard clinical protocol ($p = 1.7$), while the bottom row shows the challenging high-pitch scenario ($p = 3.4$) emulated via retrospective subsampling. Comparison methods include conventional MBIR, MBIR-TV, MBIR-UNet, diffusion-regularized MBIR (MBIR-diffusion) and PC-FLOW. Note that PC-FLOW effectively suppresses streak artifacts in the high-pitch case compared to competing approaches.

TABLE I
Quantitative Comparison of Image Quality and Computational Efficiency Under Different Helical Pitches using 3485 samples for testing

Pitch	Metric	MBIR-only	MBIR-TV	MBIR-UNet	MBIR-diffusion	PC-FLOW
$p = 2$	PSNR	41.83±0.11	38.49±0.11	42.49±0.14	42.74±0.08	**43.88±0.09**
	SSIM	0.945±0.001	0.875±0.002	0.954±0.001	0.959±0.001	**0.967±0.002**
$p = 3$	PSNR	38.28±0.09	37.00±0.09	37.89±0.12	38.04±0.08	**39.67±0.09**
	SSIM	0.905±0.001	0.854±0.002	0.902±0.002	0.916±0.001	**0.940±0.001**
$p = 4$	PSNR	36.99±0.09	36.25±0.09	36.50±0.12	**37.63±0.09**	37.56±0.06
	SSIM	0.887±0.002	0.846±0.002	0.873±0.002	**0.902±0.001**	**0.902±0.001**
$p = 5$	PSNR	35.48±0.10	35.41±0.09	35.64±0.11	35.53±0.08	**35.87±0.08**
	SSIM	0.857±0.002	0.841±0.002	0.852±0.002	0.868±0.002	**0.872±0.001**
$p = 6$	PSNR	35.29±0.10	35.14±0.09	35.16±0.11	**35.69±0.10**	35.39±0.09
	SSIM	0.842±0.002	0.830±0.003	0.840±0.003	**0.858±0.002**	0.848±0.002
Avg. Time (min)		18.7	34.6	24.9	182.3	**11.4**

References

[1] M. J. Wallace, M. D. Kuo, C. Glaiberman, C. A. Binkert, R. C. Orth, and G. Soulez, "Three-dimensional c-arm cone-beam ct: applications in the interventional suite," *Journal of Vascular and Interventional Radiology*, vol. 20, no. 7, pp. S523–S537, 2009.

[2] J. Hsieh, J. Londt, M. Vass, J. Li, X. Tang, and D. Okerlund, "Step-and-shoot data acquisition and reconstruction for cardiac x-ray computed tomography," *Medical Physics*, vol. 33, no. 11, pp. 4236–4248, 2006.

[3] R. Anxionnat, S. Bracard, X. Ducrocq, Y. Trousset, L. Launay, E. Kerrien, M. Braun, R. Vaillant, F. Scomazzoni, A. Lebedinsky *et al.*, "Intracranial aneurysms: clinical value of 3d digital subtraction angiography in the therapeutic decision and endovascular treatment," *Radiology*, vol. 218, no. 3, pp. 799–808, 2001.

[4] P. Apfaltrer, E. L. Hanna, U. J. Schoepf, J. R. Spears, S. O. Schoenberg, C. Fink, and R. Vliegenthart, "Radiation dose and image quality at high-pitch ct angiography of the aorta: intraindividual and interindividual comparisons with conventional ct angiography," *American Journal of Roentgenology*, vol. 199, no. 6, pp. 1402–1409, 2012.

[5] I. Elbakri and J. Fessler, "Statistical image reconstruction for polyenergetic x-ray computed tomography," *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 89–99, 2002.

[6] X. Liu, C. Gong, and Q. Liu, "Flow straight and fast: Learning to generate and transfer data with rectified flow," *arXiv preprint arXiv:2209.03003*, 2022.

A Framework for Ultra-High-Resolution CT via Data Synthesis and Latent Diffusion Bridge

Yang Sun, *Member, IEEE,* ShaoYang Zhang, *Member, IEEE,*Li Zhang, *Member, IEEE,*

***Abstract*—Enhancing the resolution of CT images is pivotal for diagnosis, yet it is constrained by hardware limitations and the increased processing time required for larger image scales. In this paper, we propose DDSR-CT, a novel framework for Decoupled Denoising and Super-Resolution for CT images. We decompose the task into two stages: first, a denoising module removes CT-specific artifacts; second, a Latent Diffusion Bridge module performs super-resolution in latent space to ensure computational efficiency. To alleviate the scarcity of UHR CT data for training, we further develop a pipeline that synthesizes 2048×2048 training data leveraging a natural-domain pretrained SR model. Experimental results demonstrate that our method outperforms the classical GAN-based baseline, yielding a 5.6% reduction in LPIPS score and achieves performance comparable to the advanced diffusion-based method. Notably, our framework delivers seconds-level inference efficiency without patch-wise processing.**

***Index Terms*—CT super-resolution, denoising, Diffusion model.**

I. Introduction

SUPER-RESOLUTION (SR) is the process of recovering high-resolution (HR) images from low-resolution (LR) observations. As a fundamental task in medical image processing, SR plays a pivotal role in overcoming the intrinsic trade-offs between spatial resolution, noise, and radiation dose. Moreover, SR techniques offer the potential to enhance diagnostic accuracy by clarifying subtle pathological structures.

Driven by the rapid evolution of deep learning techniques, numerous SR frameworks have emerged, ranging from early Convolutional Neural Networks (CNNs) to more recent Transformers, Generative Adversarial Networks (GANs), and Diffusion Models [1]. These methods have demonstrated remarkable success in the domain of natural image, pushing the boundaries of perceptual quality and structural restoration. However, despite these advances, the direct application of the above trained SR models to CT images remains fundamentally restricted by the substantial differences in image characteristics between CT images and natural images. Furthermore, acquiring high-quality utral high-resolution (UHR) CT data is often constrained by physical and ethical factors: hardware limitations impose a hard ceiling on spatial resolution, whereas the accessibility of authentic clinical data is often restricted by stringent clinical safety and ethical standards. Lastly, even if high-quality UHR CT data are successfully acquired, models trained on such datasets may fail when deployed on different imaging systems.

In this paper, we propose a novel framework for Decoupled Denoising and Super-Resolution for CT image (DDSR-CT). Based on this architecture, we develop a synthesis method for UHR CT data (2048×2048 resolution), which subsequently facilitates the training of the super-resolution module of DDSR-CT. Experimental results demonstrate that the proposed framework outperforms traditional methods and achieves comparable performance to a representative diffusion-based model, with improved computational efficiency.

The authors are with the Department of Engineering Physics, Tsinghua University, Beijing 100084, China (e-mail: zli@mail.tsinghua.edu.cn).

II. Method

The overview of the proposed framework is illustrated in Figure 1. Our method integrates three synergistic components: (A) CT denoising module, (B) UHR CT data synthesis pipeline, (C) Latent Diffusion Bridge SR module. First, the CT denoising module is developed as the foundational unit to suppress domain-specific noise in CT images. Once optimized, this module is employed within the data synthesis pipeline, which could generate UHR CT images from raw clinical images. Finally, these synthesized UHR data are used to train the Latent Diffusion Bridge module, which performs the super-resolution task in latent space.

A. Denoising Module

CT scans frequently contain perceptible streaking artifacts and noise. While these artifacts are often subtle enough not to impede manual clinical diagnosis, they pose a challenge for SR tasks. Specifically, we observe that most natural-domain SR models, which are optimized for sharp edge restoration, tend to misinterpret these linear streaking artifacts as high-frequency anatomical structures. To tackle this problem, we propose to utilize the model originally developed for low-dose CT (LDCT) enhancement to eliminate residual streaking artifacts and noise in normal-dose CT (NDCT) images. The feasibility of this approach stems from the fact that the artifact and noise levels in NDCT images are statistically similar to those found in mildly degraded low-dose scans. Consequently, these patterns fall within the manifold that the low-dose enhancement network is trained to identify and suppress.

Specifically, a modified version of the original U-Net [2] is employed as our denoising module. We improve the original U-Net by replacing Batch Normalization layers with Group Normalization layers and substituting Transposed Convolution layers with interpolation-based Upsampling layers to mitigate checkerboard artifacts. Moreover, we adopt the global residual connection that adds the input directly to the module's output, as formulated in Eq.1:

$$\mathcal{N}_\theta(x) = n_\theta(x) + x \tag{1}$$

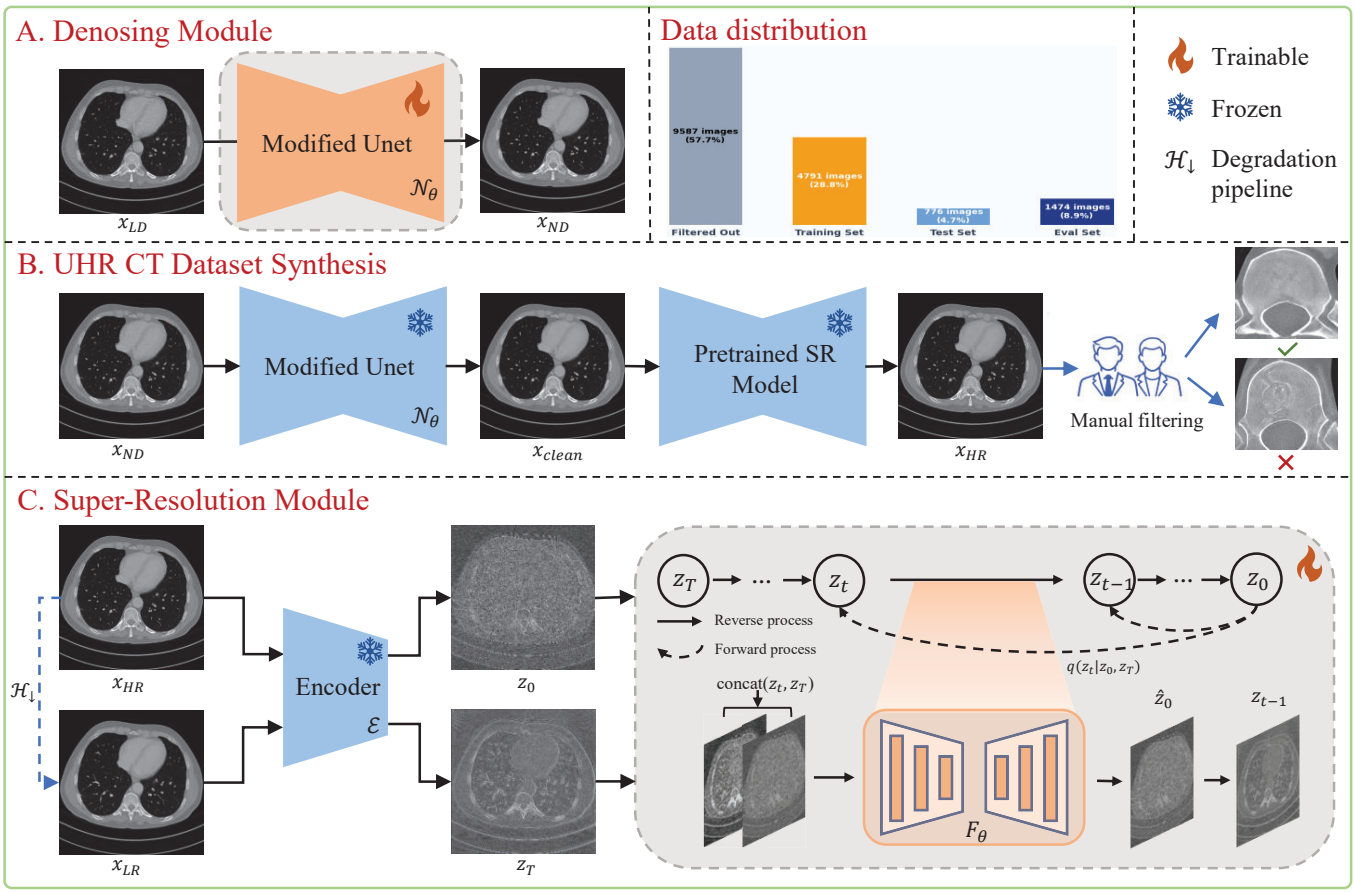


Fig. 1: Overview of the proposed DDSR-CT framework: (A) The denoising module trained on the LDCT enhancement dataset. (B) UHR CT data synthesis by super-resolving denoised normal-dose CT images using a pretrained SR model, followed by manual filtering; (C) The super-resolution module that performs super-resolution base on the diffusion bridge model in latent space.

where x denotes the LDCT image, $\mathcal{N}_\theta$ denotes the denoising module, n_θ denotes the backbone of the module. $\mathcal{N}_\theta$ is optimized by minimizing the simple L2 loss, i.e.,

$$\mathcal{L}(\mathcal{N}_\theta) = \mathbb{E}_{x,y}[\|\mathcal{N}_\theta(x) - y\|_2^2] \tag{2}$$

where y denotes the NDCT image corresponding to x. We use the "2016 NIH-AAPM Mayo Clinic Low-Dose CT Grand Challenge" (Mayo 2016) [3] dataset for training.

Despite the fantastic performance of many recent LDCT enhancement models, they often involve complex architectures or multi-step sampling processes, which hinder seamless integration into our framework. In contrast, the modified U-Net features a structurally simple design with robust noise suppression capability, making it well suited for our framework.

B. Ultra High-Resolution CT Dataset Synthesis

Super-resolving CT images from 512×512 to 2048×2048 remains a challenge due to the extreme scarcity of UHR training data. Acquiring CT images at such high spatial resolutions typically necessitates cutting-edge hardware, such as advanced X-ray sources and specialized systems (e.g., CBCT, Micro-CT). Furthermore, even when UHR scans are successfully acquired through these advanced imaging systems, ethical constraints and patient privacy concerns often restrict data sharing, resulting in UHR CT data remaining closed-source and inaccessible to the broader research community.

In this section, we propose to synthesize the UHR CT dataset by applying the SR model pretrained on natural images to the denoised CT images. This approach is built upon the hypothesis that the preceding denoising stage effectively eliminates CT-specific artifacts and noise, thereby narrowing the domain gap and shifting the CT image distribution closer to that of natural scenes. However, the pretrained SR model may still introduce hallucinated textures when applied to denoised CT images. To preserve the physical fidelity of the synthetic UHR CT images, we manually filter the generated data.

Regarding the choice of the pretrained SR model, we opt for SUPIR[1] due to its extensive training on large-scale, high-quality natural datasets and its restoration-guided sampling mechanism. These characteristics enable SUPIR to achieve superior perceptual fidelity and structural coherence in the synthesized images compared to conventional methods.

As illustrated in Figure 1, the UHR CT data synthesis process is also based on the Mayo 2016 dataset, which contains a total of 16,628 pair slices. Data from one patient (1474 slices) are reserved as the validation set for the overall framework, while the remaining slices are used to train the preceding denoising module. The trained denoising module is then applied to the NDCT images in the training set to generate clean images, which are subsequently fed into SUPIR to synthesize UHR CT data. To ensure data reliability, we perform manual quality control to filter the synthesized

dataset. Specifically, images exhibiting hallucinated artifacts or clinically implausible details are discarded. Following this procedure, 9,587 slices are excluded, and the remaining images are used for training the subsequent SR module. Specifically, slices (776) from one patient are selected as the test set, while the remaining 4,791 slices are used for training the subsequent SR module.

C. Super-Resolution Module

As discussed in the data synthesis section, SUPIR may also introduce hallucinated textures that are not faithful to the input image, which could potentially lead to diagnostic errors. Second, the massive architecture of SUPIR and the multiple sampling steps required during inference result in significant computational latency; super-resolving a single 512×512 slice can take several minutes. For clinical CT volumes that typically contain hundreds of slices, such a time cost is practically unacceptable.

To address these challenges, we introduce the Latent Diffusion Bridge (LDB) as our SR module. In this module, a variational autoencoder is used to encode images into latent representations at 1/8 spatial resolution. This encoding facilitates efficient processing of high-resolution images without resorting to patch-based tiling or aggressive model compression, while preserving both global structural consistency and fine-grained anatomical details. The diffusion bridge is a special model different from conventional diffusion models that generate images from pure noise. Instead, it directly models the image-to-image transition between data distributions, making it particularly suitable for image restoration tasks [4]. By adopting the diffusion bridge in latent space, our module bypasses the need for additional conditional control networks (e.g., ControlNet), thereby reducing parameter redundancy while ensuring high-quality reconstruction.

Practically, we leverage the Variational Autoencoder (VAE) from Stable Diffusion 3 (SD3) [5] to perform bidirectional latent space transformations, owing to its state-of-the-art reconstruction fidelity. This process can be formulated as follows

$$z = \mathcal{E}(x), \quad \hat{x} = \mathcal{D}(z) \tag{3}$$

where $\mathcal{E}$ and $\mathcal{D}$ represent the encoder and decoder respectively, z denotes the latent representation of image x. Subsequently, we prepare the training pairs by applying the Real-ESRGAN [6] degradation pipeline alongside CT forward/back-projection simulations to generate LR counterparts. This strategy is adopted because the UHR CT images may exhibit imperceptible deviations compared to the original CT scans. This data preparation process can be formulated as:

$$x_{LR} = \mathcal{H}_{\downarrow}(x_{HR}), \quad z_0 = \mathcal{E}(x_{HR}), \quad z_T = \mathcal{E}(x_{LR}) \tag{4}$$

where $\mathcal{H}_{\downarrow}$ denotes the degradation pipeline.

For the diffusion bridge, we adopt the Image-to-Image Schrödinger Bridge (I2SB) [4] to model the stochastic transition from low-resolution to high-resolution latent representations. Furthermore, by utilizing an advanced ODE-based sampling strategy [7], the inference process can be compressed to only a few dozen steps. The forward process of I2SB is defined as a diffusion bridge between two paired endpoints: the data distribution $q(z_0)$ and a conditional/prior distribution $q(z_T)$ (e.g., a degraded image). For any $t \in (0, T)$, the intermediate state z_t given the endpoints z_0 and z_T follows a Gaussian distribution:

$$q(z_t|z_0, z_T) = \mathcal{N}(a_t z_T + b_t z_0, c_t^2 \mathbf{I}) \tag{5}$$

where the coefficients are given by

$$a_t = \frac{\alpha_t}{\alpha_T}\frac{\mathrm{SNR}_T}{\mathrm{SNR}_t}, b_t = \alpha_t(1 - \frac{\mathrm{SNR}_T}{\mathrm{SNR}_t}), c_t^2 = \sigma_t^2(1 - \frac{\mathrm{SNR}_T}{\mathrm{SNR}_t})$$

and $\mathrm{SNR}_t = \alpha_t^2/\sigma_t^2$ is the signal-to-noise ratio at time t. The noise schedules can be defined as follows:

$$\alpha_t = 1, \sigma_t = \sqrt{\int_0^t g^2(\tau)\, d\tau}, g^2(t) = (\eta_T - \eta_0|2t - 1|)^2$$

with

$$\eta_0 = \frac{\sqrt{\beta_T} - \sqrt{\beta_0}}{2}, \eta_T = \frac{\sqrt{\beta_T} + \sqrt{\beta_0}}{2}$$

To train the model for the reverse diffusion process, we need to optimize the network F_θ to predict the clean data z_0 from a noisy bridge sample z_t and the condition z_T. The objective function is defined as:

$$\mathcal{L}(F_\theta) = \mathbb{E}_{t,z_0,z_T}[\frac{1}{\sigma_t^2}\|F_\theta(z_t, t, z_T) - z_0\|^2] \tag{6}$$

Once the model is trained, it can be utilized to reconstruct the clean data from the degraded input z_T. The inference process involves solving the reverse-time bridge trajectory over a discretized schedule $0 = t_0 < t_1 < \cdots < t_N = T$. The number N and the schedule of sampling steps can be made independently of the original timesteps on which the bridge model is trained. The updating rule from t_{n+1} to t_n is described by DBIM [7] sampler:

$$\begin{aligned} z_{t_n} =& a_{t_n} z_T + b_{t_n}\hat{z}_0 + \sqrt{c_{t_n}^2 - \rho_n^2} \\ & \times \underbrace{\frac{z_{t_{n+1}} - a_{t_{n+1}} z_T - b_{t_{n+1}}\hat{z}_0}{c_{t_{n+1}}}}_{\text{predicted noise } \hat{\epsilon}} + \rho_n \epsilon, \epsilon \sim \mathcal{N}(0, \mathbf{I}). \end{aligned} \tag{7}$$

where $\hat{z}_0 = F_\theta(z_{t_{n+1}}, t_{n+1}, z_T)$ denotes the predicted clean data at time 0.

III. Experiments

The denoising module is implemented based on the original U-Net architecture and trained following the training strategy described in the original work [2]. The super-resolution module is implemented by combining VAE3 [5] with I2SB and DBIM sampler, and is trained using a learning rate of 1×10^{-4} for 300k iterations.

We compare our model with the classical GAN-based Real-ESRGAN and the advanced diffusion-based StableSR [8]. Both models are trained using the synthetic data described in Section II-B and follow the identical Real-ESRGAN degradation pipeline. All images are cropped to a resolution of 512×512 during training. All model mentioned are trained on a single NVIDIA RTX 3090 GPU with 24 GB memory, using a mini-batch size of 1.

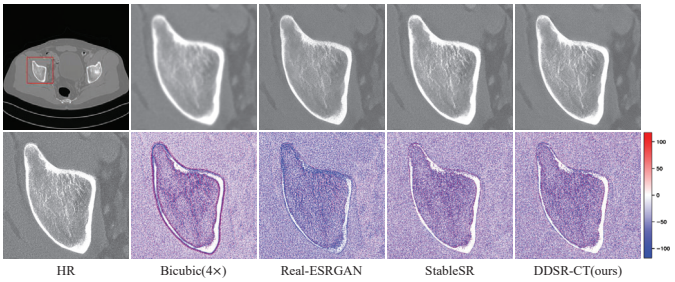


Fig. 2: Qualitative results of SR models on synthetic UHR CT test data. The display windows is [-1000,1000] HU. The red ROI is zoomed in for visual comparison. Difference maps (relative to the HR reference) are shown in the second row.

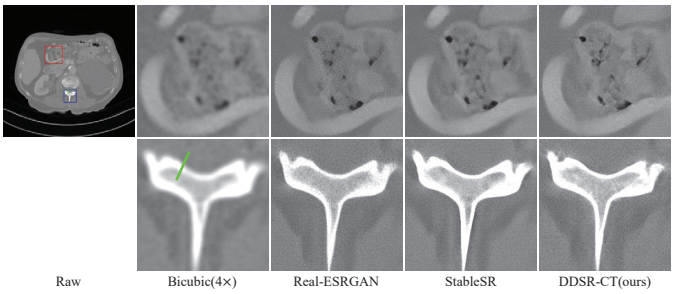


Fig. 3: Qualitative results of SR models on synthetic UHR CT eval data. The display windows is [-1000,1000] HU. The red and blue ROI is zoomed in for visual comparison.

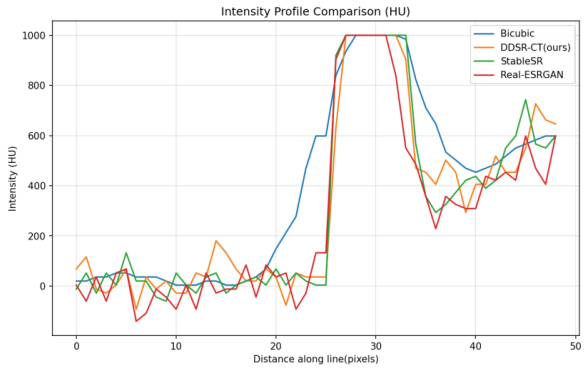


Fig. 4: Intensity profiles along the indicated green line from Figure 3 for different SR models.

Figure 2 presents qualitative comparisons on the synthetic UHR CT test set between our method and other SR models. As observed, Real-ESRGAN exhibits a tendency to overfit, leading to suboptimal detail recovery, whereas the proposed method achieves visual quality comparable to StableSR. Figure 3 further illustrates the results on the validation set, demonstrating the effectiveness of the proposed method in jointly suppressing noise and reducing blurring artifacts. The intensity profiles in Figure 4 further validate the superior edge-sharpening capability of our method. While Bicubic exhibits a gradual slope at the tissue interface, our method presents a much steeper gradient, comparable to StableSR and Real-ESRGAN.

Quantitative results are reported in Table I. While Bicubic interpolation maintains the highest PSNR/SSIM, it yields the worst LPIPS, failing to recover the detailed anatomical textures. Notably, our DDSR-CT significantly outperforms Real-ESRGAN, achieving a 5.6% reduction in LPIPS score.

TABLE I: Performance comparison of SR models on synthetic UHR CT test data

Method	PSNR↑	SSIM↑	LPIPS↓	Patch-Wise	Infer-time[2]
Bicubic	**32.409**	**0.874**	0.389	No	$< 0.01s$
Real-ESRGAN	31.710	0.847	0.154	No	$\sim 0.5s$
StableSR-200[1]	31.791	0.836	0.150	Yes	$\sim 293s$
DDSR-CT-20[1]	31.748	0.829	**0.145**	No	$\sim 5.5s$

[1] Number of diffusion sampling steps used during inference.
[2] Average inference time per slice, measured on a single NVIDIA RTX 3090 GPU.

Compared to StableSR, DDSR-CT maintains better LPIPS and CLIPIQA while achieving significant reduction in sampling steps (20 vs. 200). Furthermore, our approach avoids patch-wise inference, resulting in superior computational efficiency and improved practicality for CT imaging applications.

IV. Conclusion

In this paper, we presented DDSR-CT, a framework designed for UHR CT super-resolution. By isolating artifact suppression from resolution enhancement, our approach effectively recovers fine anatomical details while ensuring structural fidelity. Through the dedicated data synthesis pipeline, the framework alleviates the scarcity of UHR CT data. Furthermore, by performing accelerated diffusion bridge inference in latent space, our method processes a full 2048×2048 image within only 20 sampling steps. Experimental results demonstrate that DDSR-CT outperforms traditional GAN-based methods in perceptual quality and is comparable to the advanced diffusion model. Future work will focus on validating the framework across a broader range of clinical data.

References

[1] F. Yu, J. Gu, Z. Li, J. Hu, X. Kong, X. Wang, J. He, Y. Qiao, and C. Dong, "Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2024, pp. 25 669–25 680.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[3] B. Chen, S. Leng, L. Yu, D. Holmes III, J. Fletcher, and C. McCollough, "An open library of ct patient projection data," in *Medical Imaging 2016: Physics of Medical Imaging*, vol. 9783. SPIE, 2016, pp. 330–335.

[4] G.-H. Liu, A. Vahdat, D.-A. Huang, E. A. Theodorou, W. Nie, and A. Anandkumar, "I2SB: Image-to-image schrödinger bridge," *arXiv preprint arXiv:2302.05872*, 2023.

[5] P. Esser, S. Kulal, A. Blattmann, R. Entezari, J. Müller, H. Saini, Y. Levi, D. Lorenz, A. Sauer, F. Boesel *et al.*, "Scaling rectified flow transformers for high-resolution image synthesis," in *Forty-first international conference on machine learning*, 2024.

[6] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1905–1914.

[7] K. Zheng, G. He, J. Chen, F. Bao, and J. Zhu, "Diffusion bridge implicit models," *arXiv preprint arXiv:2405.15885*, 2024.

[8] J. Wang, Z. Yue, S. Zhou, K. C. Chan, and C. C. Loy, "Exploiting diffusion prior for real-world image super-resolution," *International Journal of Computer Vision*, vol. 132, no. 12, pp. 5929–5949, 2024.

Investigations on bone representation in voxelized patient models for estimation of scatter profiles in cone-beam CT

F. Noo, Z. Guo, M. B. Oktay, M. Manhart, J. DiNitto

***Abstract*—To improve the clinical value of cone-beam CT in both C-arm CT and conventional diagnostic CT, it is essential to improve the quantitative accuracy of attenuation values. Scatter is a major patient-specific, non-linear, physics effect that affects quantitative accuracy of the CT attenuation values when improperly corrected for. Recently, practical methods have emerged for patient-specific scatter correction. These methods rely on the use of voxelized patient models wherein bone representation poses challenges. In this work, we investigate the impact of bone representation on monoenergetic images and scatter profiles. Our early results indicates that many materials should be used to represent bones. However, this number needs not be proportional to the number of bone voxels; 200 materials may be largely sufficient with negligible impact on computational cost for Monte-Carlo transport with GEANT4. We also show that including or not bone marrow within the bone representation results in a small but non-negligible spatially-varying effect on scatter profiles.**

I. Introduction

Obtaining quantitatively accurate attenuation values in cone-beam CT is an essential, yet challenging, goal to improve the clinical value of C-arm CT, as well as that of conventional diagnostic CT. Patient-specific, non-linear, physics effects such as beam hardening and scatter are major factors that affect realization of this goal. In this work, we are interested in the representation of bones for patient-specific estimation of scatter profiles. See [1] for a recent investigation that we conducted on analyzing the effects of bone on patient-specific beam hardening errors within the lung parenchyma.

Two practical methods have recently emerged for patient-specific estimation of scatter profiles, one using the Boltzmann transport equation [2], and another one using deep learning convolution networks [3]. A common aspect of these methods is that realistic models of the patient anatomy are needed, either directly for the patient being imaged [2] or indirectly to produce reliable training data [3]. Such models are typically called "voxelized patient models".

A voxelized patient model is a volume of voxels regularly and finely distributed on a 3D Cartesian grid, where each voxel value specifies the patient tissue present at the spatial location of the voxel in terms of mass density and chemical composition. From such a model, it is possible, using tabulated values for atomic cross sections, to evaluate the linear attenuation coefficient for photon absorption at any photon energy, as well as to evaluate scatter and dose deposition events, all at any location within the patient. Voxelized patient models suitable for organ-specific estimation of CT radiation dose were first introduced in [4] and proved valuable for such estimation in many subsequent publications, including, e.g., [5].

Most often, the starting point used to create a voxelized patient model is a series of axial CT images obtained from a 120 kV scan. Although soft tissues can be well modeled, representation of bones from such images is particularly difficult. There are three issues that affect this representation. First, the spatial resolution is far too poor to rigorously separate bone material from bone marrow within the bones; only a composite material can be modeled. Second, there is no unique definition for such a composite material from a 120 kV scan. Third, a spatially varying composite material definition is necessary to reflect real patient anatomy.

In the DeMarco model [4], [5], bones were represented using four material options; these four materials shared the same chemical composition but differed in terms of mass density. In [6], the issue of bone representation was extensively discussed and two richer models were suggested, one in which each local bone value is modeled as cortical bone with a spatially varying mass density, and another one in which each local bone value is modeled as a linear combination of bone marrow and cortical bone, which induces spatial variations in both mass density and chemical composition. Although such models are much richer, it was also acknowledged that they could not be practically used with Monte-Carlo transport software like GEANT4 (used here) and MCNP (used in [4], [5]) as these programs are not amenable to efficiently using a number of materials proportional to the number of voxels. Categorization into a finite number of options was considered necessary, but the number of options can be much larger than that used in the DeMarco model.

In this work, we investigate bone representation using these various model options, with the two models of [6] referred to as the bone-only and the bone-mixture models along with an integer that specifies the number of categories the models are divided into.

II. Experimental set-up

A. Patient data selection

While we plan on expanding our investigations to cover various biological variables (such as sex and body mass index) and anatomical sites (such as the head and the pelvis), the investigations reported here are focused on chest CT images of a single female patient. The scan was acquired at 120 kV, and was reconstructed with a slice thickness of 3 mm. See Fig. 1

for visual appearance of the patient anatomy. To convert the HU values of the CT images into physical attenuation values, it was necessary to select an energy level that is representative of the source-detector energy spectrum involved. Using dual energy measurements through a 20 cm phantom water equivalent phantom without and with a bone insert, we established this energy level to be 78.5 keV.

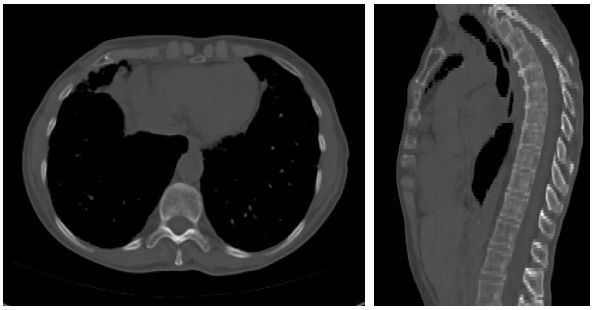


Fig. 1. Axial and sagittal slices from the CT scan used for creation of our voxelized patient models. Grayscale window: {width, level}={1800, 400} HU.

B. Voxelized patient models

We considered three voxelized patient models. The first model followed the approach of [4], [5]: all tissues were represented using 17 materials with material index 0 for air and material indices 13 through 16 for the four bone materials. The range of HU spanned by each material is shown in Fig. 2. In particular, the four bone materials were distributed over the range of $[152, 1560]$ HU; this wide range conveys the complexity of representing bone. Fig. 3 shows one sagittal slice through the voxelized patient model using color-coded display.

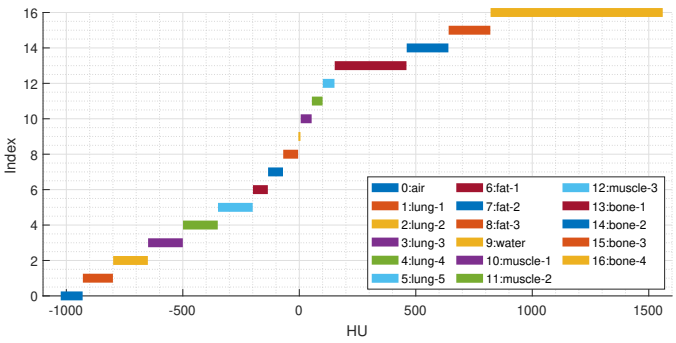


Fig. 2. Table of material indices and their HU range coverage for our first voxelized patient model, based on [4], [5]. The last four materials are those used to represent all voxels that cover the bones. These materials share the same chemical composition and thus only differ in terms of mass density.

The second and third models were modifications of the first model. We changed the description of the voxels covering the bones using the two suggestions of [6], which we refer to as the bone-only and the bone-mixture models. For these modifications, we considered categorization into 20, 200 and 2000 materials. Pre-testing assessed that 2000 materials incurs a 3% increase in computational cost for our scatter simulations

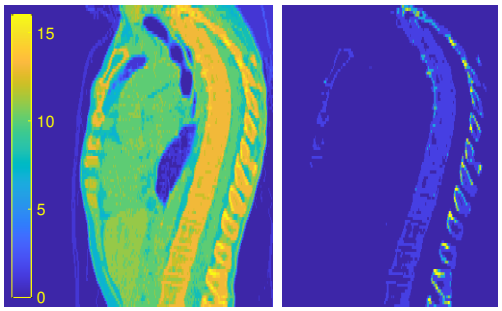


Fig. 3. Voxelized patient model using the approach of [4], [5] displayed using color coding. (Left) sagittal slice with all 17 materials color-coded over the $[0, 16]$ interval. (Right) the same sagittal slice showing only the bones with color-coding normalized from $[13, 16]$ to $[0, 1]$.

based on GEANT4. Fig. 4 shows one sagittal slice through these two modifications of the first voxelized patient model, again using color-coded display.

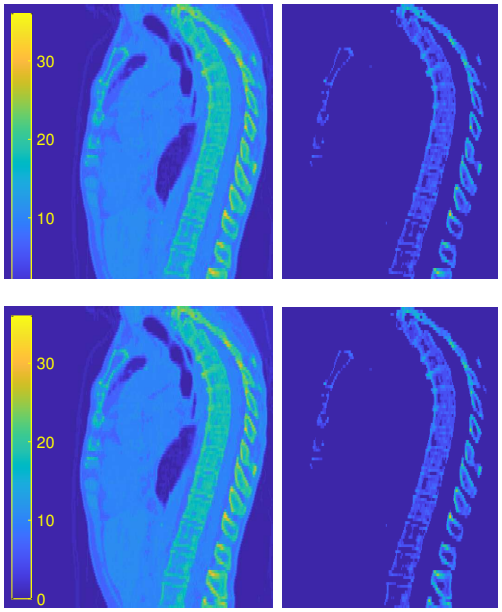


Fig. 4. Voxelized patient models using the two suggestions from [6] for bone representation, displayed using color coding. In this figure, the voxels containing bones were categorized into 20 materials. (Left) sagittal slice with all 37 materials color-coded over the $[0, 36]$ interval. (Right) the same sagittal slice showing only the bones with color-coding normalized to $[0, 1]$. (Top) bone-only model. (Bottom) bone-mixture model.

C. Scatter experiments

Scatter profiles were estimated using GEANT4. We simulated an 80 keV X-ray source randomly transmitting photons towards a 36 cm $\times$ 36 cm region in the detector plane, located at 120 cm from the X-ray source, while the source to center distance was 75 cm. The scatter events were recorded over a 42 cm $\times$ 42 cm region centered over the same point within the detector plane. The number of transmitted photons was $4,000,000$. The simulation was carried out for two source locations (anterior-posterior and lateral). To mitigate noise,

each simulation was repeated 100 times and their results were averaged over these 100 repetitions. If the 36 cm $\times$ 36 cm detector region were divided into pixels of size 0.64 mm, it could be said that 12 photons were, in average, transmitted towards each detector pixel. Hence, in our experiment, a count of 2 scattered photons hitting a location within the 42 cm $\times$ 42 cm region refers to a scatter-to-primary ratio of 0.17.

III. Results

A. Monochromatic images

To assess model accuracy, we converted the voxelized patient models into monochromatic CT images at 78.5 keV. The result is shown in Fig. 5. Comparing with Fig. 1, it is clear that the first model yields bone that are too attenuating compared to the original CT attenuation values, whereas the other two models yield attenuation values for the bones that are seemingly identical and more satisfying. For a more quantitative assessment, we examined a region of interest centered on a vertebra, as shown in Fig. 6. It can be seen that the two modified models provide highly similar results and these closely match the original CT images.

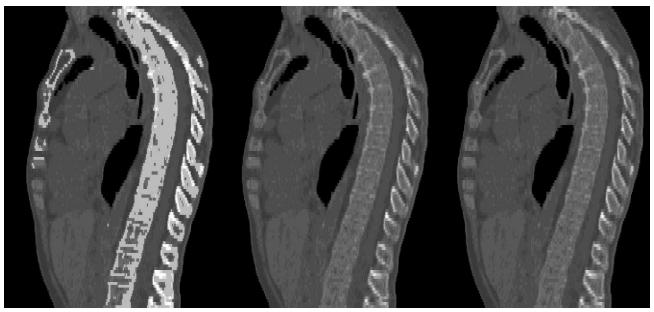


Fig. 5. Voxelized patient models convered into physical attenuation values at 78.5 keV. From left to right, the sagittal slices are from the DeMarco model, the bone-only model, and the bone-mixture model. Grayscale window: {width, level}={1800, 400} HU.

To appreciate the difference between the bone-only and the bone-mixture representation of voxels containing bones, it is necessary to examine images at other energy levels. Specifically, one might expect that the bone-only model results in higher attenuation values at lower keV level as more calcium is present in this model. Fig. 7 shows this expectation to be indeed correct. In a less predictable manner, Fig. 7 also shows that the bone-mixture model yields higher attenuation values at higher keV level; this difference is induced by the different chemical compositions involved.

B. Scatter profiles

Figure 8 shows a 2D histogram of scatter events observed on the detector. As expected, scatter has a smooth appearance. Nevertheless the variations can be non-negligible, particularly in the anterior-posterior view direction. Such variations are enough to create challenges for the problem of achieving quantitative attenuation values in image reconstruction, as illustrated in [7].

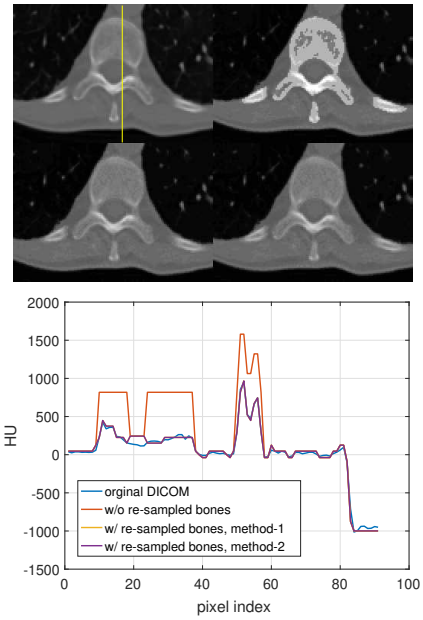


Fig. 6. (Top) Voxelized patient models converted into physical attenuation values at 78.5 keV, next to the original CT image, through which a yellow line is drawn. (Bottom) To more quantitatively compare accuracy of the models, we drew a profile along the yellow line. In the legend, w/o re-sampled bones refers to our first voxelized patient model, while method-1 and method-2 refers to our modified versions (bone only and bone-mixture resampling, respectively).

Fig. 9 shows the effect of using 20 versus 200 versus 2000 materials to represent the bones. It appears clear that 20 materials can be a little too small, but 200 materials provides converged plots with negligible impact on computational cost.

Fig. 10 shows the impact of choosing the bone-only versus the bone-mixture model for estimation of scatter. It can be seen that the overall shape of the profiles remain the same, but the magnitude is impacted by a spatially varying factor of about 4%.

IV. Conclusions

We presented early results investigating the effects of bone representation on scatter profiles estimated using Monte-Carlo transport through voxelized patient models. We observed that it is critical to use a large number of materials to represent the bones. However, this number needs not be proportional to the number of voxels in the model. In our experiments, 200 materials appear largely sufficient with the convenience that such a number has negligible impact on computational cost when using GEANT4. We also observed that there are small but non-negligible effects resulting from including or not bone marrow in the bone representation. More investigations are needed to see how these results are impacted by biological variables, anatomical sites, and scanning parameters.

References

[1] Guo, Z., Noo, F., Stierstorfer, K., and McNitt-Gray, M., "Impact of beam hardening on CT attenuation values of the lung parenchyma: analysis using computer simulations with voxelized patient models," in [*Medical Imaging 2025: Physics of Medical Imaging*], **13405**, 872–876, SPIE (2025).

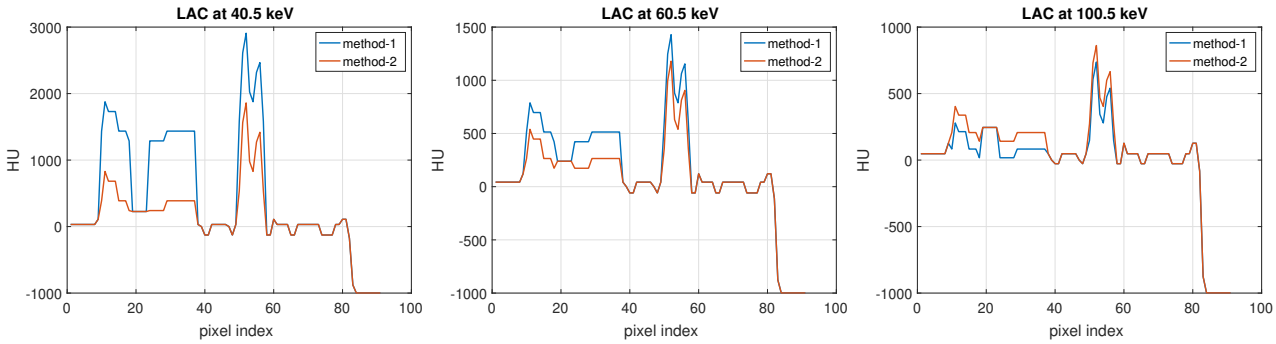


Fig. 7. Comparison of profiles for bone-only (method-1) and bone-mixture (method-2) voxelized patient models at energy levels above and below 78.5 keV. From left to right, the energy levels are 40.5 keV, 60.5 keV and 100.5 keV.

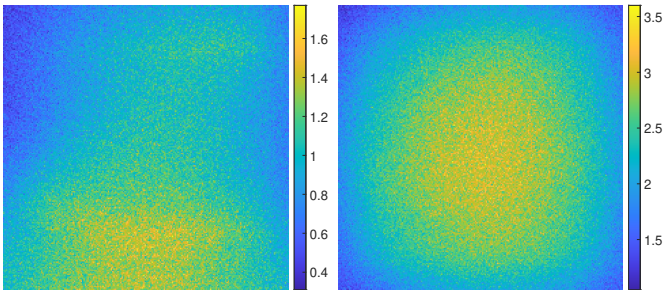


Fig. 8. Histogram of scattered events collected in the detector plane, as obtained using the voxelized patient model based on the bone-only model using 2000 materials to represent the bones. (Left) anterior-posterior beam transmission. (Right) lateral beam transmission.

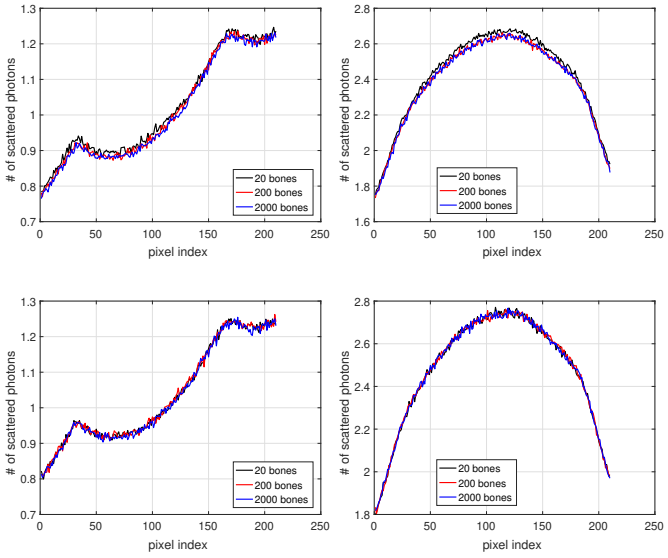


Fig. 9. Profiles through the histogram of scattered events, obtained by averaging all columns of the histogram. (Left) anterior-posterior beam transmission. (Right) lateral beam transmission. (Top) bone-only model. (Bottom) bone-mixture model. For both models, the plots show the effect of increasing the number of categories used to describe the bone materials, from 20 to 200 and then 2000.

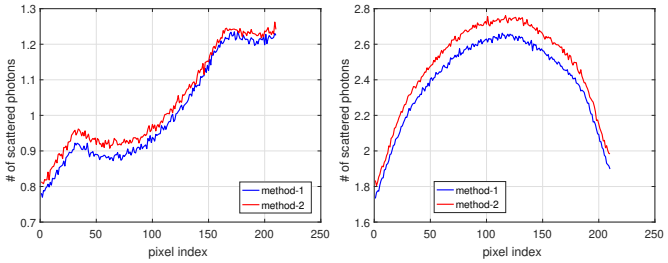


Fig. 10. Profiles through the histogram of scattered events, obtained by averaging all columns of the histogram. The bones were categorized into 200 materials. These plots compare the effect of using bone-only (method-1) versus bone-mixture (method-2) voxelized patient models.

[2] Wang, A., Maslowski, A., Messmer, P., Lehmann, M., Strzelecki, A., Yu, E., Paysan, P., Brehm, M., Munro, P., Star-Lack, J., et al., "Acuros CTS: A fast, linear boltzmann transport equation solver for computed tomography scatter–part ii: System modeling, scatter correction, and optimization," *Medical physics* **45**(5), 1914–1925 (2018).

[3] Maier, J., Eulig, E., Vöth, T., Knaup, M., Kuntz, J., Sawall, S., and Kachelrieß, M., "Real-time scatter estimation for medical CT using the deep scatter estimation: method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation," *Medical physics* **46**(1), 238–249 (2019).

[4] DeMarco, J., Solberg, T., and Smathers, J. B., "A CT-based monte carlo simulation tool for dosimetry planning and analysis," *Medical physics* **25**(1), 1–11 (1998).

[5] DeMarco, J., Cagnon, C., Cody, D., Stevens, D., McCollough, C. H., O'Daniel, J., and McNitt-Gray, M., "A monte carlo based method to estimate radiation dose from multidetector CT (MDCT): cylindrical and anthropomorphic phantoms," *Physics in medicine & biology* **50**(17), 3989 (2005).

[6] Schneider, W., Bortfeld, T., and Schlegel, W., "Correlation between CT numbers and tissue parameters needed for Monte Carlo simulations of clinical dose distributions," *Physics in Medicine & Biology* **45**(2), 459 (2000).

[7] Haase, V., Noo, F., Stierstorfer, K., and McNitt-Gray, M., "Towards enabling better evaluation of emphysema: investigation into a method to reduce the variability of lung ct attenuation values due to patient size and positioning," in [*Medical Imaging 2022: Physics of Medical Imaging*], **12031**, 288–292, SPIE (2022).

Deep Low-Rank Subspace Learning for 5D Myocardial Perfusion Photon-Counting CT Denoising

Meng Xiao, Xiao Meng, Jiangjun Peng, Jingyi Liao, Yue Liu, Hao Wang, Zhaoying Bian, Jianhua Ma, Dong Zeng

***Abstract*—Five-dimensional (5D, spatial-spectral-temporal) myocardial perfusion photon-counting CT (5D-MP-PCCT) offers clear physiological insights but suffers from severe noise due to limited photon counts across high-dimensional data. Conventional denoising methods typically treat these dimensions independently, failing to exploit the intrinsic structural correlations. To address this, we propose a Subspace Coefficient Image Denoising (SCID) network that synergizes low-rank tensor decomposition with deep residual learning. The core innovation lies in applying Singular Value Decomposition (SVD) along spatial or temporal dimensions to decompose the noisy 5D tensor into an orthogonal basis and a set of Representative Coefficient Images (RCIs). Unlike standard image-domain filtration, our method trains a deep Residual Learning-based Denoising CNN specifically on these RCIs, effectively decoupling the principal anatomical/functional information from high-frequency noise components. We extensively evaluate the topological impact of decomposition strategies on in vivo mouse cardiac PCCT data. Results demonstrate that SCID significantly outperforms conventional 2D slice-wise methods. Notably, the temporal SVD strategy yields the superior performance, i.e., PSNR gain of ~2.38 dB over spatial SVD, confirming that exploiting temporal redundancy is critical for preserving dynamic perfusion fidelity.**

***Index Terms*—Singular value decomposition, coefficient image denoising, myocardial perfusion, photon-counting CT**

I. Introduction

FIVE-dimensional (5D) myocardial perfusion photon-counting CT (5D-MP-PCCT) represents a paradigm shift in preclinical cardiovascular imaging. By enabling simultaneous multi-energy spectral separation and high-temporal-resolution dynamic scanning, 5D-MP-PCCT provides unprecedented insights into myocardial perfusion and functional hemodynamics. However, this advanced capability introduces a fundamental challenge, i.e., the "dimensionality-dose trade-off". The reconstruction of 5D volumetric data (3D spatial + 1D spectral + 1D temporal) inevitably leads to photon starvation, particularly in high-energy bins and narrow temporal windows required to freeze cardiac motion. Consequently, the reconstructed images are frequently degraded by severe quantum noise, compromising the quantification of iodine concentration and myocardial defects.

The standard approach to mitigate this issue typically involves post-reconstruction denoising. However, conventional analytical methods, such as Filtered Back-Projection (FBP) followed by spatial filtering, often sacrifice fine anatomical details to suppress noise. While iterative reconstruction algorithms incorporating regularization priors (e.g., Total Variation or Low-Rank constraints) [1] [2] have shown success, they are computationally prohibitive for high-dimensional 5D datasets. For instance, model-based iterative reconstruction methods have integrated weighted tensor low-rankness with learnable sparse representation [3], and another work has employed a prior image – induced diffusion tensor within a statistical iterative framework [4]. Nevertheless, these methods are often limited by high computational cost or the need for matched system geometry.In recent years, deep learning (DL) has dominated the landscape of CT denoising. Despite their superior performance, most existing DL architectures (e.g., U-Net [5], DnCNN [6]) treat dynamic or spectral CT data as a sequence of independent 2D slices or 3D volumes. This "slice-wise" strategy is sub-optimal because it neglects the rich intrinsic correlations embedded across the energy and time dimensions. By failing to exploit the global low-rank structure of the 5D tensor, these methods

This work was supported in part by the National Natural Science Foundation of China under Grant 62571228, the National Key R&D Program of China under Grant 2024YFA1012000, Beijing Natural Science Foundation under Grant Z250002, and Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515140146. (Corresponding author: Jingyi Liao, Zhaoying Bian.)

Meng Xiao, Xiao Meng, Jingyi Liao, Yue Liu, Hao Wang, Zhaoying Bian, Dong Zeng are with the School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China (e-mail: xiao_m@smu.edu.cn, mx_1125@smu.edu.cn, 1jy@smu.edu.cn, 18888539603@163.com , wang_hao@smu.edu.cn, zybian@smu.edu.cn and zd1989@smu.edu.cn)

Jiangjun Peng is with the School of Mathematics and Statistics, Northwestern Polytechnical University, Xi'an 710129, China (e-mail: andrew.pengjj@gmail.com).

Jianhua Ma is with the School of Life Science and Technology, Xi'anJiaotong University, Xi'an 710049, China (e-mail: jhma@smu.deu.cn)

limit their denoising potential and may introduce temporal flickering or spectral distortions.

In this work, borrowed from [7], we propose a novel Subspace Coefficient Image Denoising (SCID) network tailored for 5D-MP-PCCT. Unlike conventional methods that denoise the original image space directly, the proposed SCID integrates mathematical subspace decomposition with data-driven deep learning. Specifically, we perform Singular Value Decomposition (SVD) along the temporal or spatial dimensions to extract Representative Coefficient Images (RCIs), which condense the essential anatomical and dynamic information. A dedicated deep residual network then denoises these RCIs before reconstructing the full 5D volume via the orthogonal basis. The main contributions of this paper are:

1. We formulate the 5D-MP-PCCT denoising problem within a low-rank decomposition framework, effectively reducing the dimensionality of the noise mapping task while preserving spectral-temporal fidelity.
2. We systematically explore and compare SVD strategies along both spatial and temporal dimensions to identify the optimal representation for dynamic cardiac perfusion.
3. Experiments on preclinical 5D-MP-PCCT datasets demonstrate that the proposed SCID achieves promising results.

II. Materials And Methods

A. Low-Rank Subspace Decomposition

The core of the proposed SCID is the exploitation of global multidimensional correlations within cardiac PCCT data. We formulate the denoising problem using a low-rank subspace decomposition framework. Given the 5D tensor $Y \in R^{H\times W\times S\times E\times T}$, we employ SVD to separate the signal into a low-dimensional signal subspace and a noise subspace. We investigate two distinct decomposition strategies to leverage different structural priors.

The first strategy is spatial decomposition. We rearrange the tensor to prioritize anatomical correlations across the axial direction. By treating the spatial volume as a sequence of vectorized slices, we unfold the tensor into a matrix representation $Y_s \in R^{H\times W\times S}$, where S represents the number of axial slices across the entire cardiac volume. We then perform SVD to obtain:

$$Y_s \approx U_s \Sigma_s V_s^T$$

where $U_s \in R^{(H\times W)\times R_s}$ contains the spatial RCIs, which capture the principal anatomical structures across the volume. $V_s \in R^{S\times R_s}$ represents the slice-wise basis vectors. and Σ_s is a diagonal matrix of singular values. We set the rank R_s = 8 to retain dominant structural information while discarding high-frequency noise components.

The second strategy is temporal decomposition. This approach targets the dynamic behavior of the heart. We restructure the data to form a matrix $Y_T \in R^{H\times W\times T}$, where T represents the cardiac phases at a specific spatial location. The decomposition yields:

$$Y_t \approx U_t \Sigma_t V_t^T$$

where $U_t \in R^{(H\times W)\times R_t}$ contains the temporal RCIs, which capture the principal temporal variations across cardiac phases. Each column of U_t represents a coefficient image of size H×W. $V_t \in R^{T\times R_t}$ contains the phase-wise basis vectors, and Σ_t is a diagonal matrix of singular values. These images encode the primary motion and perfusion patterns, effectively isolating dynamic physiological signals from stochastic noise. We set R_t = 6 for this strategy. This dimension-specific factorization allows us to process the high-dimensional data in a compact subspace where signal and noise are more separable than in the original image domain.

B. The Description of SCID

To produce high-fidelity images, we designed a learning-based denoising framework that operates specifically on the extracted RCIs based on the previous study [7]. Although RCIs resemble natural images structurally, they possess distinct statistical intensity distributions. Therefore, we adapted a Residual Learning-based Denoising CNN (DnCNN) to process these coefficient maps.

The proposed SCID comprises 15 layers and utilizes residual learning to predict the noise manifold within the input RCIs. We train the network to minimize the $L1$ loss between the reconstructed images and the ground truth. The reconstruction process involves combining the network-denoised RCIs with their corresponding orthogonal basis matrices (V_s or V_t) and singular values. This hybrid approach effectively merges the mathematical robustness of low-rank models with the non-linear feature extraction capabilities of deep learning, ensuring the preservation of both fine anatomical details and dynamic perfusion trends. The proposed method framework is illustrated in Fig. 1.

C. Dataset Acquisition and Preprocessing

This study utilized in vivo mouse myocardial perfusion data acquired via a photon-counting CT (PCCT) system. We conducted all animal procedures in strict adherence to protocols approved by the Institutional Animal Care and Use Committee. The imaging protocol involved an intravenous injection of iodixanol followed by dynamic scanning to capture myocardial perfusion. The system operated at a tube voltage of 70 kVp and a current of 200 μA. We set energy thresholds at 21 keV and 31 keV, which generated

dual-energy bins of 21-70 keV and 31-70 keV. To ensure high temporal resolution sufficient for cardiac analysis, we performed dynamic scanning at 40 frames per second over ten cardiac cycles. This yielded ten temporal phases per energy channel.

Following data acquisition, we reconstructed the projection data using the FDK algorithm. We applied median and mean filtering in the projection domain prior to reconstruction to mitigate initial projection noise. The resulting dataset formed a five-dimensional (5D) tensor with dimensions corresponding to axial slices, height, width, energy channels, and time phases (189 × 1024 × 1024 × 3 × 10). To facilitate supervised training, we generated paired datasets by modeling realistic noise in the projection domain. We allocated data from eight mice for network training and reserved two independent mice for testing. For computational efficiency during the experimental phase, we downsampled images to 512 × 512 and performed center-cropping to a resolution of 256 × 256 pixels.

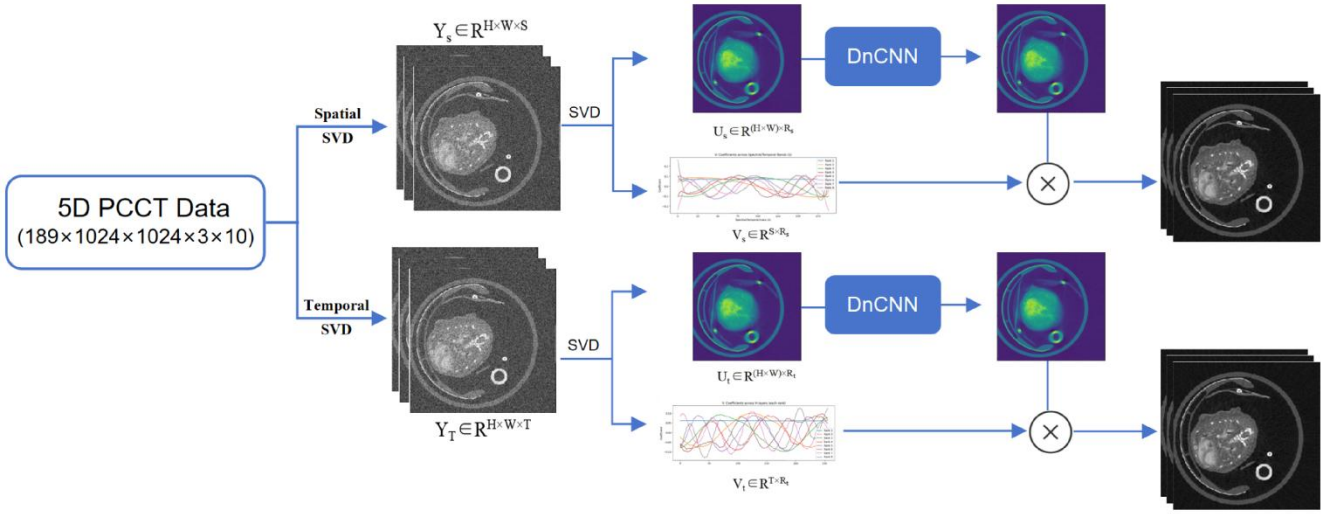


Fig. 1. The framework of the proposed method.

III. Experiments And Results

A. Training Strategies

The proposed SCID is validated using the in vivo mo 5D cardiac PCCT dataset, partitioning eight subjects for training and reserving two for independent testing. To rigorously assess the benefits of exploiting high-dimensional structure, we designed three distinct training strategies. ①2D Slice-wise (Baseline): The 5D data ($H \times W \times S \times E \times T$) were fully unfolded into independent 2D image slices, reshaping them to ($H \times W, 1$). This method processes spatial and temporal frames in isolation and ignores the correlations between them. ② Spatial SVD: To capture anatomical continuity along the axial direction, we group data by energy e, and time t into volumes ($H \times W \times S$), then apply SVD along the slice dimension S to obtain $R_s = 8$ representative coefficient images (RCIs). ③ Temporal SVD: To model dynamic perfusion patterns, we group data by slice s and energy e into temporal sequences ($H \times W \times T$), then apply SVD along the time dimension T.

All strategies employ the same deep residual denoising network (DnCNN) to ensure that performance differences arise solely from the data representation rather than the network architecture. Quantitative performance is evaluated using peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), relative dimensionless global error (ERGAS), and root mean square error (RMSE).

B. Experimental Results

1) Subspace Decomposition Analysis

We first validate the information compression capability of the proposed framework by examining the spectral components of a representative test sample. Fig. 2 displays the representative coefficient images (RCIs) extracted via spatial SVD alongside their corresponding slice-wise basis vectors. The decomposition successfully concentrates the majority of the anatomical signal energy into the initial low-rank components. Conversely, noise distributions are pushed toward the higher-order components. This energy compaction confirms that the low-rank assumption holds for myocardial perfusion data and provides a solid mathematical foundation for separating signal from noise using the proposed network.

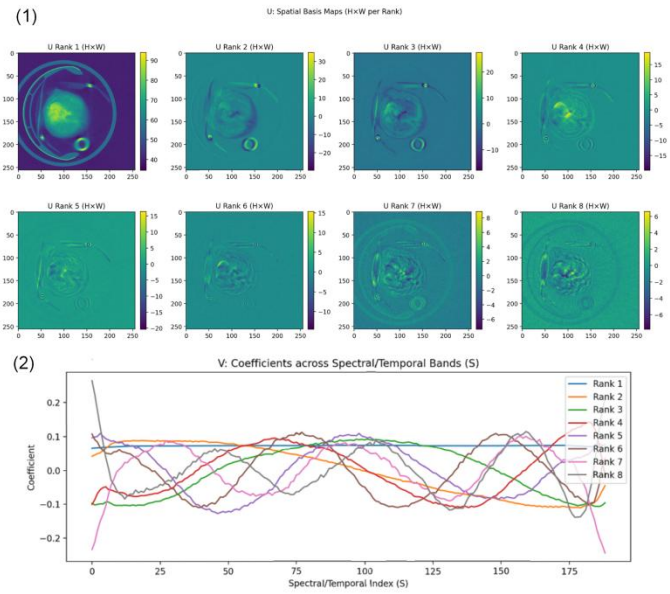


Fig. 2.(1) Representative coefficient images (RCIs, columns of U) extracted via spatial SVD, arranged in descending order of singular values (left to right). (2) Corresponding slice- wise basis vectors (rows of V).

2) Visual Quality Assessment

Fig. 3 presents a comparative visualization of the denoising results against the noisy input and the ground truth. The baseline 2D slice-wise method suppresses noise to a certain degree but compromises diagnostic quality by blurring fine structural details. This over-smoothing occurs because the method treats each image independently and ignores the intrinsic correlations between slices or time points. In contrast, the subspace-based methods preserve anatomical boundaries more effectively. Both the Spatial and Temporal SVD strategies maintain the sharpness of the myocardial wall and contrast-filled chambers while removing graininess. The Temporal SVD approach specifically yields the best visual result with minimal artifacts, suggesting it better differentiates the dynamic perfusion signal from random quantum noise.

3) Quantitative Performance

Table I summarizes the quantitative evaluation metrics for the denoising results of the three methods across the entire test set.

Table I

The quantitative evaluation metrics for the denoising results of the three methods are presented in the form of mean ± variance.

	Noisy	2d	Spatial SVD	Temporal SVD
PSNR	23.00± 4.64	26.21± 5.60	29.01± 2.39	31.39± 4.30
SSIM	0.61± 0.08	0.85± 0.06	0.86± 0.02	0.88± 0.02
ERGAS	34.31± 9.46	21.71± 5.63	22.34± 6.55	18.61± 5.71
RMSE	20.42± 3.32	8.88± 3.95	8.27± 1.43	6.96± 2.36

Note: Higher PSNR is better, SSIM closer to 1 is better, and lower ERGAS and RMSE are better.

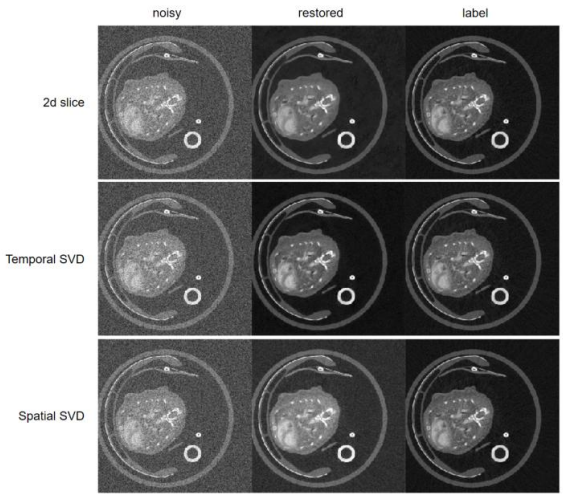


Fig. 3. presents the denoising results of the same test sample processed by three different models. From top to bottom, the rows correspond to the 2D Slice-wise, Spatial SVD, and Temporal SVD methods, respectively. Within each row, the images from left to right show the noisy input, network-denoised result, and ground truth.

All three deep learning methods improve image quality compared to the noisy input, but the subspace decomposition strategies demonstrate a clear advantage over the 2D baseline. The Temporal SVD strategy achieves the most robust performance across all metrics. It surpasses the 2D method and the Spatial SVD method by approximately 5.18 dB and 2.38 dB in PSNR, respectively. Furthermore, it attains the highest structural similarity (SSIM of 0.88) and the lowest error rates (ERGAS and RMSE). Although the temporal dimension ($T = 10$) implies less data redundancy than the spatial dimension ($S = 189$), the time-varying signals in perfusion imaging exhibit stronger statistical regularity due to the consistent physiological wash-in and wash-out of the contrast agent. Consequently, decomposing along the temporal dimension provides a more effective basis for isolating dynamic physiological information from stochastic noise. These results confirm that exploiting temporal correlations is critical for high-fidelity reconstruction in 5D-MP-PCCT.

IV. Conclusion

Accurate assessment of myocardial perfusion in PCCT is frequently hindered by quantum noise, particularly within narrow high-energy bins. We addressed this fundamental limitation by introducing the SCID network, a framework that integrates low-rank tensor theory with deep residual learning. Rather than processing isolated slices, our approach exploits the intrinsic correlations embedded in the 5D data volume through SVD. By extracting and processing

RCIs, the proposed SCID compacts the essential signal energy into a low-dimensional subspace, effectively separating anatomical and functional information from the noise floor. Validation on in vivo datasets confirms that this strategy significantly outperforms conventional image-domain filtration in terms of both structural fidelity and quantitative accuracy. Notably, the temporal decomposition scheme achieved the highest performance, as it aligns with the dynamic physiological changes of the cardiac cycle to preserve rapid perfusion variations. This work establishes that utilizing multi-dimensional structural priors is essential for ensuring high-quality reconstruction in 5D-MP-PCCT.

REFERENCES

[1] S. V. Mohd Sagheer and S. N. George, "Denoising of low-dose CT images via low-rank tensor modeling and total variation regularization," Artif. Intell. Med., vol. 94, pp. 1 – 17, 2019.

[2] Y. Tao, G.-H. Chen, T. A. Hacker, A. N. Raval, M. S. Van Lysel, and M. A. Speidel, "Low dose dynamic CT myocardial perfusion imaging using a statistical iterative reconstruction method," Med. Phys., vol. 41, no. 7, p. 071914, Jul. 2014.

[3] Y. Zhang, D. Zeng, Z. Bian, H. Lu, and J. Ma, "Weighted tensor low-rankness and learnable analysis sparse representation model for texture preserving low-dose CT reconstruction," IEEE Trans. Comput. Imaging, vol. 7, pp. 321 – 336, 2021.

[4] S. Niu, H. Liu, M. Zhang, M. Wang, J. Wang, and J. Ma, "Iterative reconstruction for low-dose cerebral perfusion computed tomography using prior image induced diffusion tensor," Phys. Med. Biol., vol. 66, no. 11, p. 115024, Jun. 2021.

[5] R. Yan, Y. Liu, Y. Liu, L. Wang, R. Zhao, Y. Bai, and Z. Gui, "Image denoising for low-dose CT via convolutional dictionary learning and neural network," IEEE Trans. Comput. Imag., vol. 9, pp. 83 – 93, 2023, doi: 10.1109/TCI.2023.3241546.

[6] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142 – 3155, Jul. 2017, doi: 10.1109/TIP.2017.2662206.

[7] J. Peng, H. Wang, X. Cao, Q. Zhao, J. Yao, H. Zhang, and D. Meng, "Learnable representative coefficient image denoiser for hyperspectral image," IEEE Trans. Geosci. Remote Sens., vol. 62, pp. 1 – 16, 2024, doi: 10.1109/TGRS.2024.3357981.

Accelerating Low-Frequency Convergence for Limited-Angle DBT via Two-Channel Fidelity in PDHG

Taro Iyadomi, Ricardo Parada, Anna Kim, Lily Jiang, Emil Sidky, and William Chang

***Abstract*—Reconstruction in limited-angle digital breast tomosynthesis (DBT) suffers from slow convergence of low spatial-frequency components when using weighted data-fidelity terms within primal–dual optimization. We introduce a two-channel fidelity strategy that decomposes the sinogram residual into complementary low-pass and high-pass bands using square-root Hanning (Hann$^{1/2}$) filter families, each driven by an independent ℓ_2-ball constraint and dual update in the PDHG (Chambolle–Pock) algorithm with He–Yuan predictor–corrector relaxation. By assigning a larger dual step-size and slightly looser tolerance to the low-frequency channel, the method delivers stronger per-iteration correction to the near-DC band without violating global PDHG stability. Experiments on a 2D digital breast phantom across multiple resolutions demonstrate that the two-channel approach yields 19%–61% RMSE improvement over the single-channel baseline, with larger gains at coarser discretizations where problem conditioning is more favorable, supporting more balanced spectral convergence in clinically realistic limited-angle regimes.**

***Index Terms*—Computed tomography, Breast tissue, Sparse matrices, Filter banks**

I. Introduction

Limited angular range (LAR) scanning has been widely studied in the literature, but only recently have discrete-to-discrete models demonstrated stable inversion at clinically relevant scan arcs. Building on the observation that directional sparsity improves LAR recovery, Zhang et al. [2021] introduced a globally convex formulation using directional total variation (DTV), enabling robust optimization-based reconstruction.

We introduce a sparsity-constrained optimization framework for LAR DBT that combines directional TV penalties with pixel/voxel sparsity, enabling practical parameterization when true DTV values are unknown. To characterize the regime in which the proposed two-channel fidelity approach is most effective, we systematically vary image discretization to modulate problem conditioning and quantify the resulting reconstruction benefit.

II. Preliminary

We consider limited angular range (LAR) digital breast tomosynthesis (DBT) in a discrete-to-discrete setting. Let f be the discretized attenuation image, g is the measured sinogram data, and X is the system matrix that encodes fan-beam (2D) or cone-beam (3D) projection. The idealized data model is

$$g = Xf \tag{1}$$

In the DBT geometry, the detector spans the in-plane directions (x, y), while the z-direction is perpendicular to the (fixed) detector plane and is referred to as the depth direction. At clinically typical scanning arcs, the resulting linear system is generally underdetermined at useful (e.g., isotropic) discretizations, motivating the use of convex regularization.

A. Directional Finite Differences and Directional Total Variation

Let $\partial_x, \partial_y, \partial_z$ denote finite-difference operators approximating derivatives along the corresponding axes. The associated directional total-variation (DTV) seminorms are $\|\partial_x f\|_1$, $\|\partial_y f\|_1$, and $\|\partial_z f\|_1$ (in 2D studies, the ∂_y term is omitted). Distinguishing in-plane and depth directional derivatives is particularly important in LAR/DBT because the data support and conditioning can differ substantially across directions.

B. Constrained Sparsity-regularized Reconstruction

Zhang et al. [2021] showed that accurate image reconstruction from LAR data can be achieved using non-smooth convex optimization with directional total variation (DTV). Their key problem minimizes a least-squares (LSQ) data-fidelity term subject to separate ℓ_1 bounds on the in-plane and depth finite-difference derivatives, forming DTV-constrained LSQ (DTV-LSQ):

$$\begin{aligned} f^* &= \arg\min_f \ \frac{1}{2}\|g - Xf\|_2^2 \\ \text{s.t.} \quad & \|\partial_x f\|_1 \le \gamma_x, \ \|\partial_z f\|_1 \le \gamma_z. \end{aligned} \tag{2}$$

Constraining DTV is more effective than the standard total variation (TV) constraint based on gradient sparsity. DTV constraints are convenient for phantom studies where true DTV values are computable, but for real digital breast tomosynthesis (DBT) data the bounds must be estimated. The present work reformulates the problem with improved parameterization and design choices to support isotropic-resolution reconstruction in DBT.

William Chang was with the Department of Mathematics, UCLA, CA, 90007, e-mail: chang314@g.ucla.edu.

Emil Sidky was with the Department of Radiology, University of Chicago, IL 60637

Manuscript received January 27, 2026.

III. The Single Channel Approach

In this section, we describe the single channel approach proposed by Sidky et al. [2025]. Let X be the discrete projection operator for a limited-angle DBT scan, g the measured sinogram, and $f \geq 0$ the unknown volume. The paper adopts a *data-discrepancy constrained, sparsity-regularized* model:

$$\begin{aligned} \min_{f\geq 0} \ & \alpha_x\|\partial_x f\|_1 + \alpha_y\|\partial_y f\|_1 + \alpha_z\|\partial_z f\|_1 + \beta\|f\|_1 \\ & \text{s.t.} \quad \left\|R[c]\,(g - Xf)\right\|_2 \leq \varepsilon\sqrt{\text{size}(g)}, \end{aligned} \tag{3}$$

where $R[c]$ is a square-root Hanning (Hann$^{1/2}$) filter along detector channels with cutoff parameter c; $\partial_\bullet$ are finite differences (directional TV). The resulting problem is convex but non-smooth (indicator constraint, ℓ_1 penalties), and is solved with a primal-dual hybrid gradient (PDHG / Chambolle–Pock) method with a He–Yuan predictor–corrector (relaxation) using $\rho \approx 1.75$ He and Yuan [2012]. The limited-angle regime reports running ~ 2000 PD iterations to reach accurate convergence.

There are two additional design choices emphasized in the paper: (i) using $R[c]$ *inside the data-constraint* to de-emphasize low spatial-frequency inconsistencies; (ii) including an optional voxel-sparsity term $\beta\|f\|_1$ in addition to DTV.

a) Notation for PDHG: With $A := R[c]\,X$ and the residual $r := g - Xf$, the data-constraint is $\|Af - R[c]\,g\|_2 \leq \varepsilon\sqrt{|g|}$. The PDHG dual variable y lives in the sinogram space; the dual step is a projection onto an ℓ_2 ball; the primal step applies the adjoint $A^\top = X^\top R[c]^\top$ followed by the proximal of the sum of ℓ_1 terms and the indicator of $f \geq 0$.

IV. Why low spatial frequencies converge slowly

The data-fidelity term can be written as $\frac{1}{2}\|W(g - Xf)\|_2^2$, where $W \equiv R[c]$. In the detector-frequency domain, this weighting strongly attenuates low spatial frequencies. Under a linearized analysis of the PDHG update, the correction applied to a residual Fourier mode at spatial frequency k scales approximately with the squared singular value s_k^2 of the weighted system operator WX. In limited-angle DBT, the singular values s_k are already small for low k due to the severe ill-conditioning of X. The additional Hann$^{1/2}$ weighting W further shrinks these low-k singular values by design, creating a large spectral gap between frequency bands.

This disparity in singular values destabilizes convergence behavior. Because PDHG must satisfy the global stability condition $\tau\sigma\|WX\|^2 < 1$ (Theorem 1 of Chambolle and Pock [2011]), the allowable step sizes are dictated by the *largest* singular values, which reside in the mid- and high-frequency bands. Step sizes that stabilize those bands therefore provide insufficient drive for near-DC modes, yielding extremely weak contraction factors for low spatial frequencies. Increasing the global step sizes to accelerate DC correction risks divergence, since PDHG cannot apply frequency-selective step-size amplification in its standard form. Consequently, mid- and high-frequency image components converge quickly, while low spatial frequencies decay slowly and dominate the tail of the iterations. This persistent low-frequency drift reveals a fundamental mode-dependent convergence bottleneck that motivates split-frequency or multi-dual fidelity strategies.

Algorithm 1 PDHG with split fidelity channels

Inputs: $g, X, R_{\text{hi}}, R_{\text{lo}}, \alpha_{x,y,z}, \beta, \varepsilon_{\text{hi}}, \varepsilon_{\text{lo}}, \sigma_{\text{hi}}, \sigma_{\text{lo}}, \tau, \rho$
Initialize f^0, $\bar{f}^0 = f^0$, $y_{\text{hi}}^0 = y_{\text{lo}}^0 = 0$
for $k = 0, 1, 2, \ldots$ **do**
  $y_{\text{hi}}^{k+1} \leftarrow \text{Proj}_{B_{\text{hi}}}\left(y_{\text{hi}}^k + \sigma_{\text{hi}} R_{\text{hi}}(g - X\bar{f}^k)\right)$
  $y_{\text{lo}}^{k+1} \leftarrow \text{Proj}_{B_{\text{lo}}}\left(y_{\text{lo}}^k + \sigma_{\text{lo}} R_{\text{lo}}(g - X\bar{f}^k)\right)$
  $f^{k+1} \leftarrow \text{prox}_{\tau\left(\alpha_x\|\partial_x\cdot\|_1 + \alpha_y\|\partial_y\cdot\|_1 + \alpha_z\|\partial_z\cdot\|_1 + \beta\|\cdot\|_1\right) + \mathbb{I}\{f\geq 0\}}\left(f^k - \tau X^\top (R_{\text{hi}}^\top y_{\text{hi}}^{k+1} + R_{\text{lo}}^\top y_{\text{lo}}^{k+1})\right)$
  $\bar{f}^{k+1} \leftarrow f^{k+1} + \rho(f^{k+1} - f^k)$ ▷ He–Yuan relaxation
end for

V. Two-channel fidelity for spectrally balanced PDHG

Single-dual weighted fidelity in limited-angle tomographic inversion entangles spatial-frequency bands with sharply different conditioning, which can destabilize convergence when the optimizer attempts to correct near-DC error using a step size dictated by high-frequency modes. To retain the noise robustness of the Hann$^{1/2}$ weighting while providing a stronger corrective signal to low spatial frequencies, we decompose the sinogram residual into complementary spectral channels and enforce independent consistency constraints:

$$\begin{aligned} \min_{f\geq 0} \ & \alpha_x\|\partial_x f\|_1 + \alpha_y\|\partial_y f\|_1 + \alpha_z\|\partial_z f\|_1 + \beta\|f\|_1 \\ & \text{s.t.} \ \begin{cases} \|R_{\text{hi}}(g - Xf)\|_2 \leq \varepsilon_{\text{hi}}\sqrt{|g|}, \\ \|R_{\text{lo}}(g - Xf)\|_2 \leq \varepsilon_{\text{lo}}\sqrt{|g|}. \end{cases} \end{aligned} \tag{4}$$

We select both filters from the square-root Hanning family. The high-frequency channel preserves the original cutoff, $R_{\text{hi}} = \text{Hann}^{1/2}(c)$, maintaining the fidelity weighting used in prior work. The low-frequency channel uses a substantially narrower cutoff, $R_{\text{lo}} = \text{Hann}^{1/2}(c_{\text{lo}})$ with $c_{\text{lo}} \ll c$, forming a complementary pair that partitions spectral correction responsibilities while keeping $R_{\text{hi}}^2 + R_{\text{lo}}^2$ close to identity over the inversion-relevant passband. The low-frequency constraint tolerance is chosen slightly looser than the high-frequency channel to allow mild model mismatch at DC while still driving large-scale bias reduction.

a) Spectral rationale: In primal–dual gradient schemes, the convergence rate of each Fourier mode is governed by the corresponding singular values of the forward operator. In limited-angle systems, high-frequency modes are well represented and produce comparatively large singular values, while low spatial frequencies induce much smaller gains. The original Hann weighting amplifies this gap by further suppressing near-DC singular values. When a single dual is used, the step size must be set to stabilize the high-frequency spectrum, which unintentionally results in an under-driven update for the low-frequency band. The optimizer therefore applies only a weak correction to DC error each iteration, but a strong correction to high frequencies, creating oscillatory or stalled behavior in the low band.

By assigning each channel an independent dual step size, the method reduces cross-band interference. The low-frequency dual ascent can safely be made larger, providing a stronger contraction factor for small k, without compromising the stability required to converge the high-frequency channel. The proximal update then combines both adjoint-filtered corrections while preserving non-negativity. This split-fidelity design yields a more spectrally uniform correction profile and mitigates frequency-induced convergence bottlenecks in limited-angle tomographic inversion.

VI. Experimental Methodology

To evaluate the proposed two-channel fidelity approach, we conduct inverse problem studies using a 2D LAR scanning configuration matching that of Sidky et al. Sidky et al. [2025]. The geometry uses 25 fan-beam projections over a 50° arc. The source is positioned 50 cm from the isocenter and 100 cm from the detector, which consists of 1024 bins spanning a length that inscribes a 10 cm × 10 cm field of view.

A. Digital Phantom

The phantom used for these studies is a 2D digital breast slice following the methodology of Sidky et al. [2025]. Glandular tissue regions are created by applying a threshold to power-law noise, with small bright specks added to simulate microcalcifications Reiser and Nishikawa [2010]. The test image is constructed as a weighted combination of three tissue components: adipose ($0.5\times$), fibroglandular ($1.0\times$), and calcification ($2.0\times$) layers.

To characterize how the two-channel fidelity benefit depends on problem conditioning, we test three array sizes: 512×512 pixels (0.2 mm pixel width), 256×256 pixels (0.4 mm pixel width), and 128×128 pixels (0.8 mm pixel width). Because the LAR inverse problem has a fixed number of measurements (25 views × 1024 detector bins = 25,600 ray integrals), coarser discretizations yield fewer unknowns and thus better-conditioned systems. This experimental design allows us to isolate the relationship between conditioning and the efficacy of the split-frequency fidelity strategy. Following Sidky et al. [2025], projection data are generated using the same pixel grid as the reconstruction, which removes discretization error and focuses the comparison on algorithmic differences. All studies use noiseless projection data to isolate the convergence behavior of the optimization algorithms from noise-related effects.

B. Algorithm Configuration

Both reconstruction methods—single-channel and two-channel—are configured with matched regularization parameters to ensure fair comparison. Following the resolution-dependent parameterization of Sidky et al. [2025], the DTV weight parameter is set to $\alpha = 1.7$ for 512×512, $\alpha = 1.9$ for 256×256, and $\alpha = 1.95$ for 128×128. The pixel sparsity parameter is set to $\beta = 5.0$ for the 512×512 case and $\beta = 10.0$ for the 256×256 and 128×128 cases.

The PDHG algorithm is run for 500 iterations using the He–Yuan predictor-corrector scheme He and Yuan [2012] with relaxation parameter $\rho = 1.75$. For the single-channel baseline, the data-discrepancy constraint employs a square-root Hanning filter with cutoff parameter $c = 4.0$. The two-channel method uses complementary filters: a low-pass channel with cutoff $c_{\text{lo}} = 8.0$ and a high-pass channel with cutoff $c_{\text{hi}} = 4.0$. The low-frequency channel employs an amplified dual step size with $\sigma_{\text{lo}} = 4.0 \times \sigma_{\text{hi}}$ and a relaxed tolerance $\varepsilon_{\text{lo}} = 1.25 \times \varepsilon_{\text{hi}}$, enabling stronger per-iteration correction to near-DC components while maintaining global PDHG stability Chambolle and Pock [2011]. Step sizes are determined via power iteration to satisfy the convergence condition $\tau\sigma\|WX\|^2 < 1$. Our implementation is publicly available at https://github.com/taroii/ct.

VII. Results

We now present quantitative and visual comparisons of the single-channel and two-channel methods. Table I summarizes reconstruction error across three image resolutions.

TABLE I
Image RMSE Comparison Across Resolutions

Resolution	Single-Channel	Two-Channel	Improvement
512×512	0.1625	0.1317	19.0%
256×256	0.0984	0.0767	22.1%
128×128	0.0332	0.0128	61.4%

The two-channel method achieves consistently lower reconstruction error across all tested resolutions. Notably, the improvement magnitude increases substantially with coarser discretization: from 19.0% at 512×512 to 61.4% at 128×128. This trend directly reflects the relationship between problem conditioning and the efficacy of split-frequency fidelity. At coarser resolutions, the ratio of measurements to unknowns improves (25,600 measurements versus 16,384 unknowns at 128×128, compared to 262,144 unknowns at 512×512), yielding a better-conditioned inverse problem. Under improved conditioning, the low-frequency error components—which the two-channel approach specifically targets—become the dominant source of residual error, and the amplified dual step size for the low-frequency channel can provide proportionally greater correction. Conversely, at finer resolutions where severe ill-conditioning introduces errors across all frequency bands, the targeted low-frequency correction constitutes a smaller fraction of the total improvement opportunity.

Figure 1 shows the convergence behavior for the 256×256 case. Both methods exhibit decrease in image RMSE, but the two-channel approach achieves lower error throughout the iteration process. The single-channel method displays pronounced oscillations in the early iterations that persist until approximately iteration 200, whereas the two-channel method stabilizes more rapidly after an initial transient. This difference is consistent with the theoretical motivation: by assigning an independent dual step size to the low-frequency channel, the two-channel approach reduces cross-band interference that would otherwise cause the optimizer to oscillate as it attempts to balance corrections across frequency bands with disparate conditioning.

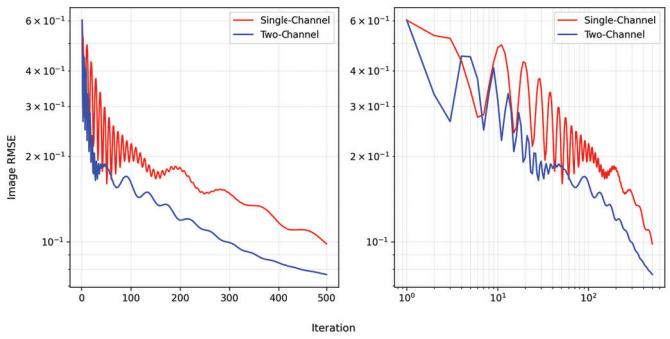


Fig. 1. Image RMSE convergence for 256×256 reconstruction. Left: linear iteration scale. Right: log-log scale. The two-channel method (blue) achieves lower error and reduced early-iteration oscillation compared to the single-channel baseline (red).

Figure 2 presents the reconstructed images and difference maps for the 256×256 case at 500 iterations. Both methods recover the gross phantom structure, including the glandular tissue distribution and microcalcification specks. The difference images reveal that the single-channel method retains more residual error distributed throughout the reconstruction, particularly in the form of large-scale intensity variations characteristic of unconverged low-frequency components. The two-channel reconstruction exhibits more uniform error reduction, with the residual errors appearing more localized to fine structural boundaries.

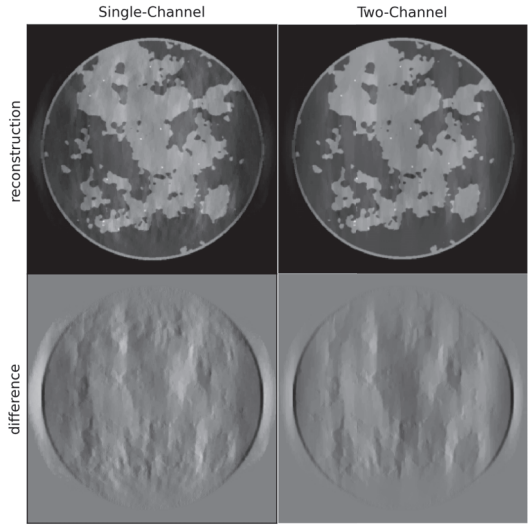


Fig. 2. Reconstruction comparison for 256×256 images. Top row: single-channel (left) and two-channel (right) reconstructions. Bottom row: corresponding difference images (reconstruction minus ground truth), displayed with grayscale range $[-0.75, 0.75]$. The two-channel result shows reduced large-scale intensity artifacts.

These results are consistent with the spectral analysis of Section 4: frequency-separated data fidelity enables more balanced convergence across the spatial frequency spectrum. The benefit is most pronounced when the underlying inverse problem is better conditioned, as occurs at coarser discretizations where the number of unknowns more closely approaches the number of measurements.

VIII. Conclusion

We analyzed the fundamental causes of slow and unstable convergence in limited-angle digital breast tomosynthesis reconstruction when using a single, globally weighted fidelity term in primal–dual optimization. The square-root Hanning weighting creates a large singular-value gap between high and near-DC modes. Because PDHG step sizes must stabilize the largest gains (high frequencies), the low-frequency channel is under-driven, leading to stalled or unstable convergence at DC.

To address this, we proposed a two-channel fidelity design that routes low and high spatial-frequency residuals through complementary square-root Hanning filters, each paired with an independent dual variable and step size. This strategy preserves noise robustness while reducing cross-band step-size interference, providing stronger corrective feedback to the ill-conditioned low-frequency channel without violating global PDHG stability. The resulting formulation better equalizes spectral corrections, mitigates low-band oscillations, and directly targets the most ill-conditioned modes in limited-angle inversion.

Experiments across multiple image resolutions reveal that the two-channel benefit scales with problem conditioning: improvements range from 19% at 512×512 (severely underdetermined) to 61% at 128×128 (moderately underdetermined). This conditioning dependence suggests the method is most impactful when low-frequency errors dominate the residual, as occurs when the inverse problem is sufficiently well-posed for high-frequency components to converge reliably.

These findings support the broader conclusion that accounting for spectral conditioning gaps in reconstruction optimizers for limited-angle tomography can help avoid unstable convergence tails. As future work, we believe that multi-dual or frequency-split fidelity designs can offer a principled and practical path to improving stability and convergence uniformity in other types of tomographic reconstruction.

References

A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

B. He and X. Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. *SIAM Journal on Imaging Sciences*, 5(1):119–149, 2012. doi:10.1137/100814494.

I. Reiser and R. M. Nishikawa. Task-based assessment of breast tomosynthesis: Effect of acquisition parameters and quantum noise. *Medical Physics*, 37(4):1591–1600, 2010. doi:10.1118/1.3357288.

E. Y. Sidky, X. Wu, X. Duan, H. Huang, W. Zhao, L. Y. Zhang, J. P. Phillips, Z. Zhang, B. Chen, D. Xia, et al. Accurate volume image reconstruction for digital breast tomosynthesis with directional-gradient and pixel sparsity regularization. *Journal of Medical Imaging*, 12(S1):S13013–S13013, 2025.

Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan. Directional-tv algorithm for image reconstruction from limited-angular-range data. *Medical Image Analysis*, 70:102030, 2021.

Unified Analysis of Uniqueness and Stability for Interior and Exterior Problems in Tomographic Image Reconstruction

Katsuya Fujii, Hiroyuki Kudo

***Abstract*—This paper investigates uniqueness and stability in computed tomography (CT) reconstruction problems in which projection data are missing both inside and outside a region. Reconstruction from externally truncated data is known as the interior problem, while reconstruction from internally missing data is referred to as the exterior problem. Although uniqueness and stability results for these problems have been established individually, they have traditionally been treated separately due to their distinct mathematical formulations. In this study, we present a unified theoretical framework for analyzing uniqueness and stability in interior, exterior, and mixed interior–exterior CT problems based on Cormack's inversion formula. Sufficient conditions for uniqueness and stability are derived under mild a priori assumptions on the object function. Numerical reconstruction examples are provided to illustrate the theoretical results.**

***Index Terms*—Computed tomography, interior problem, exterior problem, Radon transform, uniqueness, stability.**

I. Introduction

COMPUTED tomography (CT) image reconstruction is known to be stable when complete projection data are available. In practical imaging systems, however, data acquisition is often restricted due to physical limitations, geometric constraints, or requirements for radiation dose reduction [1]. As a result, reconstruction from incomplete projection data has become a central topic in modern CT research.

From the perspective of inverse problems, reconstruction from incomplete data inevitably raises fundamental issues concerning uniqueness and stability. Various restricted-data tomography problems have been extensively studied, including limited angle CT, sparse-view CT, interior CT, and exterior CT. Among these, interior CT and exterior CT are representative problems characterized by spatial truncation of projection data and exhibit complementary mathematical structures (see Fig. 1).

The theoretical foundation of CT reconstruction is provided by the Radon transform. Let f denote the attenuation (decay) function. The Radon transform $\mathcal{R}$ is defined as

$$\mathcal{R}f(t,\alpha) = p(t,\alpha) = \int_{\mathbf{R}} f\big(t\theta(\alpha) + s\theta^{\perp}(\alpha)\big)\, ds, \quad (1)$$

where $\theta(\alpha) = (\cos\alpha, \sin\alpha), \theta^{\perp}(\alpha) = (-\sin\alpha, \cos\alpha)$. This operator represents the line integral of f along a straight line determined by the angle α and the distance t from the origin. The set $P = \{p(t,\alpha) \mid t \in \mathbf{R},\ 0 \le \alpha < \pi\}$ is referred to as a sinogram or projection data. CT image reconstruction corresponds to the inverse problem of recovering f from P.

Katsuya Fujii is with the Faculty of Education, Kio University, 4-2-2 Umami-naka, Koryo-cho, Kitakatsuragi-gun, Nara 635-0832, Japan,
Hiroyuki Kudo is with the Faculty of Engineering, Information and Systems, University of Tsukuba, Tennoudai 1-1-1, Tsukuba 305-8573, Japan

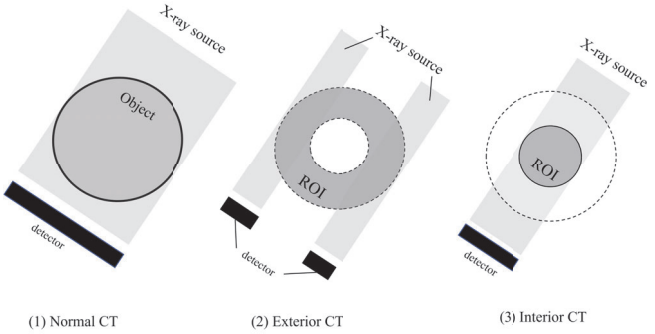


Fig. 1: The images (1) to (3) represent the normal, interior, and exterior CT with parallel beams.

In this paper, we investigate a reconstruction problem that simultaneously involves the difficulties of interior CT and exterior CT. Specifically, we consider reconstruction from projection data of the form

$$P_{\mathrm{mix}} = \{p(t,\alpha) \,|\, a \le |t| \le b,\ 0 < a < b < 1,\ 0 \le \alpha < \pi\}, \quad (2)$$

which we refer to as the mixed interior–exterior CT problem. This setting naturally arises in imaging systems that involve both a large field of view and strongly attenuating materials, such as metallic objects. The corresponding geometric configuration is illustrated in Fig. 2.

The main objective of this paper is to clarify the uniqueness and stability properties of the mixed interior–exterior CT problem, which are investigated in Section III. Since simultaneous internal and external data truncation prevents the direct use of conventional analytic inversion techniques, we develop a reconstruction strategy based on Cormack-type inversion formulas combined with a successive approximation method [2]. It is well known that both interior and exterior CT suffer from instability. In the interior case, reconstructions near the ROI boundary tend to produce low-frequency artifacts. This paper addresses these issues and proposes a method to achieve stable reconstruction. Numerical simulations of the proposed method are presented in Section IV.

II. Background on Interior and Exterior CT

This section summarizes classical interior CT and exterior CT problems from the perspective of inverse problems, which form the basis of the proposed mixed interior–exterior reconstruction framework.

A. Interior CT

In practical CT imaging, it is natural to assume that the attenuation function has compact support, and throughout this paper we assume $\operatorname{supp} f \subset B_1(0)$, where $B_1(0)$ denotes the unit disk centered at the origin. Interior CT aims to reconstruct an image from truncated projection data $P_{\mathrm{int}} = \{p(t,\alpha) \mid |t| \le b,\ 0 < b < 1,\ 0 \le \alpha < \pi\}$, corresponding to reconstruction within a region of interest (ROI), which is particularly relevant in medical CT for radiation dose reduction. It is well known that interior CT does not admit a unique solution without additional constraints. Starting with the pioneering work of [3], uniqueness can be restored if a nonzero a priori region exists inside the ROI [4], [5]. This line of research is well summarized in [6]. Subsequent studies have shown that this prior region may be approximate or piecewise constant [7], [8], or even learned from data [9]. From a stability viewpoint, reconstruction becomes more stable as the prior region approaches to the ROI boundary and increasingly unstable as the distance grows [10]. In this paper, we extend this result to the Cormack-type inversion formulas.

B. Exterior CT

Exterior CT considers the complementary situation in which projection data inside a certain detector range are missing $P_{\mathrm{ext}} = \{p(t,\alpha) \mid a \le |t|,\ 0 < a < 1,\ 0 \le \alpha < \pi\}$. This problem naturally arises when highly attenuating materials, such as metals, completely absorb X-rays. Cormack established the uniqueness of exterior CT reconstruction using an analytic framework based on circular harmonic expansions. Despite this theoretical uniqueness, exterior CT is known to be numerically unstable due to noise amplification inherent in the reconstruction formulas.

C. Mixed Interior–Exterior CT and Limitation of DBP

The mixed interior–exterior CT problem involves simultaneous data truncation both inside and outside a detector range, as defined in (2), which prevents the direct application of conventional analytic inversion methods. In interior CT, the differential back projection (DBP) method has been widely used for uniqueness analysis and reconstruction [11]. However, DBP requires continuous projection data to construct complete Hilbert lines. In the mixed interior–exterior setting, where projection data are limited to $a \le |t| \le b$, such complete Hilbert lines cannot be formed, and the associated Hilbert transform is no longer well defined.

Consequently, the DBP method is not applicable to the mixed interior–exterior CT problem. To address this issue, we instead employ Cormack-type inversion formulas, which remain theoretically valid under truncated projection data.

III. Analysis of the Mixed Interior–Exterior Problem

In this section, we analyze the uniqueness and stability properties of the mixed interior–exterior CT problem by exploiting Cormack-type inversion formulas. Cormack established explicit relations between the circular harmonic coefficients of the object function, denoted by f_ℓ, and the corresponding coefficients of the projection data, denoted by g_ℓ. These relations can be summarized as

$$g_\ell(s) = \sqrt{\frac{2}{\pi}} \int_s^1 f_\ell(r) \frac{T_{|\ell|}\left(\frac{s}{r}\right)}{\sqrt{1-\frac{s^2}{r^2}}}\, dr, \tag{3}$$

$$f_\ell(r) = -\frac{1}{\pi} \int_r^1 \frac{T_{|\ell|}\left(\frac{s}{r}\right)}{\sqrt{s^2-r^2}} g'_\ell(s)\, ds, \tag{4}$$

where $T_{|\ell|}$ are Chebyshev polynomials of the first kind. For the purpose of the following subsection, we consider the following modified inversion formula derived from (4) [1].

$$\begin{aligned} f_\ell(r) = &-\frac{1}{\pi} \int_r^1 \frac{\left(\frac{s}{r} + \sqrt{\frac{s^2}{r^2}-1}\right)^{-|\ell|}}{\sqrt{s^2-r^2}} g'_\ell(s) ds \\ &+ \frac{1}{\pi r} \int_0^r U_{|\ell|-1}\Big(\frac{s}{r}\Big) g'_\ell(s) ds, \end{aligned} \tag{5}$$

where $U_{|\ell|}$ denotes the Chebyshev polynomial of the second kind. The reason for adopting the modified Cormack-type inversion formula of the above form is that the formula (4) may lead to instability of the circular harmonic coefficients due to the influence of the Chebyshev polynomials of the first kind. Therefore, we employ the more stable reconstruction formula (5). On the other hand, although (5) requires interior projection data, we emphasize that our uniqueness and stability analyses do not require the availability of all projection data.

A. Uniqueness Results

Above Cormack's formulas provide an analytic connection between the object and the measured projection data and form the basis of the following discussion.

We consider the mixed interior–exterior data configuration (2) in which projection data are missing in both the interior and exterior regions. Let $r \in \mathbf{R}$ satisfy $r \in B \subset (a,b)$, where $B = (c,d)$ is a prior region (see Fig. 2) and assume that f_ℓ is supported on $(-1,1)$. Under the above setting, we present a simple three-step proof showing that f is uniquely determined on S. Our proof strategy follows the approach proposed by [4], [5].

[Step 1](Decomposition of f_ℓ) First, we decompose the reconstruction formula into contributions from the projection data acquisition region and the missing-data region as follows.

$$f_\ell(r) = h_1^\ell(r) + h_2^\ell(r), \tag{6}$$

where

$$\begin{aligned} h_1^\ell(r) = &-\frac{1}{\pi} \int_r^b \frac{\left(\frac{s}{r} + \sqrt{\frac{s^2}{r^2}-1}\right)^{-|\ell|}}{\sqrt{s^2-r^2}} g'_\ell(s) ds \\ &+ \frac{1}{\pi r} \int_a^r U_{|\ell|-1}\Big(\frac{s}{r}\Big) g'_\ell(s) ds, \end{aligned} \tag{7}$$

$$h_2^\ell(r) = -\frac{1}{\pi}\int_b^1 \frac{\left(\frac{s}{r}+\sqrt{\frac{s^2}{r^2}-1}\right)^{-|\ell|}}{\sqrt{s^2-r^2}} g'_\ell(s)ds + \frac{1}{\pi r}\int_0^a U_{|\ell|-1}\left(\frac{s}{r}\right) g'_\ell(s)ds. \quad (8)$$

When $r \in B$, the value of $f_\ell(r)$ on the left-hand side of (6) is assumed to be known. Moreover, since the projection-related function $g_\ell(s)$ is available on (a,b), the term $h_1^\ell(r)$ can also be evaluated on this interval.

[Step 2](Determination of f_ℓ on (a,b)) To clarify the role of h_2^ℓ, we extend it to a complex-valued function $h_2^\ell(z)$. By exploiting the analytic properties of the integrand of (8) and the continuity of $g'_\ell(s)$, it can be shown that $h_2^\ell(z)$ is analytic on the interval (a,b). As a result, $h_2^\ell(z)$ is uniquely determined on (a,b) by analytic continuation from the known values. Substituting the determined $h_2^\ell(r)$ back into (6), we conclude that $f_\ell(r)$ is uniquely determined on the interval (a,b).

[Step 3](Uniqueness of target region) Since each circular harmonic coefficient is uniquely determined, the circular harmonic expansion is uniquely determined for functions belonging to an appropriate function class. In summary, we obtain the following uniqueness result.

[Uniqueness results for interior and exterior problems] Let S be a target annular region (see Fig. 2). If projection data passing through S are available and the object function f is known on a ring-shaped prior region B located inside S, then f is uniquely determined on S.

We remark that the above uniqueness result naturally includes both the interior and exterior CT problems as special cases. Specifically, when the inner diameter a is set to zero, the problem reduces to the interior case, while as the outer diameter approaches one, it corresponds to the exterior case, in both situations ensuring uniqueness of the reconstruction within the target region.

B. Stability Results

In this subsection, we investigate the stability of the mixed interior and exterior problem. Our analysis is conducted along the same lines as the approach developed in [10]. As in the uniqueness result, we divide the domain into the projection-data acquisition region and the missing-data region (6). Specifically, we evaluate the following inequality, where the subscript "err" denotes quantities that include errors.

$$|f_{\ell,\mathrm{err}}(r) - f_\ell(r)| \le |h^\ell_{1,\mathrm{err}}(r) - h^\ell_1(r)| + |h^\ell_{2,\mathrm{err}}(r) - h^\ell_2(r)|. \quad (9)$$

Furthermore, for the stability analysis, we impose the following technical assumptions.

1) The error in the projection-data acquisition region satisfies $|h^\ell_{1,\mathrm{err}} - h^\ell_1| \le \varepsilon E_\ell$, where $\sum_{\ell\in\mathbf{Z}} E_\ell = 1$ and $\varepsilon > 0$.
2) The derivative of the projection data g_ℓ is bounded for $r \in (0,a)$ and $r \in (b,1)$.

Under the above assumptions, we obtain the following result.

[Stability results for interior and exterior problems]

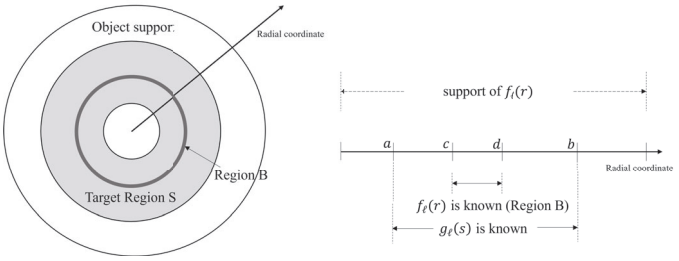


Fig. 2: Geometry of the mixed interior and exterior CT (left) and setting of parameters for our problems (right).

Let the parameters be as shown in Fig. 2. Then, the inequality (9) holds for $r \in (a,c)$ as

$$|f_{\ell,\mathrm{err}}(r) - f_\ell(r)| < \varepsilon E_\ell(r) + \frac{\tau_2(r)}{(1+\ell^2)^2}\left(\frac{\varepsilon}{|\tau_2(c)|}\right)^{\delta_1(r)}, \quad (10)$$

and for $r \in (d,b)$ as

$$|f_{\ell,\mathrm{err}}(r) - f_\ell(r)| < \varepsilon E_\ell(r) + \frac{\tau_1(r)}{(1+\ell^2)^2}\left(\frac{\varepsilon}{|\tau_1(d)|}\right)^{\delta_2(r)}. \quad (11)$$

Here, $\delta_1(r)$ satisfies $\lim_{r\to a}\delta_1(r) = 1$ and $\lim_{r\to c}\delta_1(r) = 0$, whereas $\delta_2(r)$ satisfies $\lim_{r\to b}\delta_2(r) = 1$ and $\lim_{r\to d}\delta_2(r) = 0$. Moreover, τ_1 and τ_2 are defined as

$$\tau_1(r) = \frac{\sqrt{1-b}}{\pi}\left(\frac{1}{2r}\ln\frac{(1-r)(b+r)}{(1+r)(b-r)}\right)^{\frac{1}{2}} + \frac{\sqrt{a}}{\pi}\left(\frac{1}{2c}\ln\frac{c+a}{c-a}\right)^{\frac{1}{2}},$$

$$\tau_2(r) = \frac{\sqrt{a}}{\pi}\left(\frac{1}{2r}\ln\frac{r+a}{r-a}\right)^{\frac{1}{2}} + \frac{\sqrt{1-b}}{\pi}\left(\frac{1}{2d}\ln\frac{(1-d)(b+d)}{(1+d)(b-d)}\right)^{\frac{1}{2}}.$$

The above inequality indicates that the reconstruction becomes unstable as the point approaches the boundaries a and b of the annular region. For example, since $\lim_{r\to a}\tau_2(r) = +\infty$, the estimate (10) implies that the reconstruction becomes unstable. Furthermore, in a manner similar to the results reported by [10], we also obtain the following stability estimate when the target region I is enclosed by a prior region.

$$|f_{\ell,\mathrm{err}}(r) - f_\ell(r)| < T(1+\ell^2)^{-2}\varepsilon^\delta, \quad (12)$$

where T is a constant and $\delta = \min_{r\in I}\delta(r) \in [0,1]$. This inequality implies uniform boundedness over the projection data acquisition region, from which stability in the target region follows. Here we briefly comment on the stability in the exterior case. Even in the exterior setting, the achieved stability is attributed to the use of a ring-shaped a priori region, and is fundamentally different from the approach such as [7], which relies on the strong unrealistic assumption on the function f over the whole reconstruction region. To verify these observations, numerical experiments are conducted in the next section.

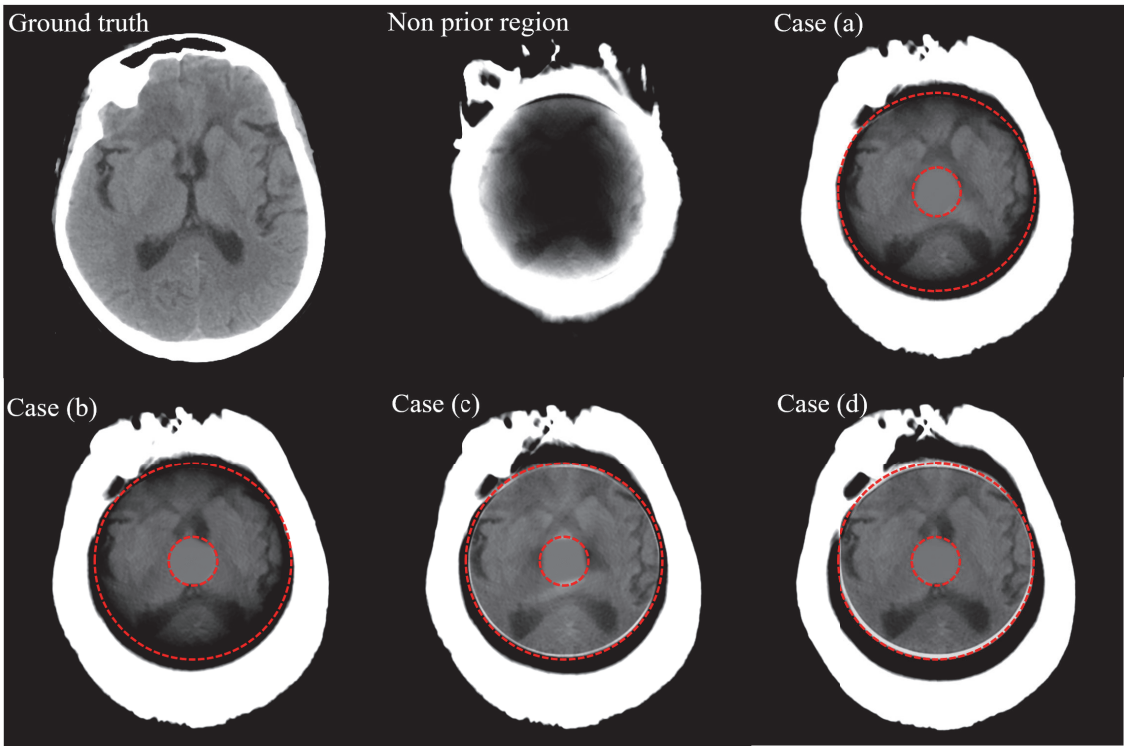


Fig. 3: Reconstruction results for various prior region configurations in Fig. 4, illustrating the effect of prior region location on reconstruction stability: (a) inner, (b) inside, (c) outside, and (d) both inside and outside prior regions.

IV. Numerical Study

In the interior case with a single a priori known region, it is well known that the reconstruction becomes unstable near the boundary of the ROI. Specifically, the intensity values tend to become darker (or brighter), leading to the emergence of low-frequency artifacts. The novelty of this paper lies in addressing these issues and propose a method to achieve stable reconstruction for the interior–exterior problem under such challenging conditions.

To apply the proposed interior and exterior tomography, we consider the following discretized total variation (TV) minimization problem:

$$\min_{\mathbf{x}} \ \|\mathbf{x}\|_{\mathrm{TV\ on\ S}} \quad \text{subject to } A\mathbf{x} = \mathbf{b}, \tag{13}$$

where $\mathbf{x}$ denotes the image vector, $\mathbf{b}$ represents the corresponding projection data, and S denotes the target region. In the formulation of Eq. (13), we introduced a weak TV regularization to suppress the amplification of discretization errors during the iterations. Numerical experiments were conducted using parallel-beam geometry with a low-contrast brain CT image of size 512×512 pixels as the target object. The target region was defined as an annular region with an outer diameter of 256 pixels and an inner diameter of 50 pixels, and only projection data passing through the target region were used. Figure 4 shows reconstruction results for four different prior configurations, indicating that, as discussed in Section III-B, instability increases as the reconstruction region moves away from the prior region, while the target region is well reconstructed when it is enclosed by the prior region. Finally,

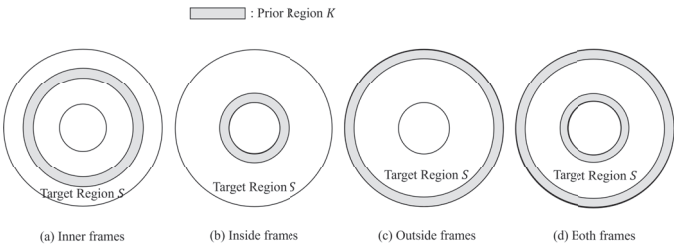


Fig. 4: Various configurations of a prior regions.

the horizontal profiles (at $y = 256$) of each reconstructed image are shown in Fig. 5. These results also indicate that the location of the prior region affects the stability of the accuracy of reconstructed regions.

V. Conclusion

This paper presents a unified framework for analyzing the uniqueness and stability properties of interior, exterior, and mixed interior–exterior CT problems using Cormack-type inversion formulas. We derive uniqueness results and stability indicators under mild a priori assumptions, and numerical experiments are presented to illustrate the theoretical findings. Finally, the extension to cone-beam geometry is expected to be challenging, as a circular harmonic expansion is not available in this case. Nevertheless, this issue should be investigated in future work.

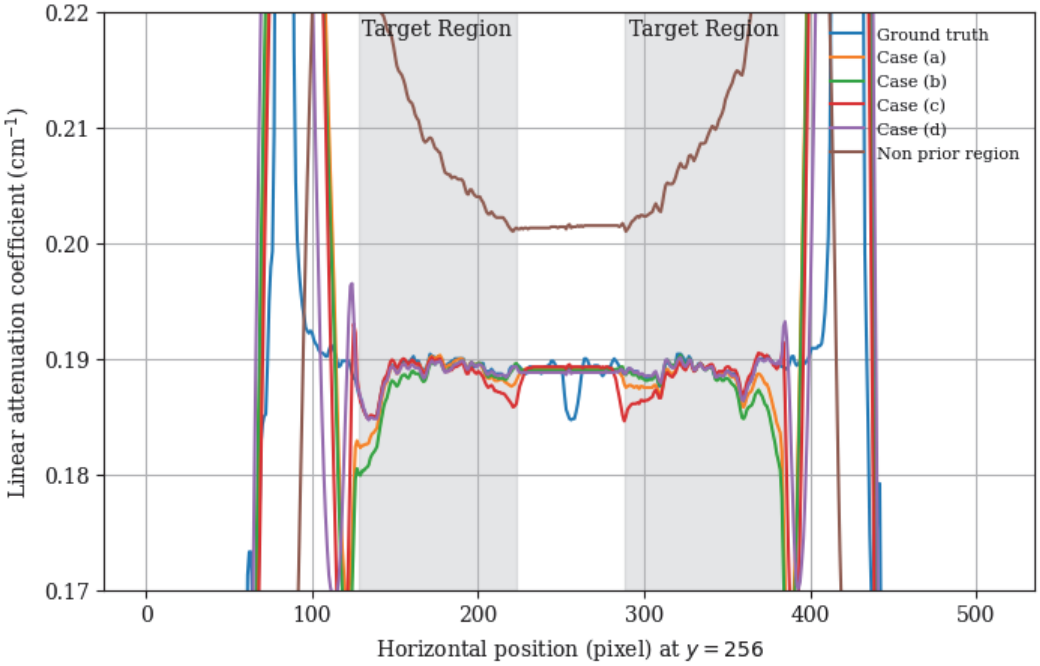


Fig. 5: Profiles along the central horizontal line $y = 256$ corresponding to Fig. 3. Each curve corresponds to Ground truth, Non-prior region, and Cases (a)–(d) shown in Fig. 3.

Acknowledgment

This research was supported by JST-CREST (JPMJCR2335), JSPS Grant-in-Aid for Scientific Research (B) (No. 23K28346), and JSPS Grant-in-Aid for Scientific Research (B) (No. 24K02995).

References

[1] F. Natterer, *The Mathematics of Computerized Tomography*, Wiley, New York, 1986.

[2] A. M. Cormack, "Representations of a function by its line integrals with some radiological applications," *J. Appl. Phys.*, vol. 34, pp. 2722–2727, 1963.

[3] M. Defrise, F. Noo, R. Clackdoyle, and H. Kudo, Truncated Hilbert transform and image reconstruction from limited tomographic data, *Inverse Problems*, vol. 22, no. 3, pp. 1037–1053, 2006.

[4] Y. Ye, H. Yu, Y. Wei, and G. Wang, "A general local reconstruction approach based on a truncated Hilbert transform," *Int. J. Biomed. Imaging*, 2007, Art. ID 63634, 8 pp., 2007, doi:10.1155/2007/63634.

[5] H. Kudo, M. Courdurier, F. Noo, and M. Defrise, "Tiny a priori knowledge solves the interior problem in computed tomography," *Phys. Med. Biol.*, vol. 53, no. 9, pp. 2207–2231, 2008, doi:10.1088/0031-9155/53/9/001.

[6] R. Clackdoyle and M. Defrise, Tomographic reconstruction in the 21st century, *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 60–80, 2010.

[7] J. Yang, H. Yu, M. Jiang, and G. Wang, "High-order total variation minimization for interior tomography," *Inverse Problems*, vol. 26, no. 3, Art. ID 035013, 2010, doi:10.1088/0266-5611/26/3/035013.

[8] H. Kudo, "Advanced compressed sensing image reconstruction for interior tomography," in *Proc. SPIE Optical Engineering + Applications*, San Diego, CA, USA, 2019.

[9] Y. Han and C. Ye, "One network to solve all ROIs: Deep learning CT for any ROI using differentiated backprojection," *Med. Phys.*, vol. 46, no. 12, pp. e855–e872, 2019.

[10] M. Courdurier, F. Noo, M. Defrise, and H. Kudo, Solving the interior problem of computed tomography using a priori knowledge, *Inverse Problems*, vol. 24, no. 6, p. 065001, 2008.

[11] F. Noo, R. Clackdoyle, and J. D. Pack, "A two-step Hilbert transform method for 2D image reconstruction," *Phys. Med. Biol.*, vol. 49, no. 17, pp. 3903–3923, 2004.

Anisotropic Total Variation Regularization for Artifact Reduction in Contrast-Enhanced Digital Breast Tomosynthesis using PDHG Optimization

Yixin Li, Xiangyi Wu, Emil Y. Sidky, Wei Zhao

***Abstract*—Digital Breast Tomosynthesis (DBT) suffers from limited-angular acquisition geometry, leading to highly ill-conditioned reconstruction and depth blurring. This work evaluates a Primal–Dual Hybrid Gradient (PDHG) optimization framework with anisotropic Total Variation (ATV) regularization for Contrast-Enhanced DBT (CEDBT). The method minimizes a least-squares data fidelity term with an ATV prior under non-negativity constraints, with algorithm step sizes determined from the system operator's largest singular value via power iteration. Using both physical phantom and clinical datasets, and comparing against Filtered Backprojection (FBP), the ATV-PDHG approach substantially reduces limited-angular artifacts and high-frequency noise, while accurately recovering iodine concentration ratios. Clinical results further show that tuning regularization parameters can enhance tumor conspicuity and margin sharpness, though with increased background texture. Overall, the proposed ATV-PDHG framework demonstrates strong potential for improving CEDBT reconstruction quality and clinical utility.**

***Index Terms*—anisotropic TV, CEDBT, clinical application, PDHG, reconstruction**

I. Introduction

Digital Breast Tomosynthesis (DBT) has become an important advancement in breast imaging, offering pseudo-3D visualization that reduces tissue superposition and improves lesion detectability compared with conventional 2D mammography [1][2]. By acquiring a limited number of projections over a restricted angular range, DBT enhances the differentiation of benign and malignant findings and decreases false-positive recalls and unnecessary biopsies, contributing to improved diagnostic performance and patient care [3][4]. However, the limited-angular geometry renders the reconstruction problem highly ill-conditioned, leading to size-dependent depth blurring and high-frequency artifacts in analytic reconstruction methods such as Filtered Backprojection (FBP) [5][6].

To address these challenges, substantial research has focused on advanced reconstruction strategies ranging from statistical iterative algorithms [7][8] to modern deep learning approaches [9][10]. Among these, Total Variation (TV) regularization has proven effective for mitigating limited-angular artifacts by exploiting gradient sparsity to recover high-fidelity images from incomplete measurements [11][12][13]. Prior studies have demonstrated that integrating TV constraints into the Primal–Dual Hybrid Gradient (PDHG) algorithm can substantially improve image quality, generating reduced noise and enhanced structural consistency compared with FBP [14].

In this study, we implement the PDHG optimization framework to Contrast-Enhanced DBT (CEDBT), which incorporates both low-energy （LE） and high-energy (HE) acquisitions to generate dual-energy（DE）images containing complementary morphological and functional information [15]. To better account for the asymmetric spatial resolution of DBT systems, we employ anisotropic TV [16] (ATV) regularization within the reconstruction model. Previously, Sidky et al. [17] established PDHG for DBT and applied the method on CEDBT phantom images. We extend this framework by implementing weighted subtraction for iodine contrast in the projection domain, accelerating the reconstruction using a differentiable projector [18] and developing the GPU-accelerated ATV-PDHG pipeline, thereby enabling clinical translation (achieving less than 5 minutes for a full 3D clinical breast image reconstruction). The proposed ATV-PDHG method is evaluated using both a physical phantom and clinical patient data to assess its quantitative accuracy, artifact suppression capability, and potential for improving lesion conspicuity in clinical CEDBT imaging.

II. Methods

A. Data Acquisition

Data acquisition was performed using an experimental CEDBT system (MAMMOMAT Inspiration, Siemens Healthineers) [19]. The imaging protocol consisted of two

This work was supported in part by the U.S. National Institutes of Health (NIH) under Grant R01EB035551. (Corresponding author: Wei Zhao)

Yixin Li is with the Department of Radiology, Stony Brook University, Stony Brook, NY 11794, USA (e-mail: yixin.li@stonybrookmedicine.edu).

Xiangyi Wu is with the Department of Radiology, Stony Brook University, Stony Brook, NY 11794, USA (e-mail: xiangyi.wu@stonybrook.edu).

Emil Y. Sidky is with the Department of Radiology, The University of Chicago, Chicago, IL 60637, USA (e-mail: sidky@uchicago.edu).

Wei Zhao is with the Department of Radiology, Stony Brook University, Stony Brook, NY 11794, USA (e-mail: wei.zhao@stonybrookmedicine.edu).

separate DBT acquisitions, a LE scan and a HE scan. For each acquisition, the X-ray tube moved continuously through an angular range of approximately 50°, acquiring 25 projection views. The LE spectrum was selected to remain below the k-edge of iodine (33.2 keV), while the HE acquisition was performed at a tube voltage of 49 kVp, utilizing a spectrum optimized to maximize iodine signal contrast for this system[20][21][22].

To validate the reconstruction methods, a 5 cm thick BR3D physical phantom was utilized. The phantom consists of five 1 cm thick semicircular slabs. The central slab contains solid iodinated contrast targets with diameters of 2, 3, 5, and 8 mm (with cylindrical height equal to diameter) and iodine concentrations of 1, 2, 3, and 5 mg/mL. Additionally, the clinical study, which was approved by our Institutional Review Board (IRB) and was fully HIPAA compliant, was utilized to demonstrate the results.

B. Image Processing and Dual Energy Subtraction

The data processing pipeline is designed to isolate the iodine signal in the projection domain prior to reconstruction. We performed spatial registration between the HE and LE raw projection data to correct any potential motion during the data acquisition. The logarithmic transformation was applied to the raw intensity data ($\boldsymbol{I}$) to obtain line integral attenuation values according to the Beer-Lambert law

$$\boldsymbol{p} = -\ln\frac{\boldsymbol{I}}{\boldsymbol{I_0}}, \tag{1}$$

where $\boldsymbol{I}_0$ is the air scan intensity under the same energy of X-ray source. The dual energy projection is calculated as

$$\boldsymbol{p}_{DE} = \boldsymbol{p}_{HE} - w \cdot \boldsymbol{p}_{LE}, \tag{2}$$

where w is the weighting factor that empirically optimized to suppress background and enhance iodine contrast.

C. Reconstruction Algorithm Design

The x-ray reconstruction is formulated as a discrete linear inverse problem, where the projection measurements are modeled as the line integrals of object's attenuation coefficient.

$$\boldsymbol{y} = A\boldsymbol{x}, \tag{3}$$

where vector $\boldsymbol{y} \in \mathbb{R}^I$ denote the model data. Vector $\boldsymbol{x} \in \mathbb{R}^J$ denotes the discrete image. $A \in \mathbb{R}^{I\times J}$ depicts the system matrix with element A_{ij} indicating the intersection length of ray i passing through voxel j.

To accelerate convergence and mitigate the ill-conditioned inverse problem arising from the limited-angular DBT geometry, we incorporated spectral weighting into the reconstruction framework [17]. The frequency domain filter operation is applied, consisting of a Ramp filter multiplied by a Hann window. The cutoff parameter for the Hann window to suppress high frequency noise was experimentally optimized as a fraction of the Nyquist frequency.

Algorithm 1. Pseudocode for N steps of the Fourier-Filtered least squares errors with the ATV-PDHG algorithm.

$L \leftarrow \|\mathcal{F}(A), \nabla\|_2; \tau \leftarrow \frac{1}{L}; \sigma \leftarrow \frac{1}{L}; \theta \leftarrow 1; n \leftarrow 0$
$\boldsymbol{u_0} \leftarrow \boldsymbol{0}_J, \boldsymbol{p_0} \leftarrow \boldsymbol{0}_I, \boldsymbol{q_0} \leftarrow \boldsymbol{0}_I$
$\bar{\boldsymbol{u}}_0 \leftarrow \boldsymbol{u_0}$
$\bar{\boldsymbol{g}} = \mathcal{F}(\boldsymbol{g})$
repeat
  $\boldsymbol{p}_{n+1} \leftarrow (\boldsymbol{p}_n + \sigma(\mathcal{F}((A\bar{\boldsymbol{u}}_n) - \bar{\boldsymbol{g}}))/(1+\sigma)$
  $\boldsymbol{v} \leftarrow \boldsymbol{q}_n + \sigma\nabla\bar{\boldsymbol{u}}_n$
  $\boldsymbol{q}_{n+1} \leftarrow \text{clip}(\boldsymbol{v}, -\boldsymbol{\lambda}, \boldsymbol{\lambda})$
  $\boldsymbol{u}_{n+1} \leftarrow pos(\boldsymbol{u}_n - \tau A^T(\mathcal{F}^T(\boldsymbol{p}_{n+1})) + \tau div(\boldsymbol{q}_{n+1}))$
  $\bar{\boldsymbol{u}}_{n+1} \leftarrow \boldsymbol{u}_{n+1} + \boldsymbol{\theta}(\boldsymbol{u}_{n+1} - \boldsymbol{u}_n)$
  $n \leftarrow n+1$
until $n \geq N$
Return $\boldsymbol{u_n}$

We formulated image reconstruction as the ATV regularized optimization problem with a non-negativity constraint.

$$\boldsymbol{u}^* = \underset{\boldsymbol{u}}{\text{argmin}}\{\tfrac{1}{2}\|\mathcal{F}(A\boldsymbol{u}) - \mathcal{F}(\boldsymbol{g})\|_2^2 + \lambda_x\|\partial_x\boldsymbol{u}\|_1 + \lambda_y\|\partial_y\boldsymbol{u}\|_1 + \lambda_z\|\partial_z\boldsymbol{u}\|_1 + \delta_P(\boldsymbol{u})\}, \tag{4}$$

where $\boldsymbol{g}$ is the measured projection data. $\mathcal{F}(\cdot)$ indicate the sinogram domain filter function. The regularization terms consist of the ATV semi-norms, where the linear operators $\partial_x\boldsymbol{u}$, $\partial_y\boldsymbol{u}$, and $\partial_z\boldsymbol{u}$ represent discrete approximations to the gradient image in x, y and z directions, respectively. The parameters λ_x, λ_y, and λ_z serve as regularization weights that control the sparsity enforcement along each dimension to preserve edges and mitigate the limited angle artifacts. The term $\delta_P(\boldsymbol{u})$ represents the indicator function for the non-negative constraints.

To solve this non-smooth convex optimization problem, we utilized a modified version of PDHG algorithm. For the sake of maintaining the presentation flow, we only show the pseudo-code in Algorithm 1.

To guarantee convergence, the primal and dual step sizes (τ and σ) are linked to the Lipschitz constant L of the combined linear operator system. We defined the composite linear operator that maps the image to both the filtered data domain and the gradient domain $K = [\mathcal{F}(A), \nabla]$, where the estimated largest singular value $L = \|K\|_2$ was estimated using the standard power iteration method [14]. Consequently, we initialized the step sizes as $\tau = \sigma = 1/L$, and the extrapolation parameter θ was set to 1. The operator $\nabla = (\partial_x, \partial_y, \partial_z)^T$ represents the discrete approximation of the image gradient, while div denotes the discrete divergence operator and it satisfies the adjoint relationship $div = -\nabla^T$.

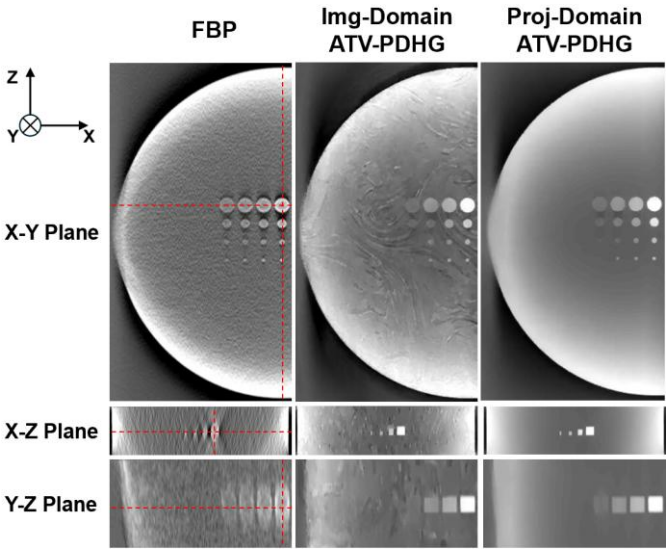


Fig 1. Comparison of reconstruction results for the BR3D phantom containing iodine targets. The columns show the FBP method, the image-domain ATV-PDHG algorithm, and the proposed projection-domain ATV-PDHG algorithm. The rows present the in-plane view (X-Y plane) and the vertical cross-sectional views (X-Z and Y-Z planes). The red dashed crosshairs indicate the centered location of the selected iodine target across all three planes.

Referring to Algorithm 1, Line 6 performs the dual variable update associated with the Fourier-Filtered data fidelity term. Line 7 executes the dual update for the ATV regularization, which involves a clip operation that limits the magnitude of the dual variables at each pixel to the corresponding directional threshold λ_i ($i \in \{x, y, z\}$), effectively implementing the proximal operator for the l_1 norm. Finally, Line 9 computes the primal update using the adjoint operators ($A^T, \mathcal{F}^T$ and div) followed by a projection onto the non-negative convex set.

All reconstructions were performed on a workstation equipped with a single NVIDIA GeForce RTX 5090 GPU. The reconstruction pipeline incorporated a differentiable forward projector built upon DrJit compiler [23], leveraging a PyTorch backend to accelerate the optimization process.

III. Results

Figure 1 displays the reconstructed CEDBT phantom images isolating the iodine contrast agent distribution. The proposed projection-domain dual-energy subtraction ATV-PDHG method demonstrates superior performance compared to both standard FBP and the image-domain subtraction approach. As observed in the in-plane (X-Y) views, while the image-domain ATV-PDHG reconstruction introduces structured background artifacts, the projection-domain method effectively suppresses background noise, yielding a highly uniform background. Furthermore, in the cross-sectional (X-Z and Y-Z) views, the projection-domain ATV-PDHG substantially reduces the limited-angle depth blurring artifacts prominent in the FBP images, resulting in improved margin sharpness and superior visualization of the iodine targets.

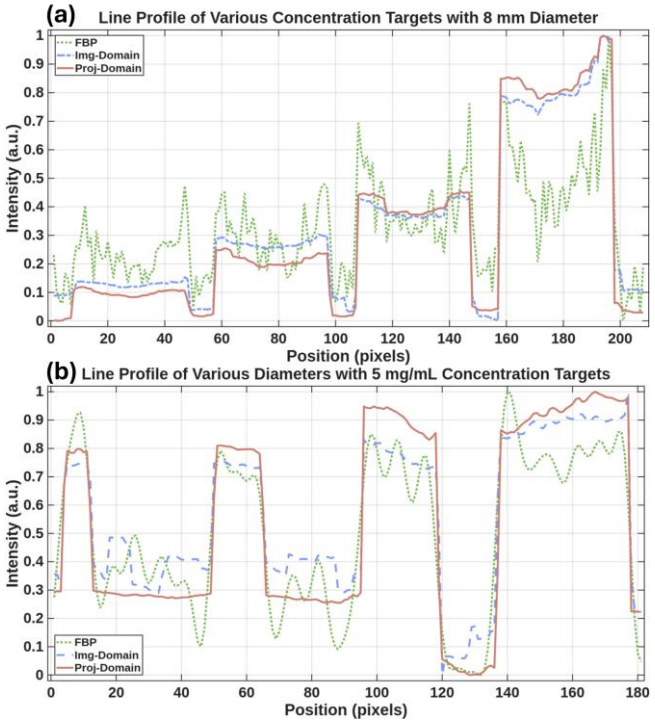


Fig. 2 Quantitative evaluation of reconstructed iodine targets using line intensity profiles comparing FBP, image-domain ATV-PDHG, and projection-domain ATV-PDHG. (a) Line profiles of the 8 mm diameter targets across varying concentrations. The profiles correspond to a vertical line traversing the center of the inserts in the X-Y plane at a depth of 28 mm. (b) Line intensity profiles traversing the 5 mg/mL concentration iodine targets across varying diameters. The plotted values represent the normalized signal intensities corresponding to a horizontal line drawn through the center of the inserts in the cross-sectional depth view (X-Z plane).

Figure 2 provides a quantitative evaluation of the reconstruction accuracy using line intensity profiles. Figure 2a presents the profiles for the 8 mm diameter targets across varying iodine concentrations at the 28 mm depth slice, with the y-axis representing the normalized signal intensity. Notably, both ATV-PDHG reconstructions demonstrate superior noise suppression, producing a stable intensity profile that clearly distinguishes the concentration levels, whereas the FBP profile exhibits significant signal fluctuations. Figure 2b compares the intensity profiles of the 5 mg/mL iodine targets across varying diameters in the depth (X-Z) plane. The FBP reconstruction exhibits severe intensity oscillations and fails to recover the uniform density of the targets due to the limited-angular sampling. In contrast, both ATV-PDHG profiles successfully recover the characteristic piecewise constant 'box' shape of the inserts and demonstrate steeper edge transitions and a consistent flat plateau, indicating superior depth resolution and structural fidelity compared to FBP. Furthermore, the projection-domain method yields noticeably cleaner background suppression than the image-domain approach.

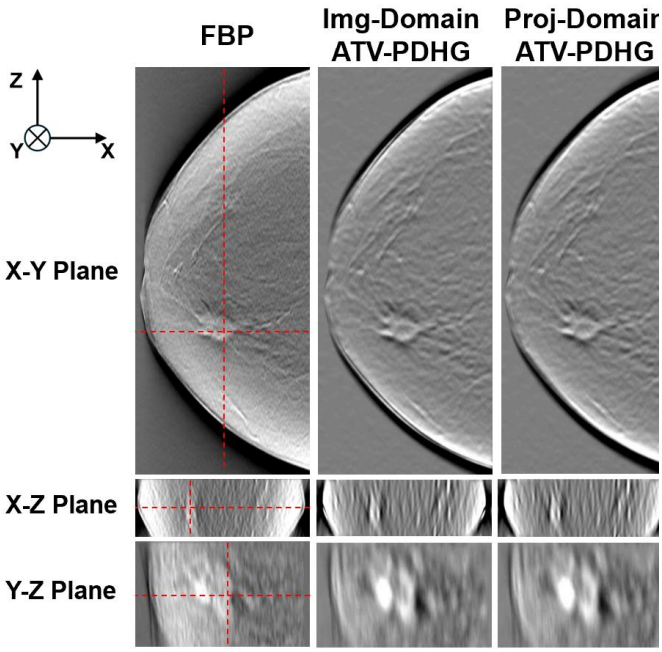


Fig 3. Reconstruction results for a clinical CEDBT case featuring a biopsy-proven invasive ductal carcinoma. The figure displays orthogonal views of the reconstructed CEDBT images centered on the mass lesion. The columns correspond to FBP against the image-domain and projection-domain ATV-PDHG frameworks. The rows show the in-plane view (X-Y plane) and the vertical cross-sectional views (X-Z and Y-Z planes). The red dashed crosshairs indicate the localization of the lesion across all three planes.

Figure 3 presents the reconstruction results for a clinical case with biopsy-proven invasive ductal carcinoma (IDC), comparing standard FBP with both the image-domain and projection-domain ATV-PDHG frameworks. Both ATV-PDHG implementations demonstrate enhanced lesion visibility relative to standard FBP as observed in the in-plane (X-Y) views. Unlike the physical phantom study, both PDHG domains exhibit a similar visual profile in this specific clinical instance, characterized by the amplification of residual background parenchymal structures (due to patient motion) under the current unoptimized parameter configuration. Nevertheless, in the cross-sectional (X-Z and Y-Z) views, both ATV-PDHG methods exhibit improved lesion contrast and depth resolution relative to FBP.

IV. Discussion

This work demonstrates the feasibility and performance of the ATV-PDHG for reconstruction of clinical CEDBT images. Phantom imaging results showed that ATV-PDHG consistently outperforms standard FBP by mitigating limited-angle artifacts and suppressing noise while preserving quantitative iodine signal.

More importantly, we demonstrated that performing dual-energy subtraction in the projection-domain prior to optimization yields superior background suppression compared to post-reconstruction image-domain subtraction. Compared with the phantom experiments, clinical data introduced additional challenges, such as increased anatomical complexity and patient motion artifacts. More aggressive filtering (cutoff 0.12 Nyquist frequency) is required to match noise characteristics. Furthermore, the unoptimized parameter configuration relied on heavy TV regularization to suppress this noise. This aggressive penalty dominated the data fidelity term, over-enhancing background parenchymal structures and effectively masking the visual differences between the image- and projection-domain subtraction methods in the clinical case. This highlights that parameter tuning critically influences the balance between lesion visibility and background texture, emphasizing that future work must explore adaptive and task-based optimization strategies for robust clinical translation.

V. Conclusion

This study presents a GPU-accelerated, projection-domain subtraction ATV-PDHG reconstruction framework for CEDBT. Both phantom and clinical results show superior artifact suppression, improved depth resolution, and enhanced lesion visibility relative to FBP. The approach demonstrates strong promise for clinical application, with ongoing development needed to optimize regularization for diagnostic performance.

VI. Acknowledgment

We gratefully acknowledge financial support from the NIH (R01EB035551).

References

[1] S. Vedantham, A. Karellas, G.R. Vijayaraghavan, and D.B. Kopans. "Digital breast tomosynthesis: state of the art." *Radiology* 277, no. 3 (2015): 663-684.

[2] R. J. Hooley, M.A. Durand, and L.E. Philpotts. "Advances in digital breast tomosynthesis." *American Journal of Roentgenology* 208, no. 2 (2017): 256-266.

[3] M. M. Goodsitt, and A.D. Maidment. "Evolution of tomosynthesis." *Journal of Medical Imaging* 12, no. S1 (2025): S13012-S13012.

[4] H. Huang et al. "Comparison of contrast-enhanced digital mammography and contrast-enhanced digital breast tomosynthesis for lesion assessment." *Journal of medical imaging* 6, no. 3 (2019): 031407-031407.

[5] A. M. Mota, N. Matela, N. Oliveira, and P. Almeida. "Total variation minimization filter for DBT imaging." *Medical physics* 42, no. 6Part1 (2015): 2827-2836.

[6] E. T. Quinto. "Artifacts and visible singularities in limited data X-ray tomography." Sensing and Imaging 18, no. 1 (2017): 9.

[7] Z. Hu et al. "An improved statistical iterative algorithm for sparse-view and limited-angle CT image reconstruction." Scientific reports 7, no. 1 (2017): 10747.

[8] C. Wang, M. Tao, J.G. Nagy, and Y. Lou. "Limited-angle CT reconstruction via the l_1/l_2 minimization." *SIAM Journal on Imaging Sciences* 14, no. 2 (2021): 749-777.

[9] T. Würfl et al. "Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems." IEEE transactions on medical imaging 37, no. 6 (2018): 1454-1463.

[10] Y. Huang et al. "Some investigations on robustness of deep learning in limited angle tomography." In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 145-153. Cham: Springer International Publishing, 2018.

[11] Z. Chen, X. Jin, L. Li, and G. Wang. "A limited-angle CT reconstruction method based on anisotropic TV minimization." *Physics in Medicine & Biology* 58, no. 7 (2013): 2119.

[12] W Zhang et al. "Limited angle CT reconstruction by simultaneous spatial and Radon domain regularization based on TV and data-driven tight frame." *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 880 (2018): 107-117.

[13] M Persson, D Bone, and H Elmqvist. "Total variation norm for three-dimensional iterative reconstruction inlimited view angle tomography." *Physics in Medicine & Biology* 46, no. 3 (2001): 853.

[14] E. Y. Sidky, J. H. Jørgensen, and X. Pan. "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm." *Physics in Medicine & Biology* 57, no. 10 (2012): 3065.

[15] X. Duan, P. Sahu, H. Huang, and W. Zhao. "Deep-learning convolutional neural network-based scatter correction for contrast enhanced digital breast tomosynthesis in both cranio-caudal and mediolateral-oblique views." Journal of Medical Imaging 10, no. S2 (2023): S22404-S22404.

[16] Y. Lou, T. Zeng, S. Osher, and J. Xin. "A weighted difference of anisotropic and isotropic total variation model for image processing." SIAM Journal on Imaging Sciences 8, no. 3 (2015): 1798-1823.

[17] E. Y. Sidky et al. "Accurate volume image reconstruction for digital breast tomosynthesis with directional-gradient and pixel sparsity regularization." *Journal of Medical Imaging* 12, no. S1 (2025): S13013-S13013.

[18] Y. Li et al. "Physically-Based Inverse Rendering Framework for PET Image Reconstruction." *arXiv preprint arXiv:2508.19956* (2025).

[19] M. D. Hörnig, L. Bätz, and T. Mertelmeier. "Design of a contrast-enhanced dual-energy tomosynthesis system for breast cancer imaging." In *Medical Imaging 2012: Physics of Medical Imaging*, vol. 8313, pp. 1325-1333. SPIE, 2012.

[20] X. Wu, X. Duan, H. Huang, and W. Zhao. "Scatter correction for contrast-enhanced digital breast tomosynthesis with a dual-layer detector." *Journal of Medical Imaging* 12, no. S1 (2025): S13008-S13008.

[21] X. Duan, H. Huang, and W. Zhao. "A novel direct-indirect dual-layer flat-panel detector for contrast-enhanced breast imaging: Monte Carlo simulation." *Medical Physics* 52, no. 9 (2025): e18122.

[22] X. Duan et al. "A novel direct-indirect dual-layer flat-panel detector for contrast-enhanced breast imaging: Experimental assessment." *Medical Physics* 52, no. 9 (2025): e18108.

[23] W. Jakob, S. Speierer, N. Roussel, and D. Vicini. "Dr. jit: A just-in-time compiler for differentiable rendering." *ACM Transactions on Graphics (TOG)* 41, no. 4 (2022): 1-19.

Basis material image reconstruction for digital breast tomosynthesis using a single kV scan

JP Phillips, Emil Y. Sidky, Ingrid S. Reiser, Buxin Chen, Zheng Zhang, Xiaochuan Pan

***Abstract*—In this work, we investigate the possibility of basis material decomposition in digital breast tomosynthesis (DBT) from a conventional single-kV scan. We employ DBT system physics modeling and the use of a pathlength constraint to enable estimation of the basis material sinograms, which can be subsequently reconstructed. Using scan data of a mammography phantom, we decompose transmission data collected over a limited angular range into pathlength basis sinograms of wax and PMMA, two materials known to comprise the bulk of the phantom; volume images are then reconstructed by a blend of DBT-adapted filtered back-projection and iterative image reconstruction. Results show that material decomposition sinograms can be estimated with quantitative accuracy using only conventional mammogram data. Quantitative accuracy in the reconstructed volumes is lacking, but the availability of reconstructed basis volumes offers the possibility of combining the basis images in a way that optimizes signal-to-noise ratio of features; i.e. microcalcifications, masses, and spiculations at the margin of masses, which may indicate malignancy.**

***Index Terms*—digital breast tomosynthesis, tomographic reconstruction, mammography, material decomposition**

I. Introduction

THE current standard imaging modality for breast cancer screening is mammography due to its advantages in detection accuracy and resource-effectiveness. [1] However, a key weakness of mammography is visibility of malignant lesions is decreased by overlapping dense fibroglandular tissue, leading to false negative diagnoses. [2] More recently, digital breast tomosynthesis (DBT) techniques have been implemented and clinically adopted by reconstructing tomographic images from low-dose mammogram projection images over a limited angular range, allowing for reconstruction of breast image slices at different depths. [3] While the potential of DBT for breast screening has been demonstrated, the augmentation of standard mammogram imaging techniques indicates a continued interest in improving cancer detection rates.

Mammogram and tomographic breast cancer screening methods may be further improved through the use of material decomposition methods by providing quantitative information about tissue composition. [5], [7] Material decomposition is traditionally achieved by acquiring energy-resolved photon information, such as using dual-energy or photon counting imaging, to leverage differences in material attenuation based on energy. However, most practical methods for acquiring energy-resolved mammography data rely on taking multiple sequential scans which are susceptible to motion artifacts and result in additional patient dose. The development of a method that would only require a standard spectral acquisition would improve the feasibility of material decomposition for mammography imaging.

Previously, material decomposition has been accomplished using a single conventional CT dataset augmented by a pathlength constraint, assuming that the total thickness of a scanned object is known. [6] The purpose of this work is to apply the pathlength constraint to mammography scan data of the Mammo 156 Phantom and decompose transmission data into wax and PMMA basis material sinograms. Subsequently, the basis sinograms are used to reconstruct DBT volume images of the basis materials.

J. P. Phillips, E.Y Sidky, I. S. Reiser, B. Chen, Z. Zheng, and X. Pan are with the Department of Radiology MC-2026, The University of Chicago, 5841. S. Maryland Ave., Chicago, IL, 60637. (e-mail:jpphillips@uchicago.edu)

II. Methods

A. Material Decomposition

For material decomposition with two basis materials, we consider the standard CT model for transmitted x-ray fluence ϕ through an object which can be expressed in the form:

$$\phi(t_1, t_2) = \phi_0 \int_0^{E_{\max}} S(E) e^{-\mu_1(E)t_1 - \mu_2(E)t_2} dE, \quad (1)$$

where $E_{\max}$ is the maximum energy integrated over, ϕ_0 is the initial photon fluence, $S(E)$ is the x-ray source energy spectrum at tube voltage V, and $\mu_1(E)$ and $\mu_2(E)$ are energy dependent attenuation coefficients corresponding to basis materials with pathlength t_1 and t_2. In this work, without any approximation, we perform the standard linearization of Eq. 1 through:

$$g(t_1, t_2) = -\log\big(\phi(t_1, t_2)\big). \quad (2)$$

In DECT, material decomposition into two materials requires a second measurement using a source spectrum operating at a different tube voltage. However, if total pathlength T of the object is known, then the second acquisition can be circumvented by applying a pathlength sum constraint to the data model [6]:

$$T = t_1 + t_2, \quad (3)$$

$$\phi(t_1) = \phi_0 \int_0^{E_{\max}} S(E) e^{-\mu_1(E)t_1 - \mu_2(E)(T - t_1)} dE, . \quad (4)$$

Using known attenuation coefficients and modeling $S(E)$ with the publicly available Python package SpekPy, Eq. 3 and Eq.

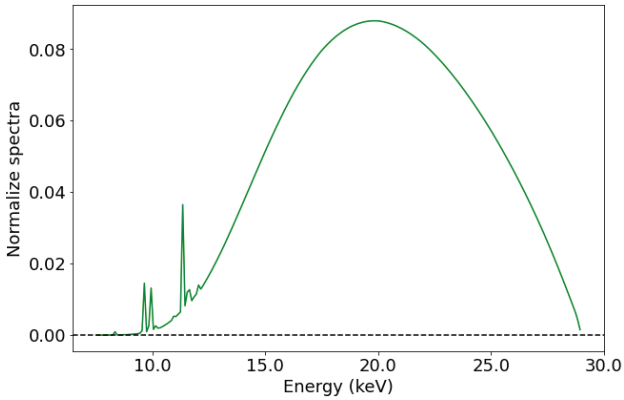


Fig. 1. 29 kV spectrum generated using SpekPy. Tungsten was used as the target material with 0.7 mm of aluminum filtering and an additional 3.25 mm of PMMA filtration to match the spectrum half-value layer reported by the imaging system (HVL $\approx$ 0.51 mm Al).

4 can be used to recover t_1 and t_2. [8] We solve for t_1 from $g(t_1) = -\log(\phi(t_1))$ by root-finding and recover the basis material pathlengths; the material path lengths for each ray are grouped together to form the basis material sinograms. Image reconstruction is then performed on each of the material sinograms using a 50%/50% blended combination of an iterative algorithm, directional total-variation (DTV) minimization [10], and filtered back-projection (FBP) with a specially designed filter for DBT [9]. The DTV-minimization algorithm is based on solving the following optimization problem

$$\arg\min_f \{\alpha_x\|\partial_x f\|_1 + \alpha_y\|\partial_y f\|_1 + \alpha_z\|\partial_z f\|_1 \quad \text{such that } \|R(Xf-b)\|_2 \le \epsilon\}, \quad (5)$$

where f is the reconstructed volume; b is the basis material sinogram, i.e. the collection of thicknesses t_1 or t_2 along all rays in the DBT scan; X is the DBT projection matrix; R is a ramp filter so that the filtered projection data is compared with the filtered estimated-projection; the data-fidelity is enforced through an inequality constraint where ϵ is a parameter that enforces an upper bound on the data discrepancy; ∂_x, ∂_y, ∂_z are the directional derivatives along the coordinate axes; and the DTV penalty parameters α_x, α_y, α_z weight the DTV terms in each direction differently.

B. Phantom Data

The Mammo 156 Phantom has a total physical thickness of 45mm, and is composed of PMMA and a 7.25 mm thick wax insert. This Phantom was imaged onsite at the University of Chicago using the Hologic Selenia Dimensions mammography system. 15 views were taken over an angular range of -7° to 7°. The x-ray tube potential was set at 29kV with a tungsten target and 0.7 mm of aluminum filtration. No scatter grid was used during data collection.

Preprocessing of the raw projection data is needed prior to applying the decomposition algorithm. First, ϕ_0 is estimated to normalize the measured values of ϕ using a negative log normalization. ϕ_0 is dependent on both scan angle and position in the detector array. To approximate each projection, a 2D

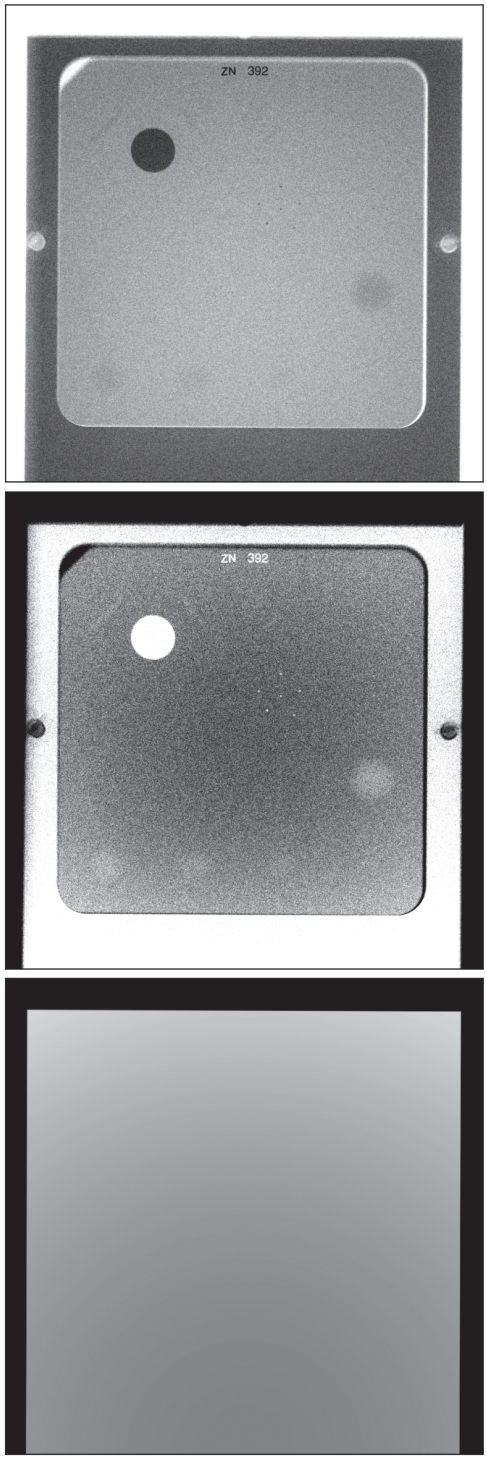


Fig. 2. Projection images of Mammo 156 Phantom for scan angle 0°. The raw transmission data (top) is shown with a window of [221, 292]. The transmission data are processed with a negative-log normalization and applying a scatter offset. The processed sinogram (middle) has a window of [2.36, 2.66]. For both processing and material decomposition, knowledge of the object thicknesses is needed. The object pathlength map (bottom) is obtained using the projection matrix from the scan and is shown with a window of [44, 46] mm.

polynomial function dependent on detector element position was fitted to ϕ values from regions where there was no object between the x-ray source and the detector. Offsets

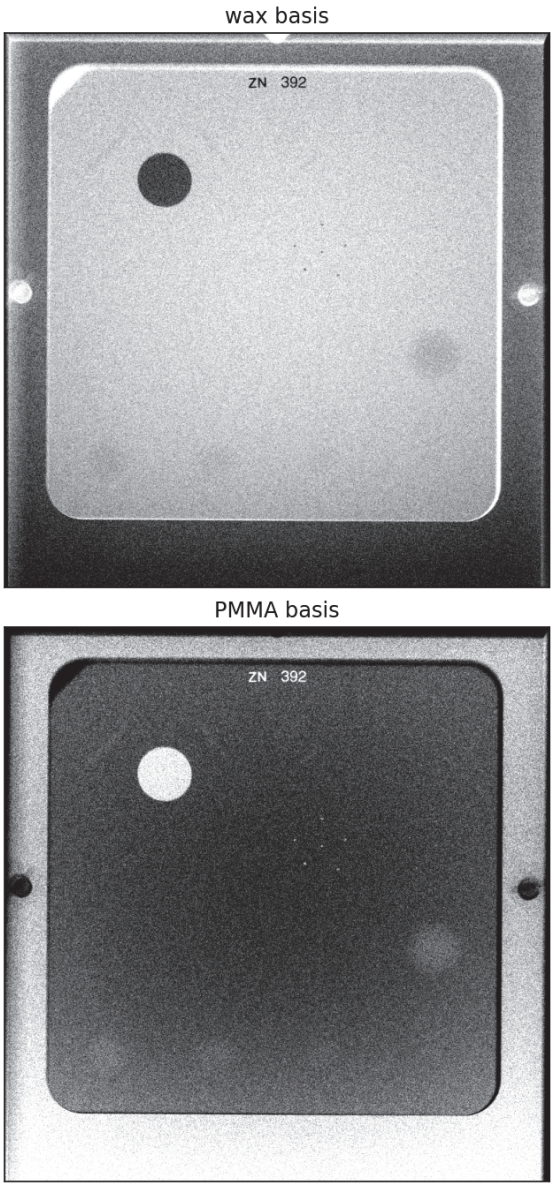


Fig. 3. Wax (top) and PMMA (bottom) basis pathlength projection images for scan angle 0°. Both images are presented in units of mm. The wax basis window is [-2, 7.5] mm and the PMMA window is [37.5, 45] mm.

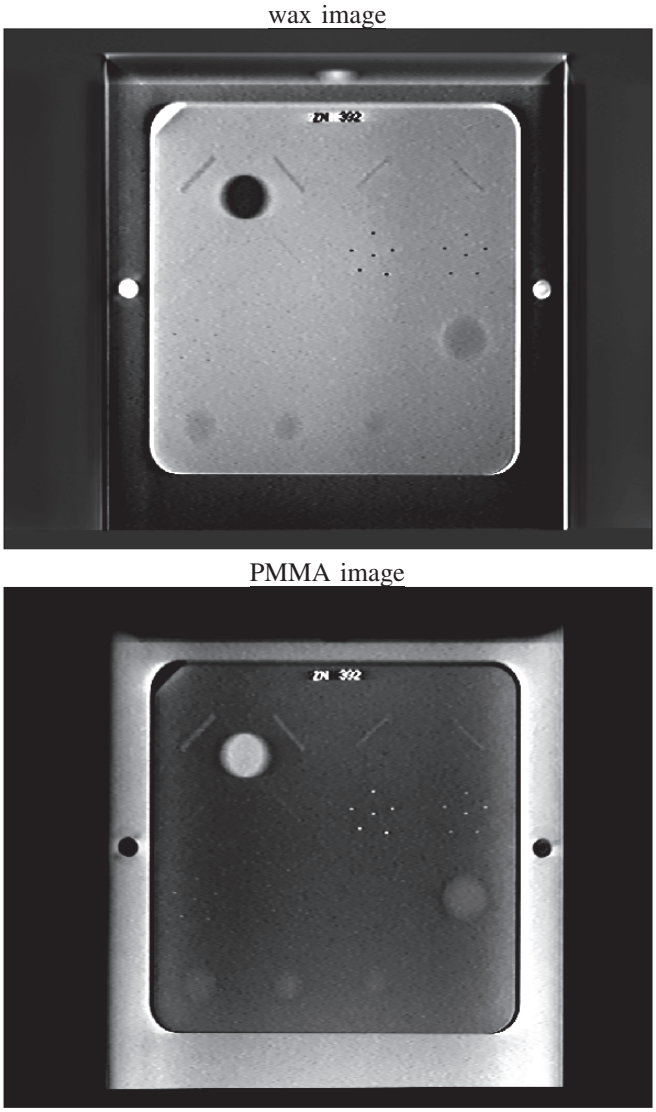


Fig. 4. Reconstructed basis material images for a plane containing the phantom signals. The gray scale for the wax image is [-0.02,0.1] and that of the PMMA image is [0.425,0.575], where the grayscale values represent fractional occupancy of the respective materials.

for signal measured due to scatter is also necessary. Scatter was approximated by comparing measured ϕ in regions only containing PMMA with expected ϕ values calculated using Eq. 1 for PMMA pathlengths between 0 and 45 mm. Exponential decay functions were fitted to the measured and estimated ϕ values. Anywhere that the object pathlength was nonzero, the difference between the fitted functions at that pathlength was subtracted from the measured ϕ to account for scatter. This processing is enabled by the fact that the dimension of the phantom are available to us; for actual clinical DBT scanning of the breast such information is available by making use of the height of the compression paddle.

The processed projection images were used as inputs to the decomposition algorithm to produce pathlength basis sinograms for two basis materials. Basis material selection is vital to ensuring quantitative accuracy of the basis sinograms. PMMA was used as one basis material as most of the phantom is composed of the material. The exact composition of the wax insert in the phantom was unknown but was found to be well approximated by using the attenuation coefficients of adipose tissue with a density of 0.2 g/cm^3. The attenuation data for both materials were acquired from the NIST database [4].

III. Results

Fig. 1 shows the model source spectrum using known scan parameters. In addition to 0.7 mm of aluminum filtration, an additional 3.25 mm of PMMA was added to account for the compression paddle used during data collection to match the HVL reported in the image DICOM header.

Raw and processed projections for the central scan angle is shown in Fig. 2 in addition to the object pathlength map. Processed projections from each view angle are decomposed into wax and PMMA basis material images. Basis images for

the central scan angle are shown in Fig. 3. Quantitative accuracy of the basis images was assessed by comparing recovered pathlengths to knowledge of the phantom's dimensions. PMMA pathlength in regions with no insert are approximately 45 mm, matching the height of the phantom. Wax pathlengths in the insert region are approximately 7 mm which is expected. The least accurate regions occur along the hard edges of the phantom where PMMA tends to be overestimated and wax pathlength is negative. The wax window is set to include negative values so the underestimation can be visualized.

Slice images from reconstructed basis volumes are shown in Fig. 4 for the plane in the Mammo 156 phantom containing the test signals. The wax image is shown for a gray scale that includes negative values so that any undershoot in this basis image can be appreciated. Both images have some minimal degree of non-uniformity due to imperfect physics modeling and image reconstruction artifacts. The wax gray values are clearly greater than zero but are significantly below 1; the sinogram decomposition of this material corresponds well to the true thickness of the wax insert but the depth blur of this thin planar object results in the low gray values. The disk, spec, and fiber signals in this image appear darker than the background because the materials comprising these signals are different than the wax/PMMA basis materials and the wax component of these signal objects comes in with a negative coefficient.

The PMMA image in the bottom panel of Fig. 4 is shown in a narrow grayscale window so that the signals are visible. The regions bordering the wax insert are pure PMMA, and the mean value of these regions is approximately 0.55, again, well short of 1. As with the wax image, the quantitative gap is caused by depth-blur, but in this case the blurring is not as severe because the PMMA in the phantoms is much thicker than the wax insert, thus occupying a greater fraction of the height of the reconstruction volume. The phantom signals are visible in the PMMA image with larger gray values than the background wax region.

IV. Conclusion

In this work we have presented a numerical method to reconstruct pathlength basis sinograms from conventional mammography data collected with a single spectrum. The pathlength sinograms are used to reconstruct basis volumes using DBT reconstruction. The ability to complete material decomposition with a standard spectral acquisition has been enabled by the implementation of a pathlength constraint in place of a second acquisition.

We apply this method to scan data of the Mammo 156 Phantom with the goal of recovering material pathlength and basis volumes that are quantitatively accurate. Processing of the raw data was key in maximizing quantitative accuracy, including estimating initial photon fluence for normalization, scatter approximation, and basis material selection to model the phantom element chemical composition. Accuracy of recovered pathlengths can still be improved, with regions near hard edges of the object suffering the most. Improvements to the processing methodology will be key in further maximizing quantitative accuracy.

Future work will include deeper studies into sources of error affecting quantitative accuracy of the material decomposition, such as implementing more robust scatter corrections and further tuning spectrum parameters and basis material qualities. Improvement to this method will increase the quantitative accuracy of material decomposition. This work focuses on the decomposition into basis sinograms and the image reconstruction into basis volumes. Future work will investigate the processing of these basis volumes to highlight features of clinical significance with the goal of enhancing screening and diagnostic performance of DBT.

Acknowledgment

This research was supported in part by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award numbers R01CA287302 and R21CA263660.

References

[1] C. J. D'Orsi. Early detection of breast cancer: mammography. *Breast Cancer Research and Treatment*, 18 Suppl 1:S107–109, May 1991.

[2] K. A. Duncan, G. Needham, F. J. Gilbert, and H. E. Deans. Incident round cancers: What lessons can we learn? *Clinical Radiology*, 53(1):29–32, January 1998.

[3] Fiona J. Gilbert, Lorraine Tucker, and Ken C. Young. Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. *Clinical Radiology*, 71(2):141–150, February 2016.

[4] J Hubbell and S Seltzer. Tables of X-ray mass attenuation coefficients and mass energy- absorption coefficients (version 1.4), 2004.

[5] M. Kahani, A. Kamali-asl, H. Ghadiri, and S. Hashemi. A method for material decomposition in dual-energy contrast enhancement digital mammography. *Measurement*, 88:87–95, June 2016.

[6] Benedek Janos Kis, Zsolt Sarnyai, Róbert Kákonyi, Miklós Erdélyi, and Gábor Szabó. Single-Energy Material Decomposition Using X-Ray Path Length Estimation. *Journal of Computer Assisted Tomography*, 36(6), 2012.

[7] Lambert T. Leong, Serghei Malkov, Karen Drukker, Bethany L. Niell, Peter Sadowski, Thomas Wolfgruber, Heather I. Greenwood, Bonnie N. Joe, Karla Kerlikowske, Maryellen L. Giger, and John A. Shepherd. Dual-energy three-compartment breast imaging for compositional biomarkers to improve detection of malignant lesions. *Communications Medicine*, 1(1):29, August 2021. Publisher: Nature Publishing Group.

[8] Gavin Poludniowski, Artur Omar, Robert Bujila, and Pedro Andreo. Technical Note: SpekPy v2.0—a software toolkit for modeling x-ray tube spectra. *Medical Physics*, 48(7):3630–3637, 2021. _eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.14945.

[9] S. D. Rose, A. A. Sanchez, E. Y. Sidky, and X. Pan. Investigating simulation-based metrics for characterizing linear iterative reconstruction in digital breast tomosynthesis. *Med. Phys.*, 44(9):e279–e296, 2017.

[10] E. Y. Sidky, X. Wu, X. Duan, H. Huang, W. Zhao, L. Y. Zhang, J. P. Phillips, Z. Zhang, B. Chen, D. Xia, I. S. Reiser, and X. Pan. Accurate volume image reconstruction for digital breast tomosynthesis with directional-gradient and pixel sparsity regularization. *J. Med. Imaging*, 12(S1):art. no. S13013, 2025.

Robust Limited-Angle CT Reconstruction via Geometry-Conditioned Multi-Environment Learning

Shujie Jin, Zilin Jiang, Ran Zhang, and Guang-Hong Chen

***Abstract*—Limited-angle CT reconstruction is inherently ill-posed due to data incompleteness and is highly sensitive to acquisition geometry, including the starting angle and angular coverage. Although deep learning–based methods have shown promise in mitigating limited-angle artifacts, most existing approaches are trained under fixed acquisition configurations and generalize poorly to unseen geometries.**

In this work, we address this limitation from a causal invariance learning perspective by explicitly modeling acquisition geometry as an environment variable rather than treating it as an implicit nuisance factor. Limited-angle reconstruction is formulated as a geometry-conditioned image-domain mapping, in which the starting angle and angular coverage are provided as conditioning inputs to the reconstruction network. A multi-environment training strategy is employed by systematically varying these parameters while holding the underlying anatomy fixed, thereby promoting geometry-invariant reconstruction behavior.

Experiments on clinical cardiac CT data demonstrate that a single trained model achieves stable reconstruction performance across a wide range of starting angles and angular coverages without retraining, consistently outperforming filtered backprojection and exhibiting improved robustness to acquisition geometry variations.

***Index Terms*—Multi-Environment Learning, Limited-View Reconstruction, Causal Invariance, Deep Learning**

I. Introduction

Accurate tomographic image reconstruction fundamentally relies on projection data that are sufficiently sampled in both view angle range and angular density. When the available angular coverage is limited, as in limited-view-angle CT or digital tomosynthesis, the acquired data are intrinsically incomplete, leading to severe image artifacts and an ill-posed reconstruction problem. Limited-view-angle acquisition arises frequently in practice due to hardware constraints, geometric restrictions, and the demand for rapid data acquisition in applications such as dynamic imaging and digital tomosynthesis.

In recent years, deep learning–based reconstruction methods have demonstrated promising performance in suppressing limited-view-angle artifacts by learning image-domain priors from data [1], [2]. However, most existing approaches are trained under a fixed acquisition configuration, typically assuming a specific view angle range and a fixed starting angle. In practical imaging scenarios, these acquisition parameters often vary across scans. As a result, models trained under a single configuration frequently exhibit substantial performance degradation when applied to unseen view angle ranges or to different starting angles within the same angular coverage.

This limitation reflects a fundamental mismatch between the acquisition process and the learning objective. While the acquisition geometry strongly influences the appearance and orientation of limited-view-angle artifacts, the underlying object content remains unchanged. From a physical perspective, the starting angle and view angle range are acquisition-dependent factors that affect how information is lost, but they are not causal determinants of the object itself. Consequently, an ideal reconstruction method should produce image quality that is insensitive to changes in acquisition geometry, provided that the underlying object remains the same.

In contrast, when deep learning models are trained to translate artifact-contaminated reconstructions—obtained using filtered backprojection or other reconstruction algorithms—into artifact-reduced images under fixed acquisition settings, the learned mappings often become implicitly tied to geometry-specific artifact patterns. Such models generally lack the ability to generalize across different view angle ranges or even across different starting angles within the same angular coverage. From a causal perspective, this behavior suggests that the learned models exploit environment-specific correlations rather than learning reconstruction features that are intrinsic to the object content, which is the true causal source of the measured data.

These observations motivate the following scientific question: *Can a single reconstruction model be trained to produce stable and high-quality limited-view-angle CT images across varying starting angles and view angle ranges by explicitly encouraging invariance with respect to acquisition geometry?*

In this work, we address this question by adopting a causal invariance learning principle, in which acquisition geometry is treated as an explicit environment variable rather than an implicit nuisance factor. Within this causal-invariance-inspired training framework, a single unified reconstruction model is trained across multiple acquisition environments characterized by different starting angles and view angle ranges. As a result, the learned reconstruction mapping is encouraged to remain stable with respect to geometry variations while adapting to geometry-dependent artifact patterns through conditioning. Experimental results demonstrate that the proposed method maintains consistent reconstruction quality under varying starting angles and view angle ranges, thereby substantially improving robustness and generalization relative to conventional reconstruction approaches.

S. Jin and Z. Jiang are with the Department of Medical Physics, University of Wisconsin-Madison, WI, USA.

R. Zhang and G. Chen are with the Department of Medical Physics and the Department of Radiology, University of Wisconsin-Madison, WI, USA

II. Method

This section presents the proposed environment-aware reconstruction framework for limited-view-angle CT. We first describe the overall image-domain strategy and geometry-conditioned formulation of the reconstruction problem. We then detail the reconstruction network architecture with explicit environment conditioning, followed by the multi-environment training strategy motivated by the causal invariance learning principle.

A. Overall Strategy and Problem Formulation

The proposed framework adopts an image-domain approach to address the limited-view-angle CT reconstruction problem. Given a limited-view sinogram, a classical filtered backprojection (FBP) algorithm is first applied to generate an image-domain reconstruction contaminated by limited-view artifacts. These artifact-corrupted images serve as inputs to a supervised learning framework, while the corresponding artifact-free reconstructions obtained from full-view data are used as training targets.

To address the scientific question formulated in the Introduction—robust reconstruction under varying starting angles and view angle ranges—we explicitly incorporate acquisition geometry into the reconstruction model. Changes in acquisition geometry, such as the starting angle and angular coverage, are viewed as interventions that alter the manifestation of limited-view artifacts without causing any change in the underlying object content, which is the desired reconstruction target.

Accordingly, limited-view-angle CT reconstruction is formulated as a geometry-conditioned mapping

$$\hat{x} = f(y; e), \tag{1}$$

where y denotes the artifact-corrupted image reconstructed from limited-view data, $\hat{x}$ is the artifact-reduced reconstruction, and $e = (\theta_s, \Delta\theta)$ represents the acquisition environment defined by the starting angle θ_s and view angle range $\Delta\theta$. By providing the acquisition geometry as an explicit conditioning input, geometry is treated as an environment variable rather than an implicit nuisance factor.

B. Reconstruction Network

To solve the problem formulated in Eq. (1), we adopt an image-domain reconstruction network with explicit environment conditioning. The overall architecture follows an encoder–decoder U-Net [3] design with skip connections, augmented with attention mechanisms [4], as illustrated in Fig. 1.

The network takes the artifact-corrupted image y as input and outputs the corresponding artifact-reduced reconstruction $\hat{x}$. A shallow convolutional layer first projects the input image into a feature space. Multi-scale feature extraction and reconstruction are then performed through symmetric downsampling and upsampling paths.

At each resolution level, feature processing is carried out by a *BasicLayer*, which consists of multiple groups of *Convolutional Residual Blocks* and *Attention Residual Blocks*.

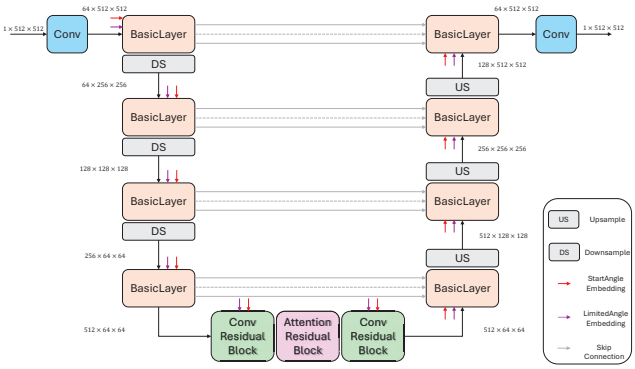


Fig. 1: Architecture of the proposed environment-aware image-domain reconstruction U-Net. The network adopts an encoder–decoder structure with skip connections. Environment information $e = (\theta_s, \Delta\theta)$ is encoded as learnable embeddings and injected additively into multiple *BasicLayer*s across different scales.

The acquisition environment $e = (\theta_s, \Delta\theta)$ is encoded into two separate learnable embeddings corresponding to the starting angle and angular coverage, respectively. These embeddings are injected additively into the intermediate feature representations within each *BasicLayer* at multiple scales in both the encoder and decoder. This multi-scale conditioning strategy allows geometry information to modulate feature representations throughout the network rather than being confined to a single layer.

As a result, the network can adapt its reconstruction behavior according to the acquisition geometry while preserving a shared reconstruction mechanism across different environments. A final convolutional layer maps the decoded features back to the image domain to produce the reconstructed image.

C. Causal Invariance Learning via Multi-Environment Training

As discussed in the Introduction, the goal of this work is to train a reconstruction model whose performance remains invariant to variations in the starting view angle and the limited view angle range. This objective naturally motivates a causal invariance learning principle for limited-view-angle CT reconstruction. From a causal perspective, the underlying object is the true source of the desired reconstruction, whereas the acquisition geometry affects only how image information is lost during data acquisition and reconstruction. Although different geometries induce distinct artifact patterns, the mapping from artifact-corrupted images to artifact-reduced images should remain invariant across acquisition environments.

To enforce this invariance, we adopt a multi-environment learning framework inspired by causal inference methods [5], [6], [7]. The key idea is that interventions on non-causal factors should not change the desired outcome. In limited-view-angle CT, the starting angle and angular coverage are treated as such non-causal factors. Training across multiple environments therefore poses counterfactual questions: if the acquisition geometry changes while the object remains fixed, should the reconstruction change?

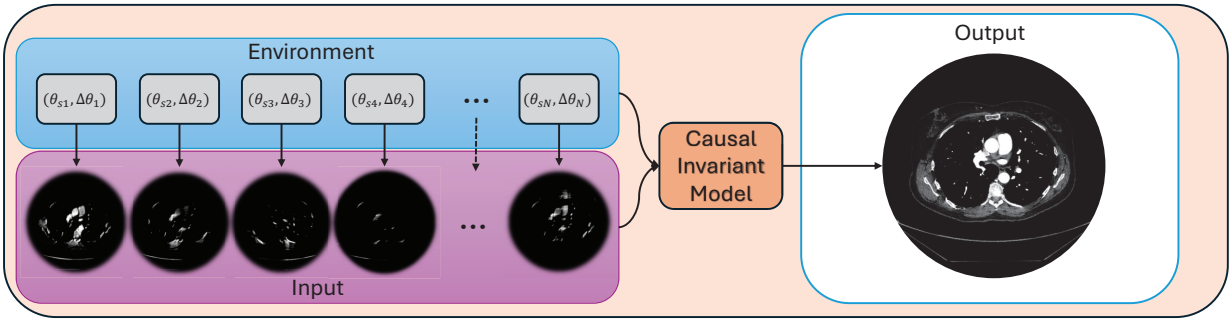


Fig. 2: Causal invariance learning for limited-view-angle CT reconstruction. Multiple acquisition environments are generated by varying the starting angle and angular coverage for the same anatomical slice. Training across these environments encourages the reconstruction network to learn geometry-invariant reconstruction features while suppressing environment-specific artifact patterns.

Specifically, for each training case, the angular coverage is sampled as $\Delta\theta \sim \mathcal{U}(60^\circ, 120^\circ)$ and the starting angle as $\theta_s \sim \mathcal{U}(0^\circ, 360^\circ)$, defining an environment e. A limited-view sinogram is extracted accordingly, and filtered backprojection (FBP) is used to generate an artifact-corrupted reconstruction. The corresponding full-view reconstruction serves as the training target.

This construction exposes the network to a diverse set of artifact patterns for the same anatomy while keeping the desired output fixed. Joint training across these environments, with explicit geometry conditioning, discourages reliance on environment-specific correlations and promotes learning of geometry-invariant reconstruction features. As a result, a single unified model can generalize across a wide range of limited-view-angle configurations without retraining.

III. Details of Training Data and Computing Facilities

Experiments were conducted using a subset of an internally curated coronary CT dataset acquired on a 64-slice CT scanner (GE Discovery CT750, Waukesha, WI). Proprietary data preprocessing software was used to decode the acquired raw data into sinogram projection data together with the corresponding gantry angles. The dataset was split at the patient level into training, validation, and test sets, ensuring that no patient data overlapped across these subsets. The training set consists of 1,600 image slices from 100 patients, the validation set contains 80 slices from 5 patients, and the test set contains 240 slices from 15 patients.

For each image slice, the decoded sinogram data include at least one full-scan angular range together with the associated acquisition geometry. By systematically varying the starting angle and the view angle range, this dataset enables the construction of a rich multi-environment training set for the proposed causal-invariance learning framework. In this work, 20 distinct acquisition environments were generated for each slice, allowing the network to observe a diverse family of geometry-induced artifact patterns during training while preserving anatomical consistency within each slice.

Model selection was performed based on validation performance. The network parameters from the epoch achieving the best validation results (epoch 6) were selected for testing. All training experiments were carried out on a local computing cluster equipped with four NVIDIA L40 GPUs.

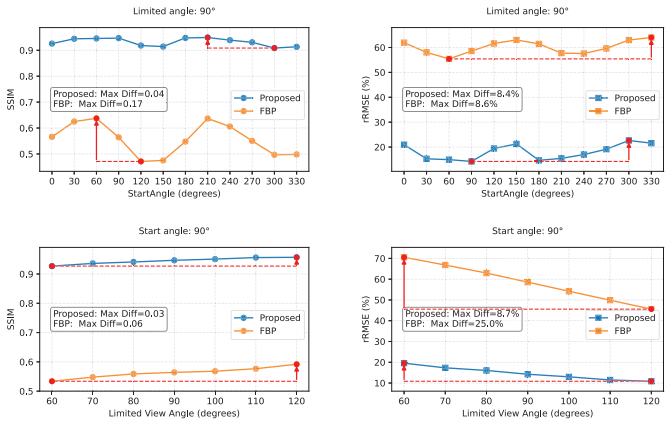


Fig. 3: Quantitative comparison of the proposed method and FBP across acquisition environments. **Top:** performance vs. starting angle θ_s at fixed $\Delta\theta = 90^\circ$. **Bottom:** performance vs. angular coverage $\Delta\theta$ at fixed $\theta_s = 90^\circ$.

IV. Results

We evaluate the generalization behavior of the proposed method under systematic variations of the acquisition environment. The experiments are designed to explicitly assess reconstruction robustness with respect to the environment variable $e = (\theta_s, \Delta\theta)$ for a fixed anatomical slice.

Two complementary testing settings are considered. In the first setting, the starting angle is fixed at $\theta_s = 90^\circ$, while the angular coverage $\Delta\theta$ is varied from 60° to 120° in steps of 10°. This setting evaluates reconstruction performance under different levels of data incompleteness while keeping the acquisition orientation unchanged.

In the second setting, the angular coverage is fixed at $\Delta\theta = 90^\circ$, and the starting angle θ_s is varied from 0° to 330° in steps of 30°. This setting isolates the effect of acquisition orientation changes under a fixed degree of missing data.

For each testing environment, the limited-angle sinogram is reconstructed using filtered backprojection (FBP) to generate

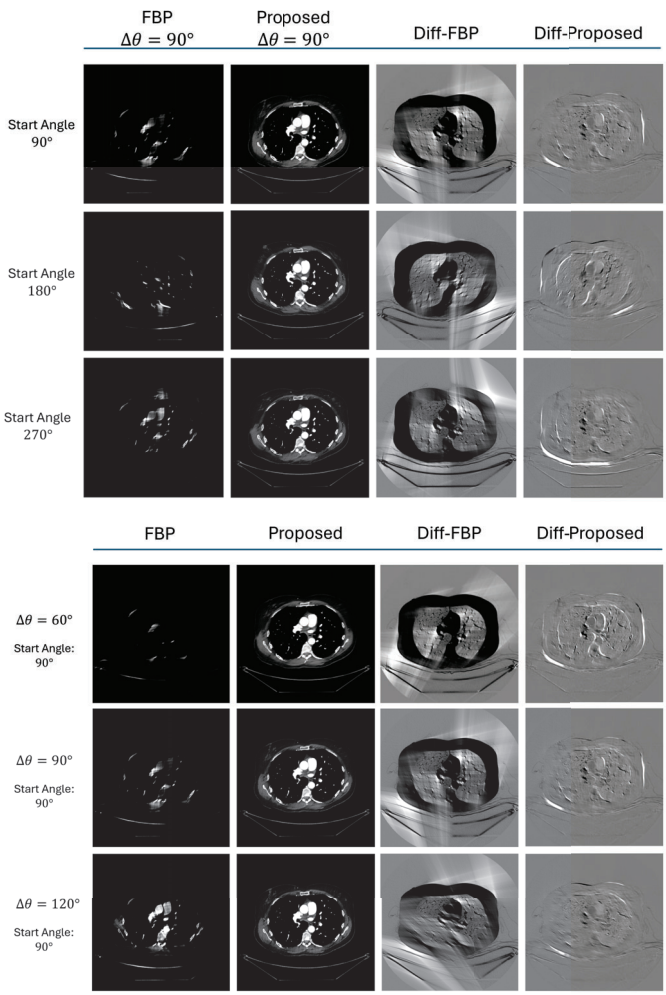


Fig. 4: Visual generalization across acquisition environments. Columns: Direct FBP, Multi-Environment Learning (Proposed), and corresponding difference maps. **Top:** $\theta_s = 90°$ with varying $\Delta\theta$. **Bottom:** $\Delta\theta = 90°$ with varying θ_s.

the network input. The proposed environment-aware reconstruction model is then applied directly without any retraining or fine-tuning.

Representative results are shown in Fig. 4. In the fixed-start-angle setting, the proposed method consistently suppresses limited-view-angle artifacts as $\Delta\theta$ decreases, while preserving anatomical structures across different degrees of data incompleteness. In contrast, FBP reconstructions exhibit increasingly severe streaking and shading artifacts as the angular coverage is reduced.

In the fixed-angular-coverage setting, the proposed method produces reconstructions that remain visually consistent across different starting angles, whereas the artifact patterns in FBP reconstructions vary substantially with acquisition orientation. These observations indicate that the proposed method is less sensitive to orientation-dependent artifact variations.

These visual observations are further supported by quantitative evaluations using rRMSE and SSIM [8], as summarized in Fig. 3. In every experiment, the proposed method significantly improves image quality relative to the input reconstructions. FBP performance, however, depends strongly on acquisition geometry—fluctuating periodically with starting angle and improving steadily as angular coverage increases. The proposed approach mitigates this geometry-induced variability, producing more consistent reconstruction quality across different starting angles and angular ranges, as evidenced by a smaller maximum spread in image-quality metrics.

V. Conclusion

In conclusion, this work shows that robustness to starting angle and angular coverage variations in limited-view-angle CT reconstruction can be achieved by adopting a causal-invariance-inspired, multi-environment learning framework. By explicitly modeling acquisition geometry as an environment variable and training a single reconstruction network across diverse acquisition conditions, the proposed method consistently suppresses limited-angle artifacts while preserving anatomical structures. Experimental results demonstrate stable reconstruction quality across a wide range of starting angles and view angle ranges without retraining, indicating improved robustness and generalization relative to conventional reconstruction approaches. These findings suggest that causal invariance provides a principled and practical framework for designing geometry-robust reconstruction models in limited-view-angle CT.

References

[1] J. Liu, R. Anirudh, J. J. Thiagarajan, S. He, K. A. Mohan, U. S. Kamilov, and H. Kim, "DOLCE: A model-based probabilistic diffusion framework for limited-angle CT reconstruction," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, pp. 10498–10508, 2023.

[2] C. Chen, Y. Xing, H. Gao, L. Zhang, and Z. Chen, "SAM's Net: A self-augmented multistage deep-learning network for end-to-end reconstruction of limited-angle CT," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2912–2924, 2022.

[3] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 234–241, 2015.

[4] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, 2017.

[5] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge, UK: Cambridge University Press, 2009.

[6] J. Peters, P. Bühlmann, and N. Meinshausen, "Causal Inference by Using Invariant Prediction: Identification and Confidence Intervals," *J. R. Stat. Soc. B*, vol. 78, no. 5, pp. 947–1012, Nov. 2016, doi: 10.1111/rssb.12167.

[7] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant Risk Minimization," arXiv:1907.02893, 2019.

[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

Efficient Projection Correction Approach for High-Density Object Artifact Reduction in Digital Breast Tomosynthesis

Subong Hyun, Seoyoung Lee, and Seungryong Cho., *Senior Member, IEEE*

Abstract— **High-density objects such as calcifications and biopsy needles often induce severe artifacts in digital breast tomosynthesis (DBT), including ripples, streaks, and undershoot artifacts, due to the limited-angle reconstruction geometry. These artifacts degrade image quality and may compromise diagnostic accuracy. In this study, we develop an efficient volume-of-interest (VOI)– based method to reduce artifacts from high-density objects in DBT. First, we generate a 2D maximum intensity projection (MIP) image along the depth direction and localize regions of interest (ROIs) for high-density objects on the MIP image. Next, we extract a 3D VOI patch by extruding each ROI box along the depth direction, ensuring the patch covers all DBT slices. We then segment high-density objects within the 3D VOI and compute a forward projection of the resulting mask to obtain a mask in the projection domain. Subsequently, we apply linear interpolation within the masked regions in the projection domain. We reconstruct two DBT volumes from the interpolated projections and the original projections, respectively. Finally, we blend the two reconstructed volumes in the image domain with the high-density object mask in order to suppress artifacts while maintain high-density objects. Experiments on a numerical phantom demonstrate favorable qualitative results and improved quantitative performance. The proposed approach combines MIP-based detection with VOI-based segmentation for efficient localization of high-density objects. We believe that this efficient framework supports its potential for clinical use.**

I. Introduction

Digital breast tomosynthesis (DBT) has emerged as an advanced imaging modality that provides quasi-three-dimensional visualization of the breast through limited-angle tomography. Compared with conventional two-dimensional digital mammography, DBT offers higher specificity, improved lesion conspicuity, and reduced tissue overlap, leading to lower recall rates in breast cancer screening. However, accurate image reconstruction in DBT is inherently constrained by its incomplete projection data acquisition, which violates the tomographic sufficiency condition. Consequently, DBT is highly susceptible to artifacts induced by high-density objects such as metallic biopsy clips, surgical markers, and microcalcifications. These artifacts commonly appear as out-of-plane and in-plane distortions such as ripples, streaks, or undershoot fringes, which obscure low-contrast structures and exaggerate the appearance of calcifications. These artifacts significantly degrade image quality and reduce the visibility and detectability of subtle lesions, thereby impairing diagnostic accuracy. Therefore, to ensure reliable clinical use of DBT, effective suppression of high-density object artifacts is crucial.

Several techniques have been developed to reduce high-density object artifacts in DBT, and these methods can be broadly categorized into voting-based and projection-correction approaches. Wu et al. proposed a voting strategy that estimates whether each voxel is affected by artifacts during backprojection [1]. However, its effectiveness becomes limited when artifacts originate from large objects or from multiple closely located objects, which often leads to degraded performance. Kim et al. integrated backprojection filtration (BPF) with a voting strategy to reduce high-density object artifacts in DBT [2]. In their framework, the voting scheme is applied during differentiated backprojection (DBP), where edge-enhanced responses allow stronger suppression of artifact-corrupted voxels through reduced voxel weights. Although this analytic approach suppresses artifacts effectively, BPF-based reconstruction poses inherent challenges in filter design and often provides less control over desirable image texture. Projection correction approaches typically segment high-density objects in the projection domain and apply interpolation within the segmented regions [3, 4]. The interpolated projections and the corresponding high-density masks are reconstructed separately, and the two reconstructed volumes are then blended in the image domain. However, accurate segmentation in the projection domain is inherently challenging due to tissue overlap, blur, and noise. In addition, the limited-angle acquisition in DBT makes the reconstruction of a projection-domain high-density mask incomplete. In related fields such as cone-beam CT (CBCT), metal segmentation is often performed in the image domain [5]. However, the inherent limited-angle design of DBT makes the reconstruction problem highly underdetermined, which leaves substantial ripple artifacts of high-density object in the reconstructed volume. These ripple artifacts make reliable

This research was supported in part by Korean National Research Foundation [NRF RS-2025-25460686, Project Number: 2710003307, RS-2023-00285089]

Subong Hyun is with the department of nuclear and quantum engineering, KAIST, Daejeon, Republic of Korea. Seoyoung Lee was with the department of nuclear and quantum engineering, KAIST, Daejeon, Republic of Korea. She is now with CT clinical innovation center, radiology, Mayo clinic, Rochester, USA. *Seungryong Cho is with the Department of Nuclear and Quantum Engineering and the Institute for IT Convergence, KAIST, Daejeon, Republic of Korea. (correspondence e-mail: scho@kaist.ac.kr).

image-domain segmentation of high-density objects difficult. Recently, Zhu et al. proposed deep learning–based approaches, such as VBVT-Net [6]. Their method automatically detects and extracts volumes of interest (VOIs) that contain high-density object artifacts and uses the corresponding 4D view-by-view backprojection (VVBP)-tensor patches as network inputs to suppress artifacts effectively. Although their approach does not require explicit segmentation of high-density objects, it still relies on threshold-based segmentation in the projection domain to generate inputs for VOI detection. This reliance limits robustness in practical settings.

In this study, we propose an efficient projection correction approach for high-density object artifact reduction in DBT. To improve the reliability of high-density object localization, we begin with a wBPF reconstruction image, which suppresses ripple artifacts induced by high-density objects. We then compute a 2D maximum intensity projection (MIP) of the wBPF volume along the depth direction and use it as the input to a detection network. The detection network localizes 2D positions of candidate high-density objects on the MIP image. Each detected region is then extruded along the depth direction to form a 3D volume of interest (VOI) that spans all DBT slices. Then, a segmentation network segments high-density object within each VOI. Subsequently, linear interpolation is applied within the masked regions in the projection domain. Two DBT volumes are reconstructed from the interpolated projections and the original projections, respectively. Finally, the two reconstructed volumes are blended in the image domain with the high-density object mask to suppress artifacts while preserving high-density objects. Extensive experiments confirm that the proposed method can effectively suppress the high-density object artifact.

II. METHOD

The overall workflow of the proposed method is shown in Fig. 1.

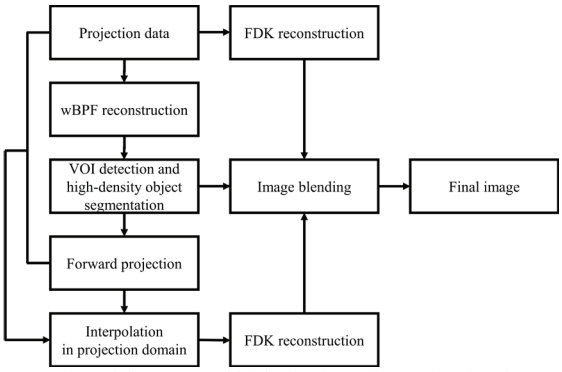


Fig. 1. Overview of the proposed high-density artifact reduction framework.

High-density objects are first identified from the wBPF reconstruction image. The wBPF reconstruction performs a voting process on the DBP image along PI-line segments. Because the DBP image mainly contains only the edge information of the object, the voting result becomes more precise and spatially compact. Figure 2 shows the example of the FDK reconstruction image and the wBPF image of the in-focus and out-of-focus plane.

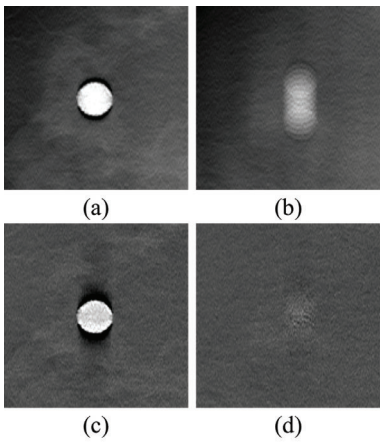


Fig. 2. Comparison of FDK image and wBPF image. (a) FDK (in-focus), (b) wBPF (in-focus), (c) FDK (out-of-focus), and (d) wBPF (out-of-focus)

A. VOI detection network

We trained a detection network to identify VOIs that contain high-density objects in the wBPF image domain. To this end, we adopted the YOLOv8 architecture [7], a well-known detection model that offers fast inference speed and relatively high accuracy. A schematic overview of the VOI detection process is shown in Fig. 3.



Fig. 3. Schematic illustration of the VOI detection process.

To improve computational efficiency, we reduced the spatial dimensionality by computing a maximum intensity projection (MIP) along the depth direction and used the resulting 2D MIP image as the network input. Network outputs are the predicted bounding-box parameters $\hat{b} = (\hat{x}, \hat{y}, \hat{w}, \hat{h})$, an objectness probability $\hat{q}$, which indicates the presence of any object, and a class probability $\hat{p}$ corresponding to the high-density object class. Finally, for each bounding box, we compute a detection confidence $c = \hat{q} \cdot \hat{p}$, and VOIs are defined as the high-confidence boxes that remain after removing overlapping boxes corresponding to the same object. Notably, because high-density object segmentation is later performed on every slice along the depth direction within each detected patch during the segmentation stage, the depth position of the VOIs does not need to be determined.

The loss function of the detection network L_{VOI} is defined as follows:

$$L_{VOI} = \lambda_{box} L_{box}(\hat{b}, b) + \lambda_{dfl} L_{dfl}(\hat{r}, r) + \lambda_{obj} L_{obj}(\hat{q}, q) + \lambda_{cls} L_{cls}(\hat{p}, p), \quad (1)$$

where L_{box} is the bounding-box regression loss that measures how well the predicted box $\hat{b}$ overlaps with the ground-truth box b using an intersection-over-union (IoU)-based metric L_{dfl} is the distribution focal loss for precise

boundary localization by comparing the predicted discrete distributions $\hat{r}$ for the four boundaries (left, top, right, bottom) with the target distributions r derived from the ground-truth box. L_{obj} is the objectness loss, implemented using binary cross-entropy (BCE) to encourage accurate discrimination between object and background regions. Finally, L_{cls} is the classification loss, also implemented using BCE, which guides the predicted class score toward the binary label of the high-density object.

B. *High-density object segmentation network*

High-density object segmentation was performed based on the bounding boxes obtained from VOI detection module. Fig. 4 illustrates the high-density object segmentation pipeline within the VOIs. For each bounding box, the corresponding 3D VOI patch was extracted by extruding the box along the depth direction to cover all slices, and the resulting 3D wBPF patch was used as the network input. The network then produced a 3D high-density object mask for each patch, which was subsequently placed back into its original location within the full 3D image volume. Finally, forward projection of this masked volume was performed to generate a projection-domain mask P_{mask}, which was used to identify regions in the projection data affected by high-density objects.

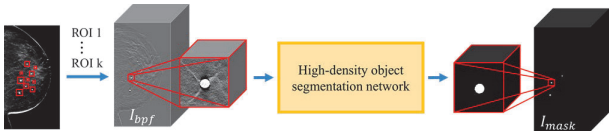


Fig. 4. Schematic illustration of high-density object segmentation process.

As the network backbone, we employed the Swin UNETR architecture [8], which combines a Swin Transformer encoder with a UNet-style decoder to capture long-range dependencies while preserving multi-scale spatial details with 3D layers, and is widely used for medical image segmentation. It is worth noting that although the spatial size of each VOI patch varied across cases, this transformer-encoder–convolutional-decoder architecture is capable of processing VOIs of various sizes.

The loss function for the segmentation network L_{seg} is defined as follows:

$$L_{seg} = \lambda_{dice} L_{dice}(\hat{y}, y) + \lambda_{BCE} L_{BCE}(\hat{y}, y), \quad (2)$$

$$L_{dice} = 1 - \frac{2\sum_{n=1}^{N} \hat{y} y}{\sum_{n=1}^{N} \hat{y}^2 + \sum_{n=1}^{N} y^2}, \quad (3)$$

$$L_{BCE} = -\frac{1}{N}\sum_{n=1}^{N} [y \log(\hat{y}) - (1-y)\log(1-\hat{y})], \quad (4)$$

where L_{seg} and L_{BCE} represent dice loss and BCE loss, respectively. y is the real class label, and $\hat{y}$ is the predicted value.

C. *Interpolation and reconstruction*

We obtain a corrected projection set by applying linear interpolation to the masked regions obtained in Section 2.B. A DBT volume is reconstructed from the corrected projections, and another volume is reconstructed from the original projections. Finally, the two reconstructed volumes are blended in the image domain with the high-density object mask. The masked region follows the volume reconstructed from the original projections to preserve high-density objects, whereas the remaining region follows the volume reconstructed from the corrected projections to suppress artifacts.

III. Experiment

A. *Dataset preparation*

We generated realistic numerical breast phantoms using the VICTRE toolkit [9] developed by the U.S. Food and Drug Administration 20. We designed two types of phantoms: fatty breasts with a glandular volume fraction (GVF) of 0.85–1.0, and dense breasts with a GVF of 0–0.35. The compressed thicknesses were set to 4 cm and 7 cm. The voxel size of the phantoms was set to 0.1×0.1×0.1 mm³. In total, 120 breast phantoms were generated. Calcification objects, as illustrated in Fig. 5, were inserted into a subset of the VICTRE phantoms. We denote the phantom set with calcifications as $\mathcal{A}^{w}$ and the corresponding set without calcifications as $\mathcal{A}^{wo}$.

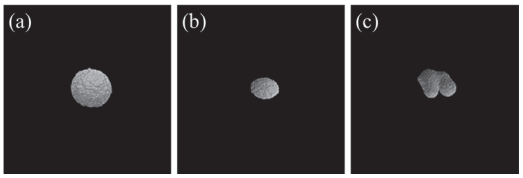


Fig. 5. Illustration of the inserted calcification-like objects, with sizes ranging from 1.0 to 3.0 mm.

Monte Carlo simulations were performed using MC-GPU. The number of projections was 9, covering a scan angle range from −7.5° to 7.5°. The detector size was 3840×3072 pixels with a pixel size of $0.076\times0.076 mm^2$.

B. *Implementation details*

The YOLOv8 model used for VOI detection was initialized with ImageNet-pretrained weights, then further fine-tuned on our DBT dataset. The width and height of the ROI were then chosen so as to fully encompass all high-density objects. The VOI detection network was trained for 300 epochs with a learning rate of 1×10^{-4}, batch size of 4. In the loss function of the detection network, L_{VOI}, we empirically set the weighting parameters to λ_{box}=7.5, λ_{dfl}=1.5, and λ_{obj}=0.5.

The segmentation network was trained for 100 epochs with a learning rate of 2×10^{-4} and a batch size of 1, due to the varying sizes of the 3D VOIs. In the loss function of the segmentation network, L_{seg}, we empirically set the weighting parameters to λ_{dice}=1.0 and λ_{obj}=1.0.

IV. Results

A. *Detection result*

Figure. 6 presents the confusion matrix for the proposed detection network. This result indicates that the proposed detector provides reliable candidate locations for subsequent VOI extraction.

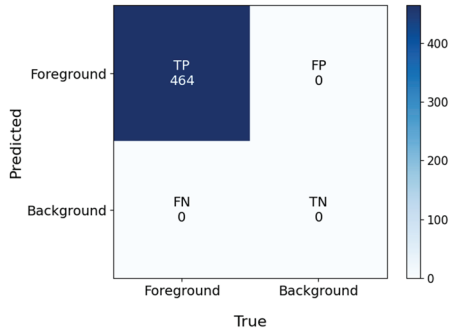


Fig. 6. Confusion matrices for VOI detection on the VICTRE phantom.

B. Segmentation result

The quantitative segmentation performance on the VICTRE dataset in terms of dice similarity coefficient (DSC), Hausdorff distance (HD), and mean surface distance (MSD) summarized in Table I.

Table I. Quantitative evaluation of segmentation on the VICTRE phantom.

	DSC	HD	MSD
VICTRE	0.962±0.027	0.610±0.485	0.052±0.044

Overall, the proposed method achieves a high DSC, which indicates a strong overlap between the predicted masks and the reference annotations. The HD results suggest that extreme boundary errors remain limited, while the low MSD further confirms accurate surface alignment between the predicted and ground-truth segmentations.

C. High-density object artifact reduction result

Figure 7 provides representative examples of high-density object artifact correction. In the uncorrected images, prominent ripple patterns appear around calcification. After the proposed correction, these ripple artifacts are markedly reduced while the high-density objects remain clearly visible.

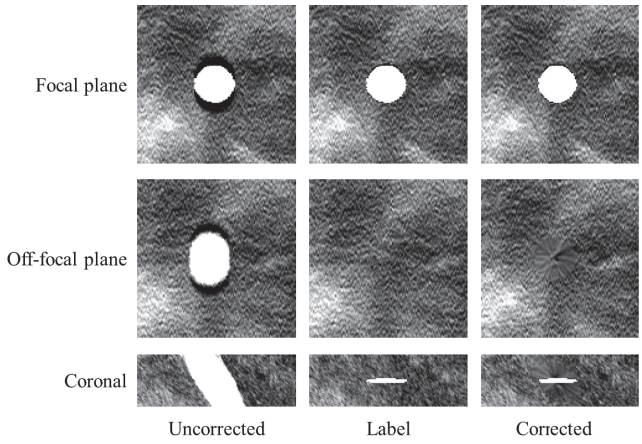


Fig. 7. High-density object artifact correction results. (a) Uncorrected image, (b) Corrected image

Table II summarizes the mean-squared error (MSE) and learned perceptual image patch similarity (LPIPS) measured in soft-tissue background regions. Compared with the uncorrected reconstruction, the corrected images show substantially lower MSE and LPIPS, which indicates reduced reconstruction error and improved perceptual similarity in artifact-affected backgrounds. Overall, these results confirm that the proposed correction effectively suppresses high-density object–induced artifacts while improving soft-tissue appearance.

Table II. Quantitative evaluation of segmentation on the VICTRE phantom.

	MSE (10^{-3})	LPIPS (10^{-3})
Uncorrected image	5.66±3.26	4.80±4.84
Corrected image	0.08±0.02	0.05±0.02

V. DISCUSSION AND CONCLUSION

This study proposes an efficient projection-correction framework for reducing high-density object artifacts in DBT. The pipeline uses a wBPF reconstruction for reliable localization and applies MIP-based VOI detection with VOI-based segmentation to avoid costly whole-volume processing. The resulting mask guides projection-domain correction and image-domain blending, which suppresses artifacts while preserving high-density structures. Experiments on numerical breast phantoms show strong detection and segmentation performance and consistent ripple artifact reduction. Nevertheless, the current study has limitations. The method was evaluated only on numerical phantoms, and the variety of high-density objects remains limited. In addition, linear interpolation can leave residual or visually unnatural artifacts. Future work will include training with more diverse objects, evaluation on clinical data, and the development of more natural projection-domain inpainting methods.

REFERENCES

[1] T. Wu *et al.*, "Voting strategy for artifact reduction in digital breast tomosynthesis," *Med Phys,* vol. 33, no. 7, pp. 2461-71, Jul, 2006.

[2] H. Kim *et al.*, "Backprojection Filtration Image Reconstruction Approach for Reducing High-Density Object Artifacts in Digital Breast Tomosynthesis," *IEEE Trans Med Imaging,* vol. 38, no. 5, pp. 1161-1171, May, 2019.

[3] Y. Lu *et al.*, "Improving image quality for digital breast tomosynthesis: an automated detection and diffusion-based method for metal artifact reduction," *Physics in Medicine and Biology,* vol. 62, no. 19, pp. 7765-7783, Oct 7, 2017.

[4] H. Tang *et al.*, "A new projection correction based voting strategy for breast calcification artifact reduction," *Phys Med Biol,* vol. 68, no. 18, Sep 11, 2023.

[5] D. Choi *et al.*, "Metal artifact reduction algorithm with conditional latent diffusion model for dental cone-beam CT," *Journal of Applied Clinical Medical Physics,* vol. 26, no. 11, Oct 24, 2025.

[6] M. M. Zhu *et al.*, "VBVT-Net: VOI-Based VVBP-Tensor Network for High-Attenuation Artifact Suppression in Digital Breast Tomosynthesis Imaging," *Ieee Transactions on Medical Imaging,* vol. 44, no. 4, pp. 1953-1968, Apr, 2025.

[7] D. Reis *et al.*, "Real-time flying object detection with YOLOv8," *arXiv preprint arXiv:2305.09972*, 2023.

[8] J. Hwang *et al.*, "Improving segmentation precision in prostate cancer adaptive radiation therapy with a patient-specific network," *Plos One,* vol. 20, no. 9, Sep 19, 2025.

[9] A. Badano *et al.*, "In silico imaging clinical trials for regulatory evaluation: initial considerations for VICTRE, a demonstration study." pp. 494-499.

Prior-Embedded Implicit Neural Representation with Dual-Domain Ordered-Subsets for High-Resolution Limited-Angle Non-Coplanar CBCT Imaging

Siqi Ye, Yu Gao, Hao Zhang, and Lei Xing*

***Abstract*—Non-coplanar radiation therapy is an emerging technique in radiation therapy to improve dose conformity and organ-at-risk sparing. However, mechanical constraints in non-coplanar treatment setups often restrict gantry motion, making limited-angle non-coplanar CBCT (LA-NC-CBCT) the only feasible volumetric imaging option. Severe angular incompleteness in LA-NC-CBCT acquisition leads to pronounced artifacts and degraded image quality, posing significant challenges for image-guided radiation therapy. To address this problem, we propose a computationally efficient, patient-specific reconstruction framework based on a prior-knowledge-embedded implicit neural representation (PINR). The INR is pretrained using a routinely available full-scan coplanar CBCT acquired at the initial patient position. In the reconstruction stage, a motion-encoded regularization combined with a data-fidelity term built upon limited-angle acquisitions is used to update the network weights to generate the target NC-CBCT image. Additionally, we propose a novel dual-domain ordered-subsets (DDOS) optimization strategy to address GPU memory limitations and enable efficient reconstruction. Experiments on physical phantom measurement data demonstrate that the proposed method consistently outperforms conventional FDK and reconstruction-by-registration baselines across a wide range of limited-angle scenarios. The runtime to reconstruct a $512 \times 512 \times 94$ image volume is less than one minute using a computer equipped with a single 48 GB GPU.**

***Index Terms*—limited-angle CBCT image reconstruction, ordered-subsets, neural representation, image registration, non-coplanar radiation therapy, image-guided radiation therapy**

I. Introduction

Limited-angle CBCT (LA-CBCT) scans are often the only viable option for non-coplanar radiation therapy to prevent potential mechanical collisions. However, the consecutive angular occlusion of projections in LA-CBCT makes reconstructing high-quality volumetric images extremely challenging, hindering efficient and precise treatment adaptation and dose delivery in radiation therapy.

Data-driven deep learning has been extensively explored for LA-CBCT imaging. Post-processing in the image domain is one of the main strategies to produce high-quality CBCT images with limited-angle or sparse acquisitions [1]–[4]. Trained on large datasets of paired low- and high-quality images, these approaches can generate enhanced images within seconds during inference. However, residual artifacts inherited from the low-quality inputs often remain and cannot be fully corrected without direct guidance from measurement data.

Physics-informed end-to-end reconstruction provides an alternative solution by learning a direct mapping from projection data to images. Using paired projection-image datasets, recent deep learning frameworks, such as encoder-decoder [5] and transformer-based models [6], [7], have demonstrated the ability to reconstruct high-quality images from extremely sparse, or even single-projection, acquisitions [5]. Despite their success, these methods rely heavily on large and diverse paired datasets, which are often impractical to acquire in clinical settings, thereby limiting their ability to achieve accurate patient-specific reconstruction [8], [9]. Deep-learning-enhanced model-based image reconstruction (MBIR) offers a promising compromise by explicitly modeling imaging physics and noise statistics while incorporating learned priors from limited data [10]–[16]. These methods enable patient-specific optimization but typically incur substantial computational cost due to repeated forward and backward projections during iterative reconstruction. Moreover, their application to non-coplanar (NC) imaging remains largely unexplored, where available projection angles can be severely restricted to less than 90°.

Early studies on NC-CBCT imaging were developed within a prior-image-constrained compressed sensing (PICCS) framework [17], and GAN-based coplanar-to-NC image mapping framework [18]. These methods demonstrate promising performance under reduced angular coverage of 60°. Implicit neural representation (INR) has recently emerged as a powerful paradigm for solving challenging

Asterisk indicates the corresponding author. S. Ye was with Stanford University, and is currently with the Department of Radiation Oncology, Mayo Clinic, Rochester, MN, 55901, USA (email: ye.siqi@mayo.edu). Y. Gao, H. Zhang, and L. Xing are with the Department of Radiation Oncology, Stanford University, Stanford, CA, 94305, USA. (emails: yugao97, hzhang96, lei@stanford.edu)

This work was partially supported by NIH (1R01CA275772 and 5R01CA256890), an interdisciplinary seed grant from Bio-X of Stanford University, and VIBE award funds from Mayo Clinic.

inverse problems, including CT image reconstruction [19]–[24]. By employing coordinate-based multilayer perceptrons (MLPs), INR models learn patient-specific mappings from spatial coordinates to image intensities and naturally support resolution enhancement by querying denser coordinate grids. IntraTomo [20] was among the first to apply INR to limited-angle tomography, achieving satisfactory results with a $45°$ scan. However, this method is restricted to 2D reconstruction at a resolution of 256×256 due to the high memory demands of MLP-based representations. To enable large-scale 3D reconstruction, NeAT [22] encodes the volumetric coordinate grid as a sparse octree and approximates X-ray line integrals as finite sums over segmented ray paths. This adaptive strategy substantially reduces memory consumption and supports gradient backpropagation through domain transformations. Nevertheless, the partial line-integral approximation can introduce inaccuracies arising from suboptimal octree partitioning, potentially leading to noise and artifacts, particularly in limited-angle scenarios with severe angular occlusion.

Motivated by these limitations, this paper presents a novel INR-based framework, termed PINR-DDOS, for efficient and high-resolution LA-NC-CBCT imaging in image-guided non-coplanar radiation therapy.

II. Method

A. Method Overview

The proposed PINR-DDOS reconstruction framework incorporates both a high-quality prior coplanar image and the geometric relationship between the coplanar and NC positions. In a patient-specific framework, an INR model is first pretrained to represent a full-scan CBCT volume acquired at the coplanar position. The INR takes spatial coordinate grids as input, which are mapped to a higher-dimensional feature space using a Fourier feature embedding (FFE) module [25] to facilitate the representation of high-frequency image details. A multi-layer perceptron (MLP) then outputs the target image volume, and the network parameters are optimized by minimizing the mean squared error with respect to the reference CBCT volume, thereby encoding prior anatomical information into the network. During reconstruction, the encoded prior image is approximately aligned to the target geometry via a rigid transformation derived from recorded treatment couch motions, yielding a pseudo reference at the NC position. The spatial coordinate grids are transformed accordingly so that the INR generates the image in the NC configuration. Projection-domain data fidelity is enforced using the acquired LA-NC-CBCT projections.

To enable efficient high-resolution reconstruction, we introduce a dual-domain ordered-subsets (DDOS) strategy. In the image domain, a novel OS scheme inspired by super-resolution principles and column-wise matrix–vector multiplication is employed, while in the projection domain, a conventional OS strategy [26], [27] is adopted based on

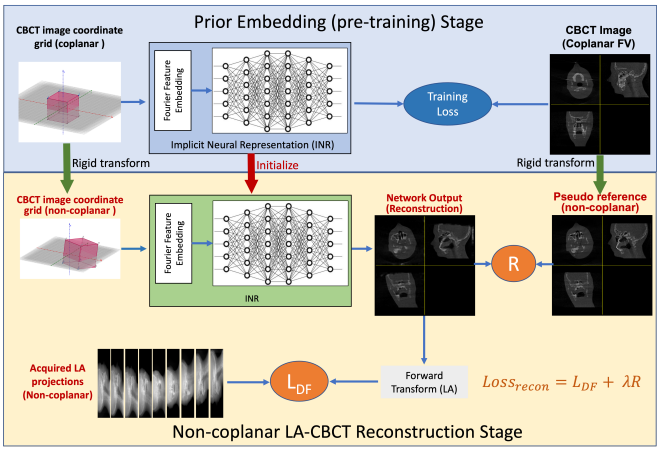


Fig. 1: Framework of LA-CBCT image reconstruction for non-coplanar radiation therapy. The abbreviations "FV" and "LA" denote full-view and limited-angle CBCT acquisitions, respectively.

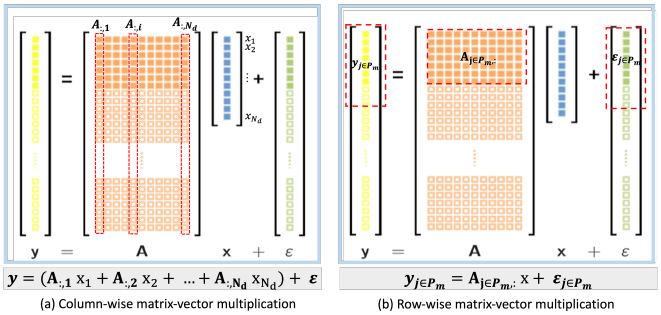


Fig. 2: Illustration of (a) image-domain OS and (b) projection-domain OS with the matrix-vector multiplication theory in the linear imaging system.

the row-wise structure of the system matrix. The DDOS partitioning is illustrated in Fig. 2.

B. Problem Formulation

1) Dual-Domain Ordered-Subsets

We design a novel image-domain ordered-subset strategy inspired by super-resolution principles: (1) a spatially downsampled image provides a low-resolution representation that preserves the global structure of the original high-resolution volume, and (2) a set of downsampled images with complementary spatial shifts jointly contains sufficient information to reconstruct the original high-resolution image.

Denoting $\mathbf{A} \in \mathbb{R}^{N_p \times N_d}$ as the forward transform operator for the original high-resolution volume, with N_p and N_d representing the number of voxels in the entire projection sets and the number of voxels in the original volumetric image. Denoting the vectorized image as $\mathbf{x} \in \mathbb{R}^{N_d}$ and the vectorized projections as $\mathbf{y} \in \mathbb{R}^{N_p}$, we formulate the data fidelity loss as $L_{DF}(\mathbf{x}) = \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2$. Applying the column-wise matrix-vector multiplication theory to this linear imaging system, we can rewrite this data-fidelity function as follows:

$$L_{DF}(\mathbf{x}) = \frac{1}{2}\|\sum_{s=1}^{S}\mathbf{A}_{:,i\in I_s}\mathbf{x}_{i\in I_s} - \mathbf{y}\|^2 \approx \frac{1}{2}\sum_{s=1}^{S}\|\mathbf{A}_{:,i\in I_s}\mathbf{x}_{i\in I_s} - \frac{1}{S}\mathbf{y}\|^2, \tag{1}$$

where $i \in I_s$ indicates the image-domain pixel i and corresponding i-th column in the forward matrix falling into the s-th subset denoted by I_s. The total number of image-domain OS is S. The "$\approx$" holds when the sampled volume can represent the global volume at a lower resolution version.

Projection-domain OS is further incorporated to reduce memory usage and accelerate the reconstruction [26], [27]. In large-scale tomographic image reconstruction, gradient computation involving forward and backward projections is computationally expensive. The projection-domain OS partitions projections into subsets corresponding to downsampled projection angles, each subset can provide an approximation of the exact gradient of the cost function. By partitioning the projections uniformly according to their scanning angles, the data-fidelity loss in Eq. (1) can be further approximated, yielding the final dual-domain ordered-subset–based data-fidelity loss for image reconstruction:

$$L_{DF}(\mathbf{x}) \triangleq \frac{1}{2}\sum_{m=1}^{M}\{\sum_{s=1}^{S}\|\mathbf{A}_{j\in P_m,i\in I_s}\mathbf{x}_{i\in I_s} - \frac{1}{S}\mathbf{y}_{j\in P_m}\|^2\}, \tag{P0}$$

where $j \in P_m$ denotes the j-th projection falling into the m-th OS in the projection domain. Iterating over each individual projection subset can lead to updates on entire input image, and the accumulated updates after one cycle pass of all the project subsets can result in significant improvement on the target, which can eventually reduce the number of outer iterations and total runtime.

2) PINR-DDOS Reconstruction Formulation

The proposed PINR-DDOS reconstruction framework incorporates two types of prior knowledge: the geometric relationship between the coplanar and NC positions, and the high-quality prior coplanar image. To encode the geometric relationship prior, we introduce a registration-involved regularization term in the reconstruction loss function. Specifically, the prior coplanar image $\mathbf{x}^0$ is warped into a pseudo-reference NC image $\mathbf{x}^r$ using the rigid motions recorded by the treatment machine, and the registration-involved regularizer is then defined as $R(\mathbf{x}) = \|\mathbf{x} - \mathbf{x}^r\|_2^2$. The high-quality coplanar image prior is modeled by the INR network.

The INR network maps the object's spatial location to the corresponding intensity value, i.e., $\mathbf{c}_i \mapsto \mathcal{M}(\phi; \mathbf{c}_i) = x_i$, where $\mathbf{c}_i \in \mathbb{R}^3$ denotes the three-dimensional spatial coordinate of a single voxel, and $\mathcal{M}(\phi)$ represents the network. We pretrain the INR network to overfit to the high-quality volumetric image obtained at the coplanar position via

$$\phi^0 = \underset{\phi}{\operatorname{argmin}}\,\|M(\phi; \mathbf{c}^0) - \mathbf{x}^0\|_2^2, \tag{2}$$

Algorithm 1 PINR-DDOS Image Reconstruction Algorithm

Input:
1: Pretrained network $\mathcal{M}(\phi^0)$ from coplanar CBCT scan;
2: Limited-angle projection acquisitions $\mathbf{y}$ under the NC setup;
3: Transformed spatial coordinate grid of 3D volume corresponding to the NC position: $\mathbf{c}^r$;
4: Imaging geometry parameters and acquisition angles to create system matrix $\mathbf{A}$;
5: Number of projection-domain ordered subsets M for projection-domain partition;
6: Number of image-domain ordered subsets S for image-domain partition;
7: Number of outer iterations N.
Output: Reconstructed CBCT image at the non-coplanar position: $\mathbf{x}^*$.
1: **for** $n = 1, 2, \cdots, N$ **do**
2: **for** $m = 1, 2, \cdots, M$ **do**
3: Sample projection subsets $\mathbf{y}_m = \mathbf{y}_{j\in P_m}$;
4: **for** $s = 1, 2, \cdots, S$ **do**
5: Create forward operator $\mathbf{A}_{j\in P_m, i\in I_s}$ for the OS-specific imaging geometry;
6: Compute image subset volume with updated network $\mathcal{M}(\phi^{(n*m)}; \mathbf{c}^r_{i\in I_s})$;
7: Compute image OS loss based on Eq. (4);
8: Accumulate losses over image OS to form Eq. (4);
9: **end for**
10: Update the network parameter $\phi^{(n*m+1)}$ based on Eq. (4) via the Adam optimizer;
11: **end for**
12: **end for**
13: Optimized network parameters $\phi^* = \phi^{(N*M)}$;
14: Reconstructed image $\mathbf{x}^* = \mathcal{M}(\phi^*; \mathbf{c}^r)$.

where $\mathbf{c}^0$ represents the coordinate grid at the coplanar position. To obtain the reconstruction at an NC position for the patient, we fine-tune the network under the transformed coordinate grid $\mathbf{c}^r$ corresponding to the NC position, i.e., $\mathbf{x}^* = \mathcal{M}(\phi^*; \mathbf{c}^r)$.

By substituting the explicit image vector $\mathbf{x}$ with its INR representation $\mathcal{M}(\phi; \mathbf{c})$, the entire reconstruction loss function for LA-NC-CBCT can be formulated as

$$L(\phi) = \frac{1}{2}\|\mathbf{A}\mathcal{M}(\phi; \mathbf{c}^r) - \mathbf{y}\|^2 + \lambda\|\mathcal{M}(\phi; \mathbf{c}^r) - \mathbf{x}^r\|_2^2, \tag{3}$$

where $\lambda > 0$ is a scalar to balance the data-fidelity term and the regularizer. By incorporating the DDOS strategy described above and further approximating the data-fidelity term for each projection-domain OS, we derive a per–projection-domain loss that supports the per-epoch up-

date of the entire network:

$$L_m(\phi) = \frac{1}{2} M \Big\{ \sum_{s=1}^{S} \|\mathbf{A}_{j\in P_m, i\in I_s} \mathcal{M}(\phi; \mathbf{c}^r_{i\in I_s}) - \frac{1}{S}\mathbf{y}_{j\in P_m}\|^2 \Big\} + \lambda \sum_{s=1}^{S} \|\mathcal{M}(\phi; \mathbf{c}^r_{i\in I_s}) - \mathbf{x}^r_{i\in I_s}\|^2. \tag{4}$$

Algorithm 1 presents the pseudo code of the proposed algorithm for PINR-DDOS-based image reconstruction.

III. Experiments

A. Experimental Setups

Experiments were performed on an anthropomorphic head-and-neck phantom scanned using the onboard CBCT system of a Varian TrueBeam linear accelerator (Varian Medical Systems, Palo Alto, CA). A full-view coplanar scan with 200^o angular coverage was acquired to generate a high-quality CBCT volume for network pretraining. The reconstructed volume has a size of $512 \times 512 \times 94$, with resolutions of 0.511 mm $\times$ 0.511 mm $\times$ 1.988 mm. The phantom was then repositioned to NC configurations, where only limited-angle projections were available. Projections were acquired on a 1424×268 detector with 0.28 mm $\times$ 1.12 mm resolution.

Results are presented for one NC configuration, with couch rotations of -9.96^o, -1.57^o, and -1.34^o (yaw, pitch, roll) and translations of 12.4 mm, 12.4 mm, and 10.9 mm (vertical, longitudinal, lateral). The INR consists of an 8-layer MLP with 256 neurons per layer. The image-domain OS size is $[128, 128, 94]$, with up to 80 projections per projection-domain OS on a 48-GB GPU. The network is optimized using Adam for 10 epochs with an initial learning rate of 2×10^{-6}, reduced by a factor of 10 every two iterations.

B. Reconstruction Results

Fig. 3 displays the volumetric reconstruction results. Each reconstruction shows the axial (top left), sagittal (top right), and coronal (bottom left) views of the volumetric image. The ground-truth (GT) image was obtained using the standard FDK reconstruction approach applied to full-view acquisitions at the NC positions. The "Pseudo-Ref" image corresponding to $\mathbf{x}^r$ was obtained by rigidly transforming the initial coplanar CBCT volume according to the couch motion parameters. While the "Pseudo-Ref" preserves high-frequency anatomical details, rigid registration that is based solely on couch motion parameters can introduce inaccuracies and misalignment artifacts. Our proposed approach can significantly improve reconstruction quality, achieving an increase of approximately $0.5\ dB$ in PSNR while preserving high anatomical structural similarity to the GT image, as indicated by the SSIM metric. Compared with the FDK reconstructions, the proposed approach achieves PSNR and SSIM improvements ranging from $4.2\ dB$ to $12.3\ dB$ and 0.01 to 0.26, respectively, as the angular coverage decreases from 90° to 15°, demonstrating strong robustness and consistent performance across varying limited-angle conditions. We implemented the experiments on a Linux server with dual Intel Xeon Silver 4316 CPUs (80 threads total) and one NVIDIA RTX A6000 GPU (48 GB). The average runtime was 7 seconds per epoch, and it took less than one minute to get a convergent reconstruction result.

C. Convergence Behavior

Figures 4a and 4b present the PSNR and SSIM convergence curves for LA-NC-CBCT reconstruction. The proposed method demonstrates fast and stable convergence across all angular ranges, with substantial improvements achieved within the first three to five iterations. Notably, even under severely limited angular coverage, the optimization process remains stable, without observable oscillation or divergence.

Figure 4c shows the evolution of the data-fidelity term during reconstruction. The loss is normalized per projection and reflects consistency with the measured data. In all cases, the data-fidelity loss decreases rapidly during the initial iterations and then gradually stabilizes, indicating reliable convergence behavior. As the angular coverage decreases below 60°, the average per-projection data-fidelity loss increases, which is expected due to the increased ill-posedness of the reconstruction problem under more limited scanning angles. Interestingly, the 90° case does not achieve the lowest loss value. This can be attributed to the stronger measurement constraints imposed by a larger number of projections, making it more challenging to simultaneously minimize residuals across all observations.

Figure 4d illustrates the evolution of the image-domain regularization term as a function of global iterations (defined as the product of outer and inner iterations). The solid markers indicate outer iterations. The regularization loss is slightly higher at the initial stages but remains on a comparable scale to the data-fidelity term. This behavior is consistent with the use of a relatively large initial regularization parameter λ_0, which helps prevent significant deviation from the high-quality prior image. Within each outer iteration, the regularization term remains nearly constant. This is expected because: (1) the image-domain discrepancy is inherently smaller in magnitude than the projection-domain residuals; and (2) the pretrained INR already provides a high-quality initialization that closely matches the pseudo-reference image obtained via warping of the coplanar prior. As the regularization parameter decays exponentially over iterations, the contribution of the regularization term correspondingly decreases. This indicates that, in later stages, the reconstruction is increasingly driven by the data-fidelity term, allowing the model to better adapt to the measured projections while maintaining anatomical consistency.

IV. Discussion

The dual-domain ordered-subsets design for large-scale image reconstruction significantly improves the computational efficiency under limited computational resources. Particularly, the image-domain partitioning, motivated by super-

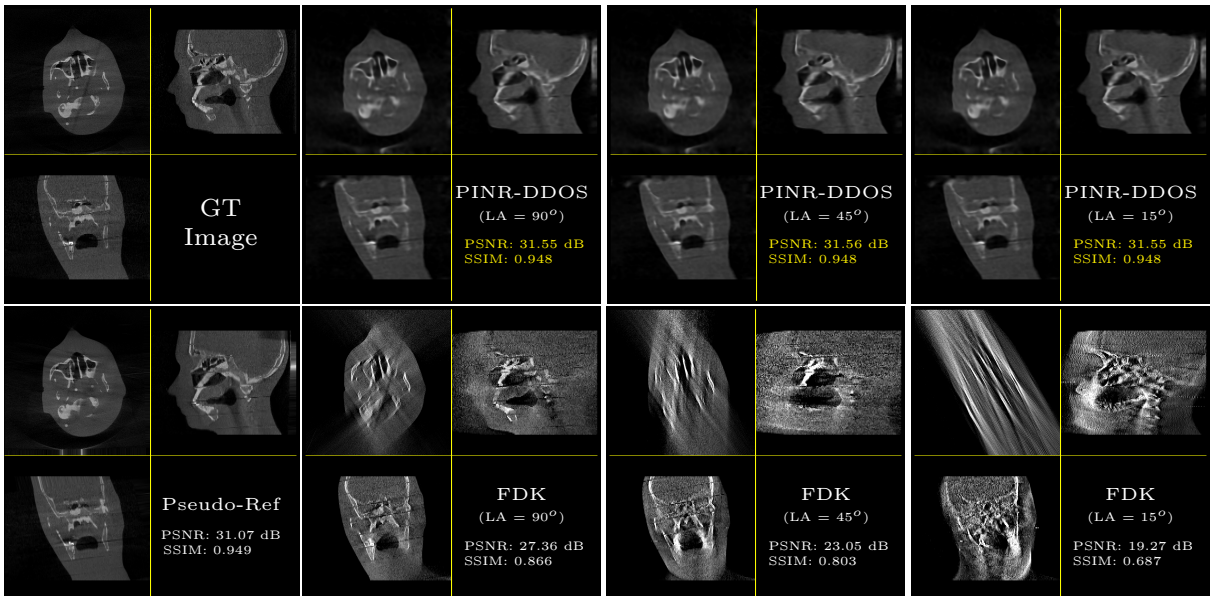
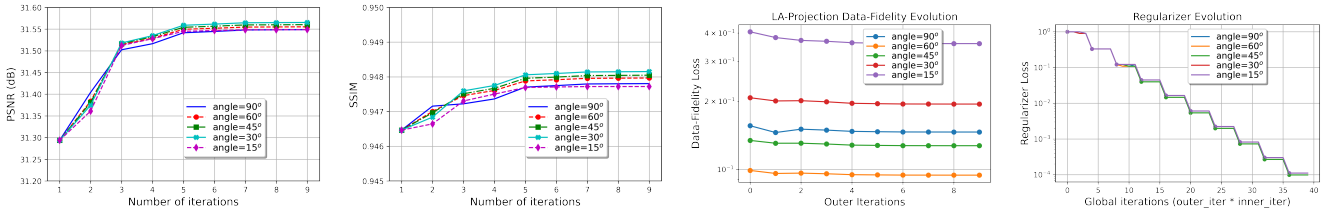


Fig. 3: Reconstruction quality comparisons of the experimental phantom study.



(a) PSNR. (b) SSIM. (c) Data-fidelity term. (d) Regularization term.

Fig. 4: Convergence behavior of the proposed PINR-DDOS method.

resolution concepts and column-wise matrix-vector multiplication, significantly reduces the per-iteration computational and memory demands associated with reconstruction problems involving domain transformations.

Additionally, the INR network offers a spatially continuous representation that is well initialized through pretraining on a full-scan coplanar CBCT image. This initialization places the optimization close to a physically plausible solution manifold, thereby reducing the number of iterations required for convergence compared with training from scratch or using conventional pixel-based representations.

Regarding scalability, in our experiments, the image volume is divided into subsets of size $[128, 128, 94]$, with 80 projections per projection-domain subset to fit within a single 48 GB GPU. This configuration can be adapted to smaller subset sizes for GPUs with more limited memory, albeit at the cost of increased iterations due to a larger number of subsets and, consequently, longer computation time. In our current setup, processing all subsets once takes approximately seven seconds, indicating that the additional computational overhead remains acceptable in practice.

Furthermore, the proposed image-domain ordered-subset strategy is readily extendable to large-scale tomographic imaging problems beyond LA-CBCT, including high-resolution sparse-view CT and tomographic microscopy. Although this work focuses on an INR-based implementation, the dual-domain strategy itself is model-agnostic and can be integrated with other neural network architectures, such as transformers.

Finally, supported by the fast, robust, and high-quality reconstructions demonstrated in this proof-of-concept study, the proposed method shows strong potential to enable more efficient and broader adoption of non-coplanar radiation therapy. Future work will extend evaluation to patient data and task-based clinical endpoints, including patient positioning accuracy, anatomical delineation accuracy, and dosimetric impact in non-coplanar treatment workflows.

V. Conclusion

We propose PINR-DDOS, a prior-knowledge-embedded implicit neural representation with a dual-domain ordered subsets framework for limited-angle non-coplanar CBCT reconstruction. The method integrates anatomical priors from routine coplanar full-view CBCT scans with accurate modeling of non-coplanar geometry and couch motion in a unified formulation. The DDOS strategy ensures computational efficiency and scalability. Phantom experiments show

consistent high-quality reconstruction across varying limited-angle conditions, outperforming FDK and registration-based methods, with high-resolution volumes reconstructed on a single 48-GB GPU within one minute, demonstrating strong potential for non-coplanar and adaptive image-guided radiation therapy.

VI. Reference

[1] D. Hu, Y. Zhang, W. Li, W. Zhang, K. Reddy, Q. Ding, X. Zhang, Y. Chen, and H. Gao, "SEA-Net: structure-enhanced attention network for limited-angle CBCT reconstruction of clinical projection data," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–13, 2023.

[2] S. Zang, Y. Zhang, D. Hu, W. Mao, X. Fei, X. Ji, Y. Yao, C. Yang, G. Coatrieux, and Y. Chen, "Coarse-to-fine learning for planning ct enhanced limited-angle CBCT reconstruction," *IEEE Trans. Instrum. Meas.*, 2024.

[3] J. Liu, R. Anirudh, J. J. Thiagarajan, S. He, K. A. Mohan, U. S. Kamilov, and H. Kim, "DOLCE: A model-based probabilistic diffusion framework for limited-angle ct reconstruction," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2023, pp. 10 498–10 508.

[4] J. Xie, H.-C. Shao, Y. Li, and Y. Zhang, "Prior frequency guided diffusion model for limited angle (LA)-CBCT reconstruction," *Phys. Med. Biol.*, 2024.

[5] L. Shen, W. Zhao, and L. Xing, "Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning," *Nat. Biomed. Eng.*, vol. 3, no. 11, pp. 880–888, 2019.

[6] B. Zhou, X. Chen, S. K. Zhou, J. S. Duncan, and C. Liu, "DuDoDR-Net: Dual-domain data consistent recurrent network for simultaneous sparse view and metal artifact reduction in computed tomography," *Med. Image Anal.*, vol. 75, p. 102289, 2022.

[7] C. Zhang, Y. Xie, and X. Liang, "XGenRecon: A New Perspective in Ultrasparse Volumetric CBCT Reconstruction Through Geometry-Controlled X-Ray Projection Generation," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 9, no. 1, pp. 95–106, 2024.

[8] V. Nagarajan, A. Andreassen, and B. Neyshabur, "Understanding the failure modes of out-of-distribution generalization," *arXiv preprint arXiv:2010.15775*, 2020.

[9] J. Henriksson, C. Berger, M. Borg, L. Tornberg, S. R. Sathyamoorthy, and C. Englund, "Performance analysis of out-of-distribution detection on trained neural networks," *Inf. Softw. Technol.*, vol. 130, p. 106409, 2021.

[10] D. Hu, Y. Zhang, J. Liu, S. Luo, and Y. Chen, "DIOR: Deep iterative optimization-based residual-learning for limited-angle CT reconstruction," *IEEE Trans, Med. Imaging*, vol. 41, no. 7, pp. 1778–1790, 2022.

[11] B. Zhou, S. K. Zhou, J. S. Duncan, and C. Liu, "Limited view tomographic reconstruction using a cascaded residual dense spatial-channel attention network with projection data fidelity layer," *IEEE Trans, Med. Imaging*, vol. 40, no. 7, pp. 1792–1804, 2021.

[12] J. Wang, L. Zeng, C. Wang, and Y. Guo, "ADMM-based deep reconstruction for limited-angle CT," *Phys. Med. Biol.*, vol. 64, no. 11, p. 115011, 2019.

[13] W. Cheng, Y. Wang, H. Li, and Y. Duan, "Learned full-sampling reconstruction from incomplete data," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 945–957, 2020.

[14] S. Barutcu, S. Aslan, A. K. Katsaggelos, and D. Gürsoy, "Limited-angle computed tomography with deep image and physics priors," *Sci. Rep.*, vol. 11, no. 1, p. 17740, 2021.

[15] D. Wu, K. Kim, and Q. Li, "Computationally efficient deep neural network for computed tomography image reconstruction," *Med. Phys.*, vol. 46, no. 11, pp. 4763–4776, 2019.

[16] M. Kellman, K. Zhang, E. Markley, J. Tamir, E. Bostan, M. Lustig, and L. Waller, "Memory-efficient learning for large-scale computational imaging," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1403–1414, 2020.

[17] B. Meng, L. Xing, B. Han, A. Koong, D. Chang, J. Cheng, and R. Li, "Cone beam CT imaging with limited angle of projections and prior knowledge for volumetric verification of non-coplanar beam radiation therapy: a proof of concept study," *Phys. Med. Biol.*, vol. 58, no. 21, p. 7777, 2013.

[18] R. Wei, Z. Song, Z. Pan, Y. Cao, Y. Song, and J. Dai, "Non-coplanar CBCT image reconstruction using a generative adversarial network for non-coplanar radiotherapy," *J. Appl. Clin. Med. Phys.*, vol. 25, no. 10, p. e14487, 2024.

[19] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," *Communications of the ACM*, vol. 65, no. 1, pp. 99–106, 2021.

[20] G. Zang, R. Idoughi, R. Li, P. Wonka, and W. Heidrich, "Intratomo: self-supervised learning-based tomography via sinogram synthesis and prediction," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 1960–1970.

[21] L. Shen, J. Pauly, and L. Xing, "NeRP: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Trans. Neural Netw. Learn. Syst.*, 2022.

[22] D. Rückert, Y. Wang, R. Li, R. Idoughi, and W. Heidrich, "Neat: Neural adaptive tomography," *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1–13, 2022.

[23] R. Zha, Y. Zhang, and H. Li, "NAF: neural attenuation fields for sparse-view CBCT reconstruction," in *Int. Conf. Med. Image Comput. Comput.-Assist. Interv.* Springer, 2022, pp. 442–452.

[24] L. Liu, L. Shen, A. Johansson, J. M. Balter, Y. Cao, L. Vitzthum, and L. Xing, "Volumetric MRI with sparse sampling for MR-guided 3D motion tracking via sparse prior-augmented implicit neural representation learning," *Med. Phys.*, vol. 51, no. 4, pp. 2526–2537, 2024.

[25] M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 7537–7547, 2020.

[26] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, 1994.

[27] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, no. 11, p. 2835, 1999.

Correction Strategies for Material Decomposition in Dual-Energy Digital Breast Tomosynthesis for Cyst-Solid Mass Differentiation

Martina Nassi, Koen Michielsen, and Ioannis Sechopoulos

***Abstract*— Accurate differentiation between cystic and solid breast lesions is critical for reducing false-positive recalls in breast cancer screening, as most cysts are benign but often appear similar to solid masses on x-ray-based images. Dual-energy digital breast tomosynthesis (DE-DBT) allows for material decomposition, but limited-angle artifacts can blur structures and mix signals between lesions and surrounding tissue.**

We propose a correction framework for DE-DBT material decomposition that mitigates these artifacts by compensating for biases related to lesion position, size, and background composition. Position-dependent errors were addressed using masking during reconstruction, background-induced signal mixing was corrected with calibration-based coefficients derived from phantom simulations, and lesion-size effects through size-dependent scaling of these coefficients.

Wide-angle DBT simulations were performed on a coronal slice of a compressed digital breast phantom containing a spherical lesion, with variations in lesion position, size (7-29 mm) and background glandularity (5-90%), and were reconstructed with material decomposition. The correction strategies, calibrated on these homogeneous-background phantoms, were further evaluated on a structured-background phantom with an 11 mm lesion insert. Masking reduced edge-related artifacts for lesions near the breast boundary, decreasing the variability of reconstructed values by 87.2% for cyst fluid and 88.7% for malignant tissue. Background correction improved accuracy, reducing the relative distance to reference values by 87.6% and 81.1%, while size-dependent scaling provided reductions of 67.9% and 75.9% for cystic and malignant lesions, respectively. Overall, the proposed correction method improves the robustness and accuracy of DE-DBT material decomposition, supporting more reliable differentiation of cystic and solid lesions.

***Index Terms*— background correction, breast tomosynthesis, dual-energy imaging, limited-angle artifacts, material decomposition, X-ray imaging.**

I. Introduction

Breast cancer screening programs are essential for reducing breast cancer mortality through early detection, but false-positive findings remain a major limitation. A substantial fraction of screening recalls, between 20% and 37% [1], are caused by cysts, which are difficult to distinguish from solid masses in x-ray imaging because both exhibit similar attenuation at typical screening energies. Since most cysts are benign, with a cancer incidence below 2% [2], improving the ability to differentiate cystic from solid lesions could reduce unnecessary recalls and their associated clinical and psychological burden.

Non-contrast dual-energy digital breast tomosynthesis (DE-DBT) offers a potential solution by combining the pseudo-three-dimensional imaging of DBT with dual-energy acquisition, which distinguishes tissue types based on their energy-dependent x-ray absorption. However, as a pseudo–three-dimensional technique, DBT suffers from limited-angle artifacts, which elongate and blur structures along the acquisition direction. These distortions reduce quantitative accuracy and can introduce errors in material decomposition, potentially leading to lesion misclassification [3].

In our previous work [4], we investigated the feasibility of distinguishing cyst fluid from malignant tissue using DE-DBT with a previously developed reconstruction algorithm incorporating material decomposition. While the algorithm showed promising baseline performance, its accuracy was limited by limited-angle artifacts. These artifacts caused signal mixing between lesions and surrounding tissue, making material decomposition sensitive to lesion size and position, and local background tissue composition. These effects were most pronounced for small lesions or those located near the edges of the field of view, where limited-angle distortions are strongest, resulting in increased overlap between cyst and malignant tissue estimates and reduced differentiation accuracy. Importantly, the observed background-related mixing showed a clear deterministic dependence on background composition and lesion size, suggesting that these effects could be addressed through simple post-processing corrections.

Building on these findings, the present study focuses on addressing these limitations. Specifically, it aims to correct for the influence of lesion size and position, as well as biases from surrounding tissues, on material decomposition.

M. Nassi, K. Michielsen, and I. Sechopoulos are with the Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands.

I. Sechopoulos is with the Dutch Expert Centre for Screening (LRCB), Nijmegen, The Netherlands.

I. Sechopoulos is with the Technical Medicine Centre, University of Twente, Enschede, The Netherlands, e-mail: ioannis.sechopoulos@radboudumc.nl.

II. Materials and Methods

A. Reconstruction with Material Decomposition

Material decomposition was performed in the reconstruction domain using a previously developed iterative algorithm that models polychromatic x-ray projections as a combination of base materials [5]. The forward model for projection line i is expressed as:

$$\hat{y}_i(\vec{w}) = \sum_e b_{ie} \exp\left(-\sum_a \mu_a^{(e)} \sum_j l_{ij} w_{aj}\right), \quad (1)$$

where b_{ie} is the energy-weighted spectrum, $\mu_a^{(e)}$ is the linear attenuation coefficient of material a at energy e, l_{ij} is the intersection length of projection line i with voxel j, and w_{aj} is the fraction of material a in voxel j. The algorithm iteratively estimates the fraction of each base material by minimizing a log-likelihood cost function [5][6].

B. Correction for Background Bias

To compensate for mixing between lesion and background signals caused by limited-angle artifacts, the reconstructed voxel values were corrected implementing a three-step approach: position-dependent correction, intensity correction, and size-dependent correction.

1) Position-Dependent Correction

To mitigate the reduced reconstruction accuracy for lesions in peripheral regions [4], a breast mask was applied during reconstruction, restricting the iterative updates to voxels within the breast. Using such masks or shape constraints has been shown to reduce edge artifacts and stabilize iterative reconstructions [7][8], providing a reliable basis for applying subsequent intensity corrections consistently.

2) Intensity Correction

Signal mixing between lesions and surrounding tissue was modeled as a linear combination of true lesion and background intensities:

$$L_m = \alpha L_t + (1 - \alpha) B_t, \quad (2)$$

where L_m is the measured (artifact-affected) lesion intensity, L_t is the true lesion intensity, B_t is the background intensity measured in a region assumed to be minimally affected by artifacts, and α quantifies the relative contribution of the lesion signal. Separate mixing coefficients, α, were obtained for each base material channel in the reconstruction. Using this model, the corrected lesion intensity for each channel was computed as:

$$L_{corr} = \frac{L_m - (1 - \alpha) B_t}{\alpha} \quad (3)$$

The values of α and the measured and true intensities were estimated from phantom simulations and literature values (see Section C.4), providing background-compensated lesion intensities.

3) Size-Dependent Correction

Even after background compensation, the effective mixing coefficient α depends on lesion size. Smaller lesions show reduced effective α due to limited-angle blurring [4]. To compensate for this, a size-scaling factor f_{size} was applied to the background-calibrated coefficient α of the reference size lesion used for background calibration:

$$\alpha_{corr} = f_{size} \cdot \alpha \quad (4)$$

This preserves background calibration while compensating for different lesion sizes. Corrected lesion intensities were then calculated from Eq. (3) using the adjusted α_{corr}.

C. Study design

This study evaluates the impact of the proposed correction strategies on material decomposition and their ability to improve differentiation between cystic and malignant tissue.

1) Phantoms

A 2D coronal slice of a compressed digital breast phantom with a thickness of 66 mm and an isotropic voxel size of 0.273 mm was used [9]. The phantom featured a homogeneous background composed of a mixture of glandular and adipose tissue, and a spherical insert simulating a round lesion, assigned either cyst fluid or malignant tissue in separate simulations. Elemental compositions of cyst fluid [10] and malignant tissue [4][11] were extracted and converted to energy-dependent attenuation coefficients $\mu_a^{(e)}$ [12].

A baseline phantom configuration was defined with a 15% background glandularity and an 18 mm-diameter lesion located at the center of the slice. Starting from this baseline, lesion position was varied as in our previous work [4] to assess position-dependent effects, while background glandularity (5–90%) and lesion size (7–29 mm) were varied to derive the calibration curves for the background and size correction strategies.

To evaluate the robustness of the correction framework, a structured-background phantom was also simulated. Starting from the same coronal slice, this phantom featured spatially varying tissue composition and a 11 mm-diameter lesion, representing a more challenging test case beyond the homogeneous phantoms and lesion sizes used for calibration.

2) Image Simulation

Phantoms were projected using a distance-driven DE-DBT simulation. The x-ray source was positioned 60.85 cm above the center of rotation (CoR), the phantom's lower edge was 3 cm below the CoR, and the source-to-detector distance was 65.55 cm. The source rotated over a limited angular range of ±25° relative to a fixed detector, producing 25 projections. Low- and high-energy tungsten spectra were set to 25 kV with

50 µm Rh filtration [13] and 49 kV with 1 mm Ti filtration [14], respectively [4]. The detector pixel pitch was 0.5 mm. Simulations excluded noise, scatter, and resolution loss.

3) Image Reconstruction

Reconstruction followed the iterative material decomposition method described in Section A, using aluminum (Al) and polymethyl methacrylate (PMMA) as base materials, with 10^3 iterations and voxel size 0.273 mm × 1.092 mm [4].

4) Correction for Background Bias

To determine the mixing coefficients α, phantom simulations were performed for a range of background glandularities. In the reconstructed volumes, voxel values within the lesion mask were used as the measured lesion signal (L_m), while background voxels adjacent to the lesion in the lateral direction were used as an estimate of the background signal (B_t), as limited-angle artifacts mainly occur along the acquisition direction. True lesion intensities (L_t) were obtained from literature [10][11]. Using these data and Eq. (2), α was computed separately for each base material, and calibration curves relating background glandularity to the mixing coefficients were generated.

These calibration curves were then applied both to the homogeneous-background phantoms used for calibration and to the structured-background phantom, allowing the estimation of α for arbitrary background composition and the computation of background-compensated lesion intensities using Eq. (3).

5) Size-Dependent Correction

To account for the effect of lesion size on the effective mixing coefficient α, phantom simulations with lesions of varying diameters were performed. Following the same approach used for intensity correction, measured voxel values were extracted from the reconstructions and combined with literature-based lesion intensities [10][11]. Using these data, together with the reference mixing coefficients α obtained for a reference lesion size of 18 mm (same used to generate the calibration curves in the previous section), a size-scaling factor f_{size} was computed for each base material. Size-dependent scaling curves were then derived to map lesion size to f_{size}.

These scaling curves were applied to both the homogeneous-background phantoms and the structured-background phantom. For an arbitrary lesion size, the appropriate f_{size} was estimated from the curves and used to compute size-compensated lesion intensities using the adjusted mixing coefficient α_{corr} from Eq. (4). This results in reconstructions that are corrected for both background composition and lesion size.

6) Analysis

The impact of the correction strategies was evaluated by comparing reconstructed Al and PMMA values ($\mu \pm 1.96\sigma$) for cystic and malignant lesions to literature reference values [10][11]. Deviations from reference values were quantified as relative distances in μ values in the Al/PMMA space. Correction performance was evaluated as the percent reduction of this distance between uncorrected and corrected values.

For background and size-dependent corrections, these deviations were assessed for both homogeneous phantoms used for calibration and structured phantoms used for validation. For position-dependent correction, the analysis focused on the lesion closest to the phantom edge, where reconstruction artifacts were most evident. For this lesion, both the percent reduction of the relative distance to reference values and the variability of reconstructed values, measured by the area of the $\pm 1.96\sigma$ box, were computed to evaluate how masking corrections improved accuracy and reduced position-dependent spread.

III. Results

Masking substantially reduced edge-dependent artifacts. For the lesion closest to the breast edge, the relative distance to reference Al and PMMA mean values decreased by 29.9% for cyst fluid and 28.6% for malignant tissue, while variability was reduced by 87.2% and 88.7%, respectively.

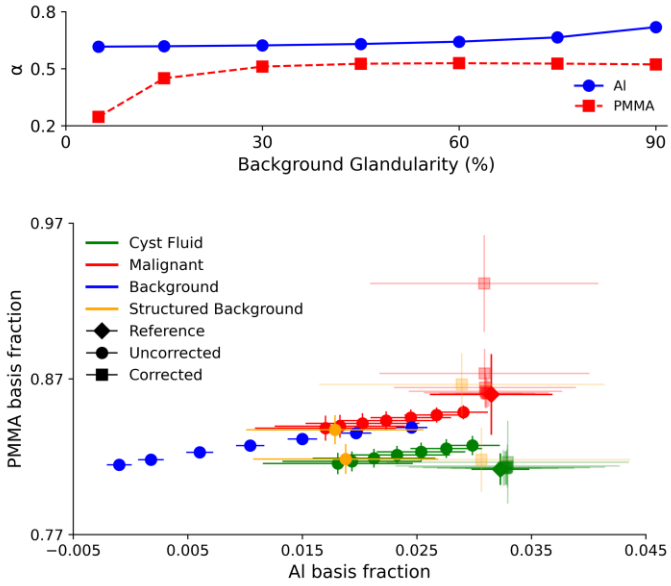


Fig. 1 Top panel: Mixing coefficients, α, as a function of background glandularity for Al (blue solid line) and PMMA (red dashed lines). Bottom panel: Material fractions of Al (x axis) and PMMA (y axis) for cyst fluid (green), malignant tissue (red), and background tissue (blue). Results are shown for varying glandularities (5 to 90%) before (circles) and after (squares) intensity correction, and are compared with literature values (diamonds). Values for the structured-background phantom are shown in orange. Markers represent mean values, with error bars indicating the 5th-95th percentile range.

Fig. 1 illustrates the background bias correction, showing calibration curves of the mixing coefficients as a function of background glandularity and the corresponding lesion material fractions before and after intensity correction. Across homogeneous-background phantoms with different glandularities, the median reduction in relative distance to reference values was 94.2% for cyst fluid and 93.8% for malignant tissue, while for the structured-background phantom the reductions were 87.6% and 81.1%, respectively.

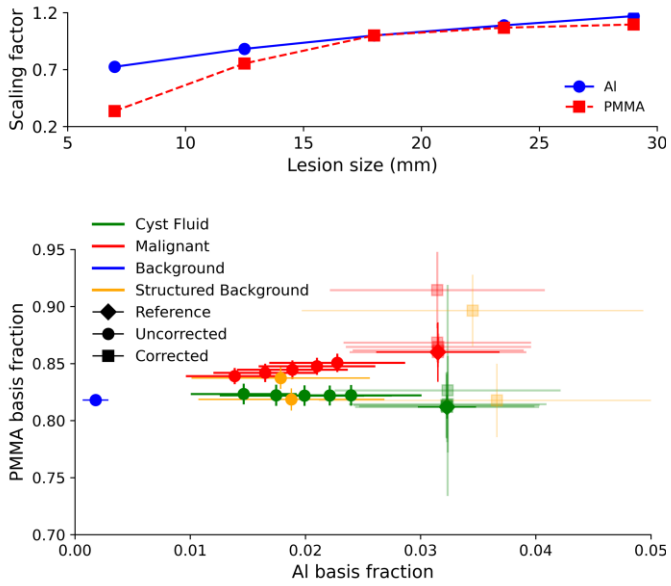


Fig. 2 Top panel: Size-scaling factor, f_{size}, as a function of lesion size for Al (blue solid line) and PMMA (red dashed line). Bottom panel: Material fractions of Al (x axis) and PMMA (y axis) for cyst fluid (green), malignant tissue (red), and background tissue (blue). Results are shown for varying lesion sizes (7 to 29 mm) before (circles) and after (squares) size correction, and are compared with literature values (diamonds). Values for the structured-background phantom are shown in orange. Markers represent mean values, with error bars indicating the 5th-95th percentile range.

Fig. 2 shows the size-dependent correction, presenting scaling curves of f_{size} as a function of lesion size along with the lesion material fractions before and after size correction. Across homogeneous-background phantoms with different lesion sizes, the median relative distance reductions were 99.3% for cyst fluid and 98.0% for malignant tissue, while for the structured-background phantom the reductions were 67.9% and 75.9%, respectively.

IV. Discussion

This study evaluates a post-processing correction aimed at reducing the influence of lesion position, size, and surrounding tissue on material decomposition in DE-DBT, with the goal of improving lesion characterization accuracy and reliability.

Masking effectively reduced position-dependent artifacts, particularly for lesions near the breast edge, where deviations from reference values were largest, bringing the reconstructed signals closer to the baseline.

Intensity correction using calibrated mixing coefficients effectively compensated for bias introduced by local background glandularity, with corrected lesion values showing improved agreement with reference values (Fig. 1). Larger errors were observed for the lowest glandularities. While the average breast glandularity is around 15%, lesions are typically embedded in glandular tissue or at the interface with glandular regions, so local glandularity near the lesion should often be higher, falling in the mid-to-high range where the correction performs more accurately.

Size-dependent correction provides accurate compensation for medium- to large-sized lesions, but smaller lesions, down to 7 mm, remain challenging to correct, highlighting a potential clinical challenge (Fig. 2).

The structured-background phantom, combining a small lesion with heterogeneous tissue composition, served as a further validation of the framework. Corrected values for both background and size generally followed the trends observed in homogeneous phantoms, though slight deviations were noticed. These deviations likely reflect the interdependence of background and size effects, which the current approach treats separately, suggesting that future improvements could include joint correction of these effects.

Despite these limitations, the results demonstrate that the proposed corrections improve the robustness and accuracy of material decomposition across a wide range of relevant lesion sizes and background compositions.

V. Conclusion

The proposed correction framework improves DE-DBT material decomposition by reducing biases associated with lesion position and size, and surrounding tissue. The corrections improve stability and agreement with reference values, supporting more reliable differentiation between cystic and malignant lesions.

Future work will focus on validating this approach using a larger set of structured-background phantoms and extending it to 3D breast phantoms and physical phantom studies to assess its robustness under realistic conditions.

Acknowledgment

This project has received funding from the Dutch Research Council (NWO) under the NWO Talent Vici AES 2021 research program (grant agreement No. 19403).

References

[1] Kooi T, Van Ginneken B, Karssemeijer N, Den Heeten A. Discriminating solitary cysts from soft tissue lesions in mammography using a pretrained deep convolutional neural network. *Medical Physics*. 2017;44(3):1017-1027. doi:10.1002/mp.12110

[2] Erhard K, Kilburn-Toppin F, Willsher P, et al. Characterization of Cystic Lesions by Spectral Mammography: Results of a Clinical Pilot

Study. *Investigative Radiology*. 2016;51(5):340-347. doi:10.1097/RLI.0000000000000246

[3] Hu YH, Zhao W. Experimental quantification of lesion detectability in contrast enhanced dual energy digital breast tomosynthesis. In *Medical Imaging 2012: Physics of Medical Imaging*. 2012;8313:112-121. SPIE. doi: 10.1117/12.911433

[4] Nassi M, Michielsen K, Sechopoulos I. Material Decomposition in Dual-Energy Breast Tomosynthesis: Cysts versus Solid Masses. In: *Proc. Fully3D 2025; Shanghai, China.* p. 415-418.

[5] Michielsen K, Rodríguez-Ruiz A, Reiser I, Nagy JG, Sechopoulos I. Iodine quantification in limited angle tomography. *Medical Physics*. 2020;47(10):4906-4916. doi:10.1002/mp.14400

[6] Bustamante VM, Nagy JG, Feng SSJ, Sechopoulos I. Iterative Breast Tomosynthesis Image Reconstruction. *SIAM J Sci Comput*. 2013;35(5):S192-S208. doi:10.1137/120881440

[7] Michielsen, K., Rangelova, T. and Sechopoulos, I. Patient evaluation of breast shape-corrected tomosynthesis reconstruction. In *Medical Imaging 2019: Physics of Medical Imaging*. 2019;10948:256-261. SPIE. doi: 10.1117/12.2513436

[8] Michielsen K, Rodriguez-Ruiz A, Sechopoulos I. Estimating the compressed breast-shape using deep learning. *In15th International Workshop on Breast Imaging (IWBI2020)*. 2020:11513:487-491. SPIE. doi: 10.1117/12.2563635

[9] García E, Fedon C, Caballo M, Martí R, Sechopoulos I, Diaz O. Realistic compressed breast phantoms for medical physics applications. In: Van Ongeval C, Marshall N, Bosmans H, eds. *15th International Workshop on Breast Imaging (IWBI2020)*. SPIE; 2020:73. doi:10.1117/12.2564273

[10] Fredenberg E, Dance DR, Willsher P, et al. Measurement of breast-tissue x-ray attenuation by spectral mammography: first results on cyst fluid. *Phys Med Biol*. 2013;58(24):8609. doi:10.1088/0031-9155/58/24/8609

[11] Fredenberg E, Kilburn-Toppin F, Willsher P, et al. Measurement of breast-tissue x-ray attenuation by spectral mammography: solid lesions. *Phys Med Biol*. 2016;61(7):2595-2612. doi:10.1088/0031-9155/61/7/2595

[12] Boone JM, Chavez AE. Comparison of x-ray cross sections for diagnostic and therapeutic medical physics. *Medical Physics*. 1996;23(12):1997-2005. doi:10.1118/1.597899

[13] Boone JM, Fewell TR, Jennings RJ. Molybdenum, rhodium, and tungsten anode spectral models using interpolating polynomials with application to mammography. *Medical Physics*. 1997;24(12):1863-1874. doi:10.1118/1.598100

[14] Boone JM, Seibert JA. An accurate method for computer-generating tungsten anode x-ray spectra from 30 to 140 kV. *Medical Physics*. 1997;24(11):1661-1670. doi:10.1118/1.597953

Memory-efficient optimization of implicit neural representations for CT reconstruction

Mahrokh Najaf, Gregory Ongie

***Abstract*—Implicit neural representations (INRs) provide a parameter-efficient and fully differentiable image model for CT reconstruction. However, optimizing INRs for CT reconstruction using standard auto-differentiation techniques can be prohibitively GPU memory-intensive, especially in 3D imaging, due to the large number of INR evaluations needed to simulate ray projections. To address this issue, we propose a memory-efficient stochastic gradient approximation based on decomposing the gradient into a Jacobian–vector product that is amenable to stochastic subsampling. This approximation allows the user to trade-off between GPU memory usage and gradient approximation accuracy. Our experiments on synthetic 2D data demonstrate that gradient approximation uses far less GPU memory than standard INR training, while yielding reconstructions that are comparable in convergence behavior and mean squared error. Finally, we demonstrate that the proposed approach allows for memory-efficient 3D cone beam CT reconstruction in a sparse-view setting.**

***Index Terms*—Implicit Neural Representations, CT Reconstruction, Memory-efficient Optimization**

I. Introduction

Implicit neural representations (INRs) provide a parameter-efficient, nonlinear, differentiable model for image reconstruction tasks in CT. In an INR, the unknown attenuation image is represented by a small-scale neural network $f_\theta : \mathbb{R}^d \to \mathbb{R}$ that takes spatial coordinates as input and outputs the corresponding grayscale value. Fitting an INR to CT measurements can be viewed as an unsupervised, model-based iterative reconstruction (MBIR) approach, where the unknown is parameterized by the network weights θ rather than by pixels/voxels. Recent work has explored INRs for low-dose and sparse-view CT reconstruction [1], [2], [3], [4], demonstrating that INRs can yield high-quality reconstructions while implicitly regularizing against noise and artifacts.

Given measured sinogram data, INR parameters are optimized to minimize a data-fidelity loss that enforces consistency with the forward CT model. While conceptually appealing, this approach introduces a significant computational challenge: training an INR typically requires differentiating through the CT forward operator. When auto-differentiation is used end-to-end, the computational graph must store intermediate states associated with both the neural network and the projection operator. As a result, peak GPU memory consumption grows rapidly with image resolution, number of views, and dimensionality. These memory constraints force practitioners to either use very small INRs with limited representational capacity, or to severely restrict the number of ray projections simulated per iteration. In 3D cone-beam CT, this limitation is particularly acute: computing and storing the gradients associated with even a single view can exceed available GPU memory on typical hardware (e.g., reconstruction of a 501^3 voxel volume on a 24 GB GPU; see Section IV-B). This motivates the need for alternative optimization strategies that reduce memory usage while preserving reconstruction quality.

In this paper, we propose a memory-efficient gradient approximation strategy for training INRs that avoids back-propagation through the CT forward model. Rather than relying on auto-differentiation to compute gradients of the loss with respect to the INR parameters, we explicitly form a Jacobian–vector product that separates the projection operator from the network derivative. This allows the use of efficient projection and backprojection routines, combined with stochastic coordinate subsampling, to approximate the gradient with substantially reduced peak GPU memory requirements.

We illustrate the proposed approach for sparse-view CT reconstruction on synthetic 2D data and real 3D data with a variety of INR architectures. Our results show that the proposed gradient approximation requires far less peak GPU memory than standard auto-differentiation, and the resulting reconstructions are comparable to auto-differentiation in terms of convergence behavior and final image mean squared error.

II. Problem formulation

Let $\boldsymbol{y} \in \mathbb{R}^m$ denote measured CT data (a sinogram in 2D or cone-beam projections in 3D), rearranged as a vector. Let $f_\theta : \mathbb{R}^d \to \mathbb{R}$ denote an INR with parameters $\theta \in \mathbb{R}^p$ that maps d-dimensional ($d = 2, 3$) spatial coordinates to attenuation values. Also, let $\mathcal{P}$ denote the forward projection operator that approximates ray integrals of a continuously defined attenuation function.

INR-based CT reconstruction is posed as the optimization problem:

$$\min_\theta \mathcal{L}(\mathcal{P}\{f_\theta\}, \boldsymbol{y}), \tag{1}$$

where $\mathcal{L} : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}$ is a given loss function. In this work, we focus on two loss functions: the least squares (LS) loss $\mathcal{L}_{\text{LS}}(\boldsymbol{z}, \boldsymbol{y}) = \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{y}\|^2$ and the filtered least squares (FLS) loss $\mathcal{L}_{\text{FLS}}(\boldsymbol{z}, \boldsymbol{y}) = \frac{1}{2}(\boldsymbol{z} - \boldsymbol{y})^\top \boldsymbol{F}(\boldsymbol{z} - \boldsymbol{y})$, where $\boldsymbol{F} \in \mathbb{R}^{m \times m}$ is a matrix that represents convolution with a ramp filter along detector columns, as used in filtered back-projection. The FLS loss can be viewed as a preconditioned least squares objective that improves conditioning of the optimization problem. This loss was investigated for INR reconstruction in our prior work [5]. While we focus on the LS and FLS loss in our experiments, we emphasize that the gradient approximation

Marquette University, Department of Mathematical and Statistical Sciences, Milwaukee, WI, USA. Corresponding author: gregory.ongie@marquette.edu

approach proposed below applies to *any* loss $\mathcal{L}(\cdot,\cdot)$ that is differentiable in its first argument.

The most direct approach to minimizing (1) is to use auto-differentiation to obtain gradients of the loss, enabling the use of first-order optimizers such as Adam. However, auto-differentiation requires retaining the computational graph (and saved intermediate outputs) for the forward projector so that gradients can be computed during the backward pass. This can consume substantial GPU memory, and the cost is amplified in 3D due to the larger volume size and the increased number of projection rays.

III. Proposed Gradient Approximation

We now describe our proposed memory-efficient approximation to gradients of (1). First, we decompose $\mathcal{P}(f_\theta) = \boldsymbol{P}\mathcal{E}\{f_\theta\}$, where $\mathcal{E}\{f_\theta\} = [f_\theta(x_1), ..., f_\theta(x_n)]^\top \in \mathbb{R}^n$ denotes the evaluation operator that forms an image/volume by sampling f_θ on a regular grid of coordinates $x_1, ..., x_n \in \mathbb{R}^d$ within a field-of-view (FOV) mask, and $\boldsymbol{P} \in \mathbb{R}^{m\times n}$ denotes a discrete forward projection operator. Let $L(\theta) = \mathcal{L}(\boldsymbol{z}, \boldsymbol{y})$ be the loss function, where $\boldsymbol{z} = \mathcal{P}(f_\theta) = \boldsymbol{P}\mathcal{E}\{f_\theta\}$ is implicitly a function of INR parameters θ, and let $\boldsymbol{r}^\top = \partial_{\boldsymbol{z}}\mathcal{L}(\boldsymbol{z}, \boldsymbol{y})$. Then by the Jacobian chain rule we have

$$\nabla L(\theta)^\top = \frac{\partial L}{\partial \boldsymbol{z}}\frac{\partial \boldsymbol{z}}{\partial \theta} = \boldsymbol{r}^\top \boldsymbol{P}\left[\frac{\partial}{\partial \theta}\mathcal{E}\{f_\theta\}\right].$$

Defining $\boldsymbol{v} = \boldsymbol{P}^\top \boldsymbol{r}$, the gradient can be written as

$$\nabla L(\theta) = \left[\frac{\partial}{\partial \theta}\mathcal{E}\{f_\theta\}\right]^\top \boldsymbol{v} = \sum_{i=1}^{n} v_i \nabla_\theta f_\theta(x_i).$$

Our key observation is that the sum structure in the gradient above enables stochastic approximation via subsampling. Let $I \subset [n]$ denote a random subset of INR evaluation indices (e.g., random coordinate batches within the FOV mask). We approximate the gradient by

$$\widehat{\nabla L}(\theta) = \frac{n}{|I|}\sum_{i\in I} v_i \nabla_\theta f_\theta(x_i). \tag{2}$$

If the index sets I are drawn uniformly at random, this is an unbiased estimator:

$$\mathbb{E}_I\left[\widehat{\nabla L}(\theta)\right] = \nabla L(\theta).$$

For implementation purposes, observe that if $\boldsymbol{v}$ is treated as constant with respect to θ, then $\widehat{\nabla L}(\theta) = \nabla_\theta \tilde{L}(\theta)$ where

$$\tilde{L}(\theta) = \frac{n}{|I|}\sum_{i\in I} v_i f_\theta(x_i) \tag{3}$$

is a "virtual" loss whose gradient matches the desired stochastic gradient estimate. Therefore, auto-differentiation can be applied to the (3) to obtain the gradient approximation $\widehat{\nabla L}(\theta)$. Algorithm 1 summarizes this approach.

Note that for the LS loss, $\partial_{\boldsymbol{z}}\mathcal{L}_{\text{LS}}(\boldsymbol{z}, \boldsymbol{y}) = (\boldsymbol{z} - \boldsymbol{y})$, and for the FLS loss $\partial_{\boldsymbol{z}}\mathcal{L}_{\text{FLS}}(\boldsymbol{z}, \boldsymbol{y}) = \boldsymbol{F}(\boldsymbol{z} - \boldsymbol{y})$. Therefore, Step 2 in Algorithm 1 reduces to backprojection of the residual $\boldsymbol{z} - \boldsymbol{y}$ for the LS loss, and a filtered backprojection of the residual for the FLS loss, which can be efficiently computed outside the computational graph.

Algorithm 1 Stochastic Gradient Estimation for INR-based CT Reconstruction

Require: INR parameters θ, measurements $\boldsymbol{y}$, batch size $|I|$

1: Evaluate $\boldsymbol{z} = \boldsymbol{P}\mathcal{E}\{f_\theta\}$ {no gradient tracking}
2: Compute $\boldsymbol{v} = \boldsymbol{P}^\top \partial_{\boldsymbol{z}}\mathcal{L}(\boldsymbol{z}, \boldsymbol{y})$ {no gradient tracking}
3: Sample random index set $I \subset [n]$, $|I| \ll n$
4: Compute virtual loss: $\tilde{L} = \frac{n}{|I|}\sum_{i\in I} v_i f_\theta(x_i)$
5: **return** $\nabla_\theta \tilde{L}$ via auto-diff. {gradient tracking on θ only}

Finally, note that the proposed gradient approximation in Algorithm 1 is also compatible with ordered subsets-like approaches where the gradient of the loss is evaluated with respect to a partial subset of views.

A. Memory and computational considerations

The memory-requirements of computing gradients of the loss in (1) using auto-differentiation scales proportionally to the total number of INR evaluations used to estimate ray projections. For example, to estimate the projections associated with one view may require millions of INR evaluations, which can lead to large peak GPU memory usage. In contrast, the gradient approximation strategy presented computes the weight vector $\boldsymbol{v}$ with optimized CT operators without gradient tracking, and backpropagates only through the INR evaluated on a coordinate minibatch. Consequently, peak GPU memory usage is dominated by the memory requirements to evaluate the INR over the minibatch, rather than the memory requirements to compute ray projections.

	FFNs	SIREN	Hash Enc.
Auto-differentiation	2.625	3.453	0.869
Proposed gradient approx.	0.708	0.767	0.534

TABLE I
Peak GPU memory usage (in GB) across different INR architectures for 2D reconstructions

Table I illustrates the peak GPU memory usage for auto-differentiation versus our proposed gradient approximation across INR architectures for our 2D reconstruction experiments (see Section IV-A for more details). Observe that the gradient approximation requires far less peak GPU memory than auto-differentiation, allowing for substantially larger problem sizes.

IV. Experiments

Here we demonstrate the proposed approach for sparse-view CT reconstruction. We present results for three popular INR architectures: fully-connected ReLU networks with an initial Fourier features encoding (FFNs) [6], fully-connected networks with sinusoidal activation functions (SIREN) [7], and a MLP with an initial multi-resolution hash encoding (Hash Enc.) [8].

We use the LivermorE AI Projector for Computed Tomography (LEAP) software package[1] for GPU-enabled forward and backward projections and FBP reconstructions [9]. All

[1] https://github.com/llnl/LEAP

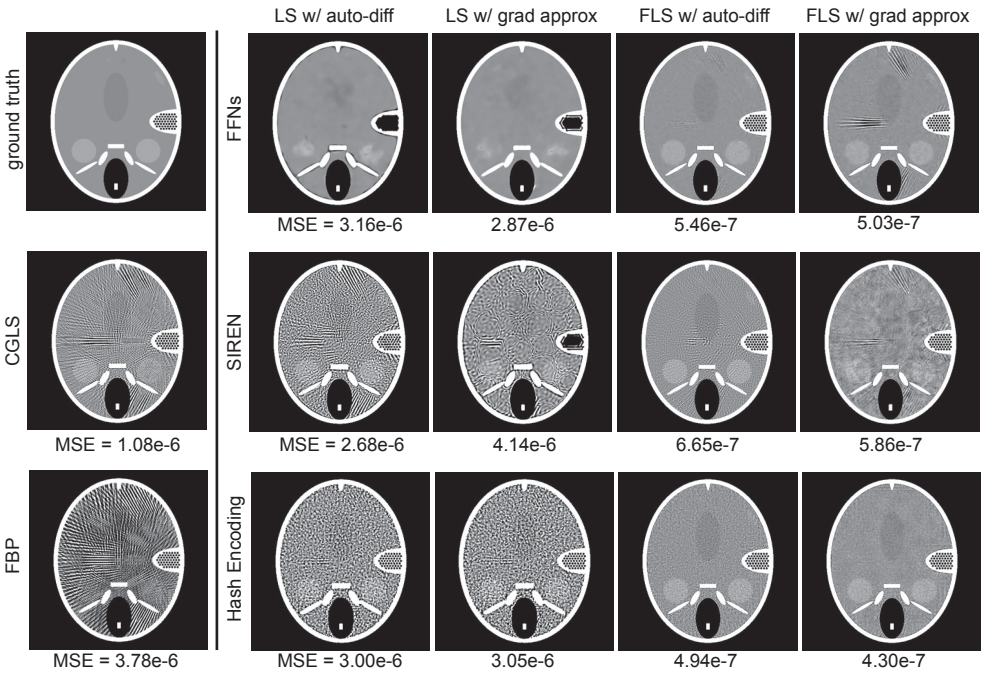


Fig. 1. 2D FORBILD sparse-view reconstructions obtained with conjugate gradient least squares (CGLS) and filtered backprojection (FBP), alongside reconstructions produced by training an INR using FFNs, SIREN, and Hash Encoding architectures. All images are shown on the scale [0.0200,0.0220] mm^{-1}.

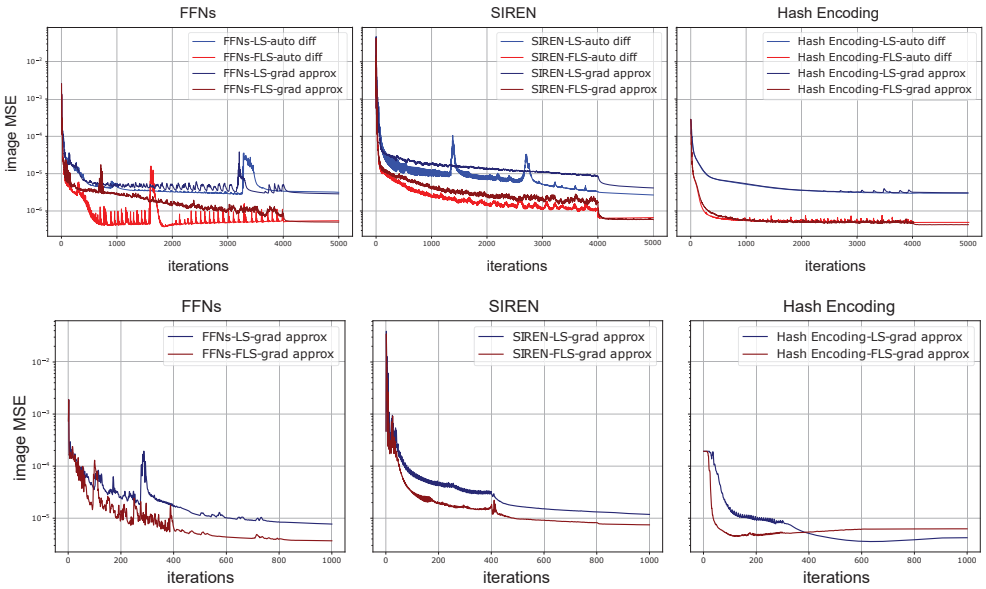


Fig. 2. Top row (2D FORBILD): Image MSE plots for INR training with FFNs, SIREN, and hash encoding, comparing auto-differentiation and gradient approximation (for both LS and FLS). Bottom row (3D Walnut): Image MSE plots for INR training with FFNs, SIREN, and hash encoding using gradient approximation (for both LS and FLS).

experiments were conducted using a single NVIDIA GeForce RTX 4090 GPU with 24 GB RAM.

As baselines, we compute reconstructions using analytical methods, i.e., filtered backprojection for 2D fan-beam and FDK for 3D cone-beam, both referred to as "FBP" in this paper. We also compare against a pixel/voxel-space least squares reconstruction obtained via the conjugate gradients algorithm (CGLS).

A. 2D FORBILD Phantom

In our first set of experiments, we synthetically generate a sparse-view 2D fanbeam acquisition of the 2D FORBILD phantom (180 equi-spaced views along a 360 degree arc, with 512 detector pixels). To avoid the inverse crime, we generate projection data from the phantom rasterized to a 2048×2048 pixel grid, and for reconstruction use projection operators acting on a 512×512 pixel grid.

We compare four training configurations for each INR architecture: (1) LS loss with auto-differentiation (LS w/auto-diff), (2) LS loss with the proposed gradient approximation (LS w/grad-approx), (3) FLS loss with auto-differentiation (FLS w/auto-diff), (4) FLS loss with the proposed gradient approximation (FLS w/grad-approx). For the gradient approximation cases, we use a minibatch size of $n/16$, where n is the number of pixel locations inside the FOV.

Figure 1 shows reconstructions obtained after training an INR for 5000 iterations with the Adam optimizer, which are comparable to those produced by CGLS or FBP, while in some cases achieving higher quality. The top panel of Figure 2 plots the pixel-wise mean-squared-error (MSE) of the INR reconstructions versus the ground truth phantom over all iterations. Note that for a given loss function and INR architecture, we see similar convergence behavior with and without the proposed gradient approximation approach.

B. 3D Walnut Dataset

We also illustrate the scalability of our approach by reconstructing real cone-beam CT data of a walnut [10]. The original dataset contains 1201 projection views with 972×768 detector pixels, which we retrospectively subsample to 120 views. The reconstructed volume size is 501^3 voxels. We use a CGLS reconstruction obtained from the full set of 1201 views as the ground truth.

For this dataset, auto-differentiation using the subsampled set of 120 views was not feasible due to GPU memory constraints. Therefore, we report results only for the proposed memory-efficient gradient approximation approach. We used a minibatch size of $n/512$, where n is the number of voxel locations inside the FOV. This resulted in the following peak GPU memory usages: 7.96 GB for FFNs, 8.43 GB for SIREN, and 6.36 GB for Hash Encoding.

Figure 3 shows reconstructions obtained after training an INR for 1000 iterations using the Adam optimizer. All three INR architectures (FFNs, SIREN, and Hash Encoding) produce reconstructions that are comparable to those from CGLS and FBP. The bottom panel of Figure 2 plots the pixel-wise MSE of the INR reconstructions versus the ground truth phantom over all iterations, which shows stable convergence of the proposed approach.

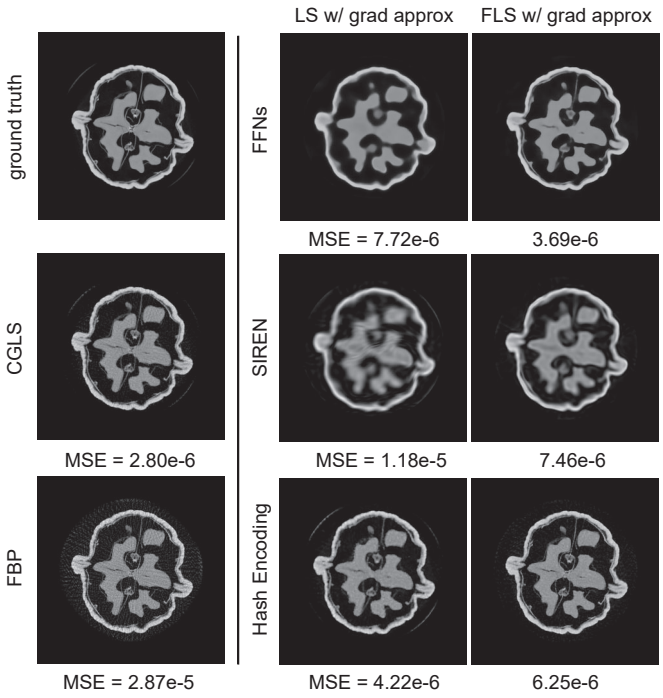


Fig. 3. 3D sparse-view cone-beam CT reconstructions obtained with conjugate gradient least squares (CGLS) and filtered backprojection (FBP), alongside reconstructions produced by training an INR using FFNs, SIREN, and Hash Encoding architectures. All images are shown on the scale [0.00,0.08] mm^{-1}.

V. Conclusion

We introduced a GPU memory-efficient gradient approximation strategy for optimizing INRs in CT reconstruction. Our experiments show that the proposed gradient approximation dramatically reduces peak GPU memory while producing reconstructions that are comparable to auto-differentiation in convergence behavior and final MSE. Finally, we also demonstrate that the proposed approach allows for easy scalability to 3D cone beam reconstruction, where the GPU memory demands of auto-differentiation through the CT projector can be prohibitive.

References

[1] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S Kamilov, "CoIL: Coordinate-based internal learning for tomographic imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1400–1412, 2021.

[2] Ruyi Zha, Yanhao Zhang, and Hongdong Li, "NAF: neural attenuation fields for sparse-view CBCT reconstruction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2022, pp. 442–452.

[3] Liyue Shen, John Pauly, and Lei Xing, "NeRP: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 1, pp. 770–782, 2022.

[4] Zhiyang Fu, Hsin Wu Tseng, and Srinivasan Vedantham, "An attenuation field network for dedicated cone beam breast CT with short scan and offset detector geometry," *Scientific Reports*, vol. 14, no. 1, pp. 319, 2024.

[5] Mahrokh Najaf and Gregory Ongie, "Accelerated optimization of implicit neural representations for CT reconstruction," in *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2025, pp. 1–5.

[6] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

[7] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7462–7473, 2020.

[8] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM transactions on graphics (TOG)*, vol. 41, no. 4, pp. 1–15, 2022.

[9] Hyojin Kim and Kyle Champley, "Differentiable forward projector for x-ray computed tomography," in *ICML 2023 Workshop on Differentiable Almost Everything: Differentiable Relaxations, Algorithms, Operators, and Simulators*.

[10] Henri Der Sarkissian, Felix Lucka, Maureen Van Eijnatten, Giulia Colacicco, Sophia Bethany Coban, and Kees Joost Batenburg, "A cone-beam X-ray computed tomography data collection designed for machine learning," *Scientific data*, vol. 6, no. 1, pp. 215, 2019.

Limited-Angle CT Reconstruction via Conditional Adversarial Diffusion with Dynamic Multi-Scale Injection

Yutong He, Dimeng Xia, Shusen Zhao

***Abstract*—Limited-angle CT (LACT) reconstruction is a severely ill-posed inverse problem. In this problem, incomplete angular sampling causes obvious streak artifacts, structural distortion and loss of clinically useful details. Existing conditional diffusion methods usually add auxiliary information in a static or shallow way. This makes conditional guidance weaker during deep denoising, so their effect is limited in limited-angle cases. To solve this problem, we propose a conditional adversarial diffusion framework with dynamic multi-scale injection for LACT reconstruction. It turns conditional guidance into a mechanism that continuously adjusts reverse diffusion dynamics. The method first learns separate structural and textural features from limited-angle reconstructed images, then combines them via cross-modal interaction, uncertainty-aware modeling and multi-scale contextual aggregation. This builds a hierarchical conditional representation, which is dynamically added into the diffusion generator at multiple U-Net scales and each denoising timestep. We also combine adversarial learning with direct image prediction. This improves high-frequency fidelity and reduces over-smoothing. This design stably suppresses artifacts and enhances structural consistency, achieving better reconstruction performance than other existing reconstruction and diffusion-based methods. (about 4%).**

I. Introduction

COMPUTED Tomography (CT) plays a crucial role in clinical diagnosis and interventional procedures due to its high spatial resolution and quantitative imaging capability. In many practical scenarios, however, the acquisition of full-angle projection data is restricted by physical constraints, dose limitations, or system geometry, leading to limited-angle CT (LACT) reconstruction. The incomplete angular coverage violates the data sufficiency condition of analytic reconstruction methods and results in characteristic streak artifacts, loss of structural integrity, and degraded contrast, particularly along directions corresponding to missing views.

Existing approaches for limited-angle CT reconstruction can be broadly categorized into three groups: image-domain methods, projection-domain completion methods, and dual-domain approaches.Image-domain methods directly process reconstructed images obtained from limited-angle data to suppress artifacts and restore structural information. Traditional image-domain techniques are mainly based on prior regularization models, such as total variation (TV) minimization [1] and sparse representation methods. Although these approaches can effectively reduce noise and streak artifacts, they often lead to over-smoothed edges and fail to preserve fine structural details. More recently, deep learning-based image-domain methods have been widely explored, including convolutional neural network (CNN)-based artifact reduction [2], [3], residual learning frameworks [4], and Transformer- [5] or generative-model-based [6]–[9] approaches. Nevertheless, pure image-domain methods typically do not explicitly model the directional information loss caused by limited-angle acquisition, resulting in insufficient recovery of image boundaries and unobservable regions.

Projection-domain completion methods aim to restore missing projections in the sinogram domain and then reconstruct images using conventional algorithms such as FBP [10] or FDK [11]. Early approaches rely on analytical extrapolation or symmetry assumptions to fill missing projections. With the advent of deep learning, data-driven models have been proposed to learn the mapping from incomplete to complete sinograms, for example, U-Net-based sinogram completion networks. However, unlike sparse-view CT, limited-angle CT suffers from a fundamental lack of physical constraints between the known and missing projection regions. The available projections provide insufficient information to reliably infer the missing data, causing the completed sinograms to heavily depend on learned priors and leading to noticeable discrepancies between reconstructed images and ground truth.

Dual-domain approaches attempt to exploit the complementary advantages of both projection and image domains. Typically, projection-domain completion is first performed, followed by image-domain refinement. Traditional dual-domain methods iteratively propagate constraints between the two domains, while deep learning-based approaches often adopt cascaded or jointly trained networks [12], [13]. Although dual-domain strategies can improve reconstruction quality to some extent, they usually involve complex architectures. Errors may accumulate during repeated domain transformations, resulting in unstable training and limited robustness under severe limited-angle conditions.

Diffusion probabilistic models have recently emerged as powerful generative frameworks capable of modeling complex data distributions through a gradual denoising process. Their inherent stochasticity and iterative refinement make them particularly attractive for inverse problems such as CT reconstruction. To further improve controllability, conditional diffusion models introduce auxiliary information to guide generation. However, existing conditional diffusion approaches

Yutong He, and Dimeng Xia are with the National Center for Applied Mathematics Shenzhen (NCAMS), Southern University of Science and Technology, Shenzhen 518055, China. Shusen Zhao is with the National Center for Applied Mathematics Shenzhen (NCAMS), Southern University of Science and Technology and Detection Institute for Advanced Technology Longhua-Shenzhen(DIATLHSZ), Shenzhen 518000, China.

typically integrate conditions via simple concatenation or shallow feature modulation, treating the condition as a static input [14]. In the context of limited-angle reconstruction, such designs are suboptimal: conditional influence tends to weaken as denoising progresses, and early-stage guidance may be overridden by accumulated noise and model bias, leading to residual artifacts or structural inconsistency.

Specifically, we introduce a dual-branch conditional encoder that disentangles structure-sensitive and texture-sensitive representations from the limited-angle reconstruction. This decoupling alleviates semantic interference between structural cues and artifact-corrupted textures. The extracted features are further fused through cross-modal interaction to enforce alignment, followed by uncertainty-aware variational modeling to account for spatially varying reliability inherent in limited-angle data. A multi-scale contextual aggregation module is then employed to construct a hierarchical conditional representation that jointly captures global structural context and local texture details.

Unlike existing methods that inject conditional information only once, the proposed framework injects the fused conditional features into the diffusion generator at multiple U-Net scales and at every denoising timestep through zero-initialized residual connections. This dynamic multi-scale injection ensures that conditional information continuously shapes the denoising trajectory, effectively preventing the degradation of conditional influence in deep diffusion steps and enabling consistent structural guidance throughout reconstruction.

II. Methodology

A. Problem Definition

We address the limited-angle CT (LACT) reconstruction problem, aiming to recover a full-angle-quality CT image from a limited-angle reconstruction.

Given a limited-angle reconstruction $x_{LA} \in R^{H\times W}$, we aim to produce $\hat{x}_{FA} \in R^{H\times W}$, approximating the full-angle target $x_{FA} \in R^{H\times W}$. The paired training dataset is defined as $\mathcal{D} = \{(x_{LA}, x_{FA})\}$. Due to incomplete projection data, limited-angle reconstruction introduces severe directional artifacts and structural distortions. Relying solely on x_{LA} is insufficient for faithful recovery; therefore, we implicitly introduce structural priors extracted from x_{FA} to guide the generation process. The real edge mask M_{mask} is not available in the inference stage, so we position the edge prior as supervision in the training stage, which is used to shape the conditional representation of structure sensitive, rather than as the inference input. The framework is shown in Fig. 1

B. Dual-Branch Interactive Variational Fusion Module

To make full use of the conditional images, we propose a mechanism-driven fusion strategy, in which structural and textural conditions are decoupled and injected into the reverse diffusion process in a multi-scale and temporally consistent manner. The goal is not merely to define an algorithmic pipeline, but to explicitly model how conditions modulate the denoising dynamics.

Specifically, we use the dual-branch feature extraction method, and the limited angle image x_{LA} is encoded by the structural branch and the texture branch, respectively, to generate multi-scale feature pyramids $\{F_s^{(l)}\}$ and $\{F_t^{(l)}\}$ respectively. This decoupling design prevents texture artifacts from polluting the structural prior, while retaining a wealth of intensity information. At each scale l, the variational uncertainty cross-modal interaction (VUCI) module explicitly aligns and exchanges information between structural and texture features to form a constraint aware representation rather than a simple feature concatenation. Furthermore, we introduce variational modeling on the interacted features to represent conditional fusion as a stochastic variable.

Crucially, the fused conditional features are repeatedly integrated into the generator at every diffusion timestep and multiple U-Net scales. At timestep t and scale $i(l)$, the generator feature is modulated as:

$$h_{i(l)}^t \leftarrow h_{i(l)}^t + \lambda_{inj} \cdot \mathcal{Z}_{i(l)} \left(F_{\text{fuse}}^{(l)} \right) \tag{1}$$

where $Z_{i(l)}$ denotes a zero-initialized convolution layer, which allows the conditional information to progressively influence the denoising trajectory. Here, the fused representation $\{F_{\text{fuse}}^{(l)}\}$ serves as the sole conditional signal for the diffusion generator, while raw inputs and privileged edge supervision no longer directly participate in the reverse process.

Through this design, conditional information becomes an intrinsic component of the diffusion dynamics rather than an external input. The model therefore learns how to denoise under conditional modulation.

1) Dual-branch Condition Encoding: We obtain structural features and textural features from different branches, respectively. The structure branch takes the limited-angle image x_{LA} as input and adopts a lightweight ResNet-18 backbone. Given input resolution $H\times W$, the ResNet stages produce multi-scale features $\{F_s^{(l)}\}_{l=0}^{L}$. The texture branch takes the reconstructed image with limited angle x_{LA} as the input, also. Then uses U-Net as the feature extractor to capture the global gray distribution and local texture details simultaneously, which is particularly important for the limited angle image containing directional artifacts. The multi-scale texture feature pyramid $\{F_t^{(l)}\}_{l=0}^{L}$ is obtained after fusion through jump connection. In order to facilitate fusion, the two features are projected into 512 dimensional channel space. Although both branches take x_{LA} as input, they are optimized under different inductive biases and supervision signals: the structure branch emphasizes boundary-aware representations guided by edge supervision, while the texture branch focuses on intensity and local texture reconstruction.

2) Variational Uncertainty Cross-Modal Interaction Module: **Bidirectional Cross-attention** At each scale l, structure feature $F_s^{(l)}$ and texture feature $F_t^{(l)}$ are fused via a cross-modal interaction module to get $F_{int}^{(l)}$, followed by uncertainty-aware fusion and multi-scale context aggregation:

Variational Uncertainty Modeling Considering the non-uniform reliability of the limited angle texture information in

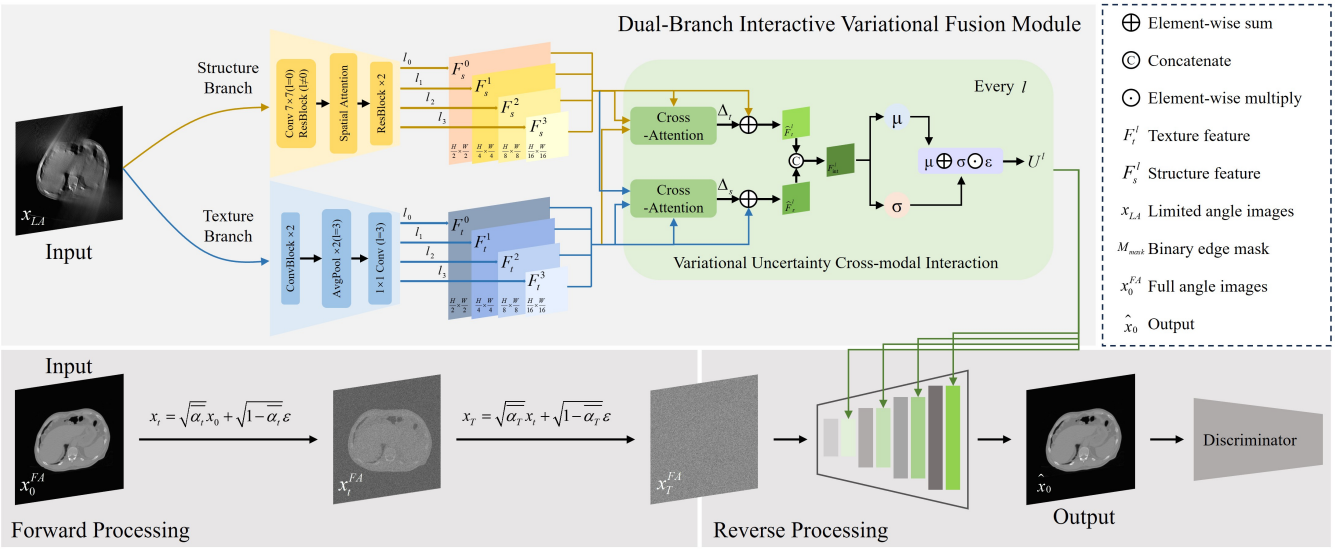


Fig. 1: An overview of the Conditional Adversarial Diffusion with Dynamic Multi-Scale Injection framework.

space, this paper introduces variational uncertainty modeling on the interactive feature $F_{int}^{(l)}$ of each scale:

$$\begin{aligned}\left(\mu^{(l)}, \log \sigma^{2(l)}\right) &= \mathrm{Conv}_{1\times1}\left(F_{int}^{(l)}\right), \\ U^{(l)} &= \mu^{(l)} + \sigma^{(l)} \odot \epsilon, \epsilon \sim \mathcal{N}(0, I).\end{aligned} \tag{2}$$

Then, build a multi-scale context branch on $U^{(l)}$:

$$G_{d=1,2,3}^{(l)} = \mathrm{Conv}\left(U^{(l)}\right), \tag{3}$$

Adaptive aggregation by pixel level weight:

$$\begin{aligned}\left[w_1^{(l)}, w_2^{(l)}, w_3^{(l)}\right] &= \mathrm{Softmax}\left(\mathrm{Conv}_{1\times1}\left(U^{(l)}\right)\right), \\ F_{fuse}^{(l)} &= \sum_{k=1}^{3} w_k^{(l)} \odot G_k^{(l)}\end{aligned} \tag{4}$$

In the training phase, KL regularization [15] is accumulated for all scales:

$$\mathcal{L}_{KL} = \sum_{l} \mathrm{KL}\left(\mathcal{N}\left(\mu^{(l)}, \sigma^{2(l)}\right) \| \mathcal{N}(0, I)\right). \tag{5}$$

Multiscale conditional injection The fused feature $F_{fuse}^{(l)}$ at each scale is injected into the corresponding level $i(l)$ of the generator U-Net through a zero-initialized residual connection.

This design preserves the original generative behavior at the early stage of training while enabling stable and progressive integration of conditional information, as described in Eq. 1.

The unified framework of decoupled double branch feature extraction and cross modal interactive fusion explicitly separates the representation learning of structural information and texture information to avoid the mutual interference of different modal features in the early fusion. On this basis, variational uncertainty modeling and multi-scale aggregation are introduced to fully characterize the differences between finite angle reconstruction and structural prior in semantic attributes, so as to achieve robust condition injection.

III. Experiments

A. Experimental Settings

1) Dataset and Pre-Processing: The experiments were conducted using the Data from 4D Lung Imaging of NSCLC Patients dataset obtained from The Cancer Imaging Archive (TCIA) [16]. The TCIA data were acquired using a 16-slice helical CT scanner (Brilliance Big Bore, Philips Medical Systems, Andover, MA) with a slice thickness of 3 mm. From this dataset, a total of 2,380 two-dimensional slices were extracted and divided into 2,000 training images, 240 validation images, and 140 test images. Incomplete limited-angle (LA) projection data over the angular range of [0°, 120°] were generated using the FDK algorithm provided by the TIGRE library.

2) Methods in comparison: Six representative CT reconstruction approaches from three categories are utilized as comparative methods: 1) two model-based methods (FDK [11], SIRT [17]); 2) four DL methods (FBPConvNet [2], DDIM [18], Fast-DDPM [19] and DiffusionMBIR [20]).

3) Evaluation Metrics: We utilize four standard metrics to quantitatively assess the performance of the compared methods: Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM).

B. Contrast Experiments

We compare the proposed method with six representative reconstruction methods on the TCIA dataset for limited-angle CT reconstruction with a scanning range of [0°, 120°]. Fig. 2 presents both qualitative and quantitative comparisons of different methods on the TCIA lung dataset under limited-angle acquisition. As expected, traditional analytical and iterative reconstruction methods exhibit limited performance in this severely ill-posed setting. FDK achieves the lowest PSNR and SSIM due to its strong reliance on complete angular coverage, which results in pronounced streak artifacts and substantial loss of structural information when projections are incomplete.

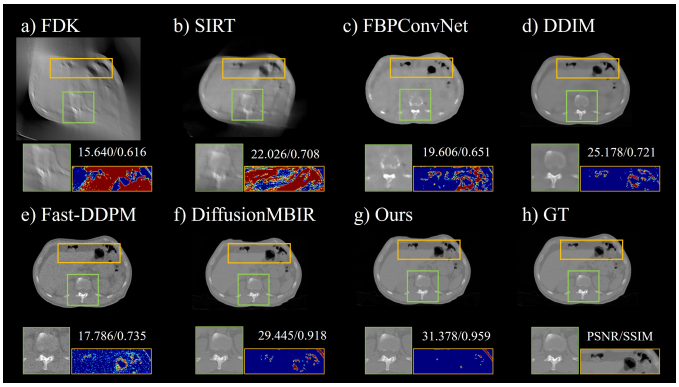


Fig. 2: The comparison results of different methods on TCIA dataset. The display window is [0,1]. The orange box shows the differences between the results obtained by different methods and GT. The values above the orange box represent the PSNR and SSIM values. The green box shows the ROI area.

SIRT improves reconstruction quality by iteratively enforcing data fidelity, yielding higher PSNR and SSIM than FDK; however, its inherent averaging behavior leads to over-smoothed reconstructions, causing blurred anatomical boundaries and loss of fine structural details in the absence of explicit prior constraints.

Learning-based image-domain and diffusion-based methods show varied performance depending on their modeling assumptions and conditioning mechanisms. FBPConvNet effectively suppresses global streak artifacts but produces overly smooth reconstructions with noticeably blurred edges, indicating limited capability in recovering sharp anatomical boundaries from severely incomplete data. DiffusionMBIR exhibits limited structural similarity, as reflected by its relatively low SSIM values, suggesting that insufficient or shallow conditioning causes structural constraints from the limited-angle input to gradually weaken during the reverse diffusion process. DDIM is able to recover the coarse anatomical layout and overall object contours; however, without conditional image guidance, its denoising process lacks explicit structural constraints, resulting in unclear and inconsistent edge delineation. Consequently, while global shapes are preserved, fine boundary structures remain ambiguous. Fast-DDPM demonstrates improved recovery of edges and fine details compared to other diffusion-based baselines; however, its reconstructions contain substantial high-frequency noise. This behavior is intrinsic to its denoising strategy rather than a result of insufficient training, highlighting the difficulty of balancing detail preservation and noise suppression under severe ill-posedness without explicit structural guidance.

The proposed method achieves the best quantitative performance across all evaluated metrics, significantly outperforming existing approaches. This superiority can be attributed to the proposed conditional adversarial diffusion framework, which explicitly models the complementary roles of structure and texture in limited-angle CT reconstruction. By employing a dual-branch feature extraction strategy, the method separately captures structure-sensitive and texture-sensitive cues from the limited-angle input, effectively preventing artifact-contaminated textures from degrading structural priors. The extracted features are fused through the VUCI module to enable cross-interaction and uncertainty-aware integration of complementary information. More importantly, the fused conditional features are progressively injected into the U-Net-based generator at multiple scales throughout the reverse diffusion process, ensuring persistent conditional guidance during denoising. This design enables superior artifact suppression while preserving anatomical fidelity, resulting in substantial improvements in both PSNR and SSIM under severe limited-angle conditions.

IV. Conclusion

This work proposes a Conditional Adversarial Diffusion framework with Dynamic Multi-Scale Injection, which transforms conditional guidance into a persistent mechanism for modulating reverse diffusion dynamics. The framework adopts a dual-branch conditional encoder to separate structure-sensitive and texture-sensitive representations from limited-angle reconstructions, reducing semantic interference between structural cues and artifact-corrupted textures. After fusing these features via cross-modal interaction for alignment, uncertainty-aware variational modeling is applied to address the spatially varying reliability of limited-angle data, and a multi-scale contextual aggregation module constructs a hierarchical conditional representation capturing both global structure and local textures. Unlike existing methods with one-time conditional injection, the proposed framework injects fused conditional features into the diffusion generator at multiple U-Net scales and each denoising timestep through zero-initialized residual connections. This dynamic multi-scale injection ensures conditional information continuously guides the denoising process, preventing the weakening of conditional influence in deep diffusion steps and maintaining consistent structural guidance throughout LACT reconstruction.

Acknowledgments

This work is in part by the grants from the National Key Research and Develop Program of China (2023YFA1011400) and the Shenzhen General Program under Grant Ref. JCYJ (20240813154703005).

References

[1] Beck, A. and Teboulle, M., "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences* **2**(1), 183–202 (2009).

[2] Jin, K. H., McCann, M. T., Froustey, E., and Unser, M., "Deep convolutional neural network for inverse problems in imaging," *IEEE transactions on image processing* **26**(9), 4509–4522 (2017).

[3] Anirudh, R., Kim, H., Thiagarajan, J. J., Mohan, K. A., Champley, K., and Bremer, T., "Lose the views: Limited angle ct reconstruction via implicit sinogram completion," in [*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*], 6343–6352 (2018).

[4] Kang, E., Chang, W., Yoo, J., and Ye, J. C., "Deep convolutional framelet denosing for low-dose ct via wavelet residual network," *IEEE transactions on medical imaging* **37**(6), 1358–1369 (2018).

[5] Wang, C., Shang, K., Zhang, H., Li, Q., and Zhou, S. K., "Dudotrans: dual-domain transformer for sparse-view ct reconstruction," in [*International workshop on machine learning for medical image reconstruction*], 84–94, Springer (2022).

[6] Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., Kalra, M. K., Zhang, Y., Sun, L., and Wang, G., "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE transactions on medical imaging* **37**(6), 1348–1357 (2018).
[7] Unal, M. O., Ertas, M., and Yildirim, I., "An unsupervised reconstruction method for low-dose ct using deep generative regularization prior," *Biomedical Signal Processing and Control* **75**, 103598 (2022).
[8] He, Z., Zhang, Y., Guan, Y., Guan, B., Niu, S., Zhang, Y., Chen, Y., and Liu, Q., "Iterative reconstruction for low-dose ct using deep gradient priors of generative model," *IEEE Transactions on Radiation and Plasma Medical Sciences* **6**(7), 741–754 (2022).
[9] Guan, B., Yang, C., Zhang, L., Niu, S., Zhang, M., Wang, Y., Wu, W., and Liu, Q., "Generative modeling in sinogram domain for sparse-view ct reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences* **8**(2), 195–207 (2023).
[10] Kak, A. C. and Slaney, M., [*Principles of computerized tomographic imaging*], SIAM (2001).
[11] Feldkamp, L. A., Davis, L. C., and Kress, J. W., "Practical cone-beam algorithm," *Journal of the Optical Society of America A* **1**(6), 612–619 (1984).
[12] Zhu, G., Zhang, Y., Mao, W., Zhu, S., Xi, Y., Quan, G., Coatrieux, G., Ji, X., Xie, S., and Chen, Y., "Multistage dual-domain networks with informative prior and self-augmentation for dual-energy limited-angle ct reconstruction," *IEEE Transactions on Instrumentation and Measurement* **73**, 1–12 (2024).
[13] Shi, Y., Lyu, T., Zhang, X., Zhong, X., Yang, Y., Zhang, P., Wu, Z., and Chen, Y., "Dpd-net: a dual-domain progressive diffusion network for joint limited-angle and metal artifact removal," in [*Optics in Health Care and Biomedical Optics XIV*], **13242**, 69–76, SPIE (2024).
[14] Dong, G., He, Y., Liu, X., Dai, J., Xie, Y., and Liang, X., "Better cone-beam ct artifact correction via spatial and channel reconstruction convolution based on unsupervised adversarial diffusion models," *Bioengineering* **12**(2), 132 (2025).
[15] Kullback, S., "Kullback-leibler divergence," *Tech. Rep.* (1951).
[16] Hugo, G. D., Weiss, E., Sleeman, W. C., Balik, S., Keall, P. J., Lu, J., and Williamson, J. F., "Data from 4d lung imaging of nsclc patients," *(No Title)* (2016).
[17] Trampert, J. and Leveque, J.-J., "Simultaneous iterative reconstruction technique: Physical interpretation based on the generalized least squares solution," *Journal of Geophysical Research: Solid Earth* **95**(B8), 12553–12559 (1990).
[18] Song, J., Meng, C., and Ermon, S., "Denoising diffusion implicit models," *arXiv:2010.02502* (October 2020).
[19] Jiang, H., Imran, M., Zhang, T., Zhou, Y., Liang, M., Gong, K., and Shao, W., "Fast-ddpm: Fast denoising diffusion probabilistic models for medical image-to-image generation," *IEEE Journal of Biomedical and Health Informatics* (2025).
[20] Chung, H., Ryu, D., McCann, M. T., Klasky, M. L., and Ye, J. C., "Solving 3d inverse problems using pre-trained 2d diffusion models," in [*Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*], 22542–22551 (2023).

Ensemble learning of different diffusion priors for sparse view CT reconstruction

Sho Ozaki, Shizuo Kaji, Toshikazu Imae, Kanabu Nawa, Hideomi Yamashita, Keiichi Nakagawa

***Abstract*—Diffusion models have recently achieved remarkable success in natural image processing, inspiring their application to medical image reconstruction. However, diffusion priors often introduce structural distortions, which are particularly detrimental in medical contexts. In this study, we propose a reconstruction framework that integrates information from multiple time steps of a diffusion model—each corresponding to a distinct noise level—and apply it to sparse-view CT reconstruction. In diffusion models, small time steps represent low noise levels and capture fine-grained details, whereas large time steps correspond to high noise levels and capture coarse structures. We observe that reconstructions from small time steps preserve anatomical fidelity but suffer from streak artifacts, while those from large time steps reduce artifacts at the cost of structural integrity. To reconcile this trade-off, we introduce an ensemble method that combines reconstructions across different noise levels, effectively preserving the advantages of each while mitigating their drawbacks. Experimental results demonstrate that the proposed ensemble method outperforms existing techniques, including iterative reconstruction (IR), IR with total variation regularization (IR+TV), diffusion posterior sampling (DPS), and stable DPS (SDPS).**

***Index Terms*—Diffusion model, ensemble learning, sparse-view CT reconstruction**

I. Introduction

With the recent advancements in diffusion models for natural image processing, several methods have been proposed for medical image reconstruction using diffusion priors [1], [2], [3], [4]. The authors of Diffusion Posterior Sampling (DPS) [5] introduced a method that utilizes diffusion priors for solving inverse problems in the image domain. They demonstrated that DPS outperforms other diffusion model-based methods in various inverse problem tasks. Although diffusion models can generate high-quality images, they also introduce diversity, leading to structural variations in the reconstructed images. Such variations are considered undesirable in medical applications. To address this issue, the authors of [6] proposed Stable DPS (SDPS) for CT reconstruction, which mitigates structural changes induced by diffusion priors and accelerates the reconstruction process.

In this study, we propose a reconstruction method that leverages contributions from different time steps in a diffusion model, corresponding to varying noise levels, and apply it to sparse-view CT reconstruction. In diffusion models, low noise levels capture fine-grained details, whereas high noise levels represent coarse-grained information. The proposed reconstruction method integrates diffusion priors at different noise levels. Using a small time step (low noise level) preserves fine anatomical structures but retains artifacts in the reconstructed image. Conversely, using a large time step (high noise level) produces cleaner images but introduces structural alterations. To effectively harness the advantages of different noise levels while mitigating their respective drawbacks such as artifacts and structural changes, we propose an ensemble method that combines contributions from multiple noise levels.

Sho Ozaki is with Graduate School of Science and Technology, Hirosaki University, 0368561 Aomori, Japan, and also with Department of Radiology, University of Tokyo Hospital, 1138655 Tokyo, Japan (e-mail:sho.ozaki@hirosaki-u.ac.jp).

Shizuo Kaji is with Center for Science Adventure and Collaborative Research Advancement, Kyoto University, 6068502 Kyoto, Japan, and also with Institute of Mathematics for Industry, Kyushu University, 8190395 Fukuoka, Japan.

Kanabu Nawa is with Kansai BNCT Medical Center, Osaka Medical and Pharmaceutical University, 5698686 Osaka, Japan, and also with Department of Radiology, University of Tokyo Hospital, 1138655 Tokyo, Japan.

Toshikazu Imae, Hideomi Yamashita, and Keiichi Nakagawa are with Department of Radiology, University of Tokyo Hospital, 1138655 Tokyo, Japan.

II. Method

The diffusion prior is derived from a pre-trained diffusion model trained on a large dataset of high-quality images, which encourages the reconstruction of clean images by reducing noise and artifacts. Diffusion models involve two processes: a forward process and a reverse process [7], each indexed by a discrete time step t. During training, noise is gradually added to the image through the forward process, while in the reverse process, noise is removed at each step t using deep neural networks. The strength of the added noise at each step is controlled by a noise schedule β_t. During inference, the trained denoising network ϵ_{θ^*} generates high-quality images.

We introduce a method to reconstruct an image from each time step of the diffusion process. We first sample a noised image x_t at time step t from the forward process, starting from an initial image reconstructed by IR: $x'_0 = \mathrm{IR}(y)$. Here, y is sparse-view projection data. The noised image x_t at an arbitrary time t can be expressed using x'_0 and Gaussian noise $\epsilon \sim \mathcal{N}(0, I)$ [7] as

$$x_t = \sqrt{\bar{\alpha}_t}\, x'_0 + \sqrt{\bar{\beta}_t}\, \epsilon, \tag{1}$$

where $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$ with $\alpha_t = 1 - \beta_t$ and $\bar{\beta}_t = 1 - \bar{\alpha}_t$. We denoise x_t using the Tweedie's formula

$$\hat{x}_0^{(t)} = \frac{1}{\sqrt{\bar{\alpha}_t}} \left(x_t - \sqrt{1 - \bar{\alpha}_t}\, \epsilon_{\theta^*}(x_t, t) \right), \tag{2}$$

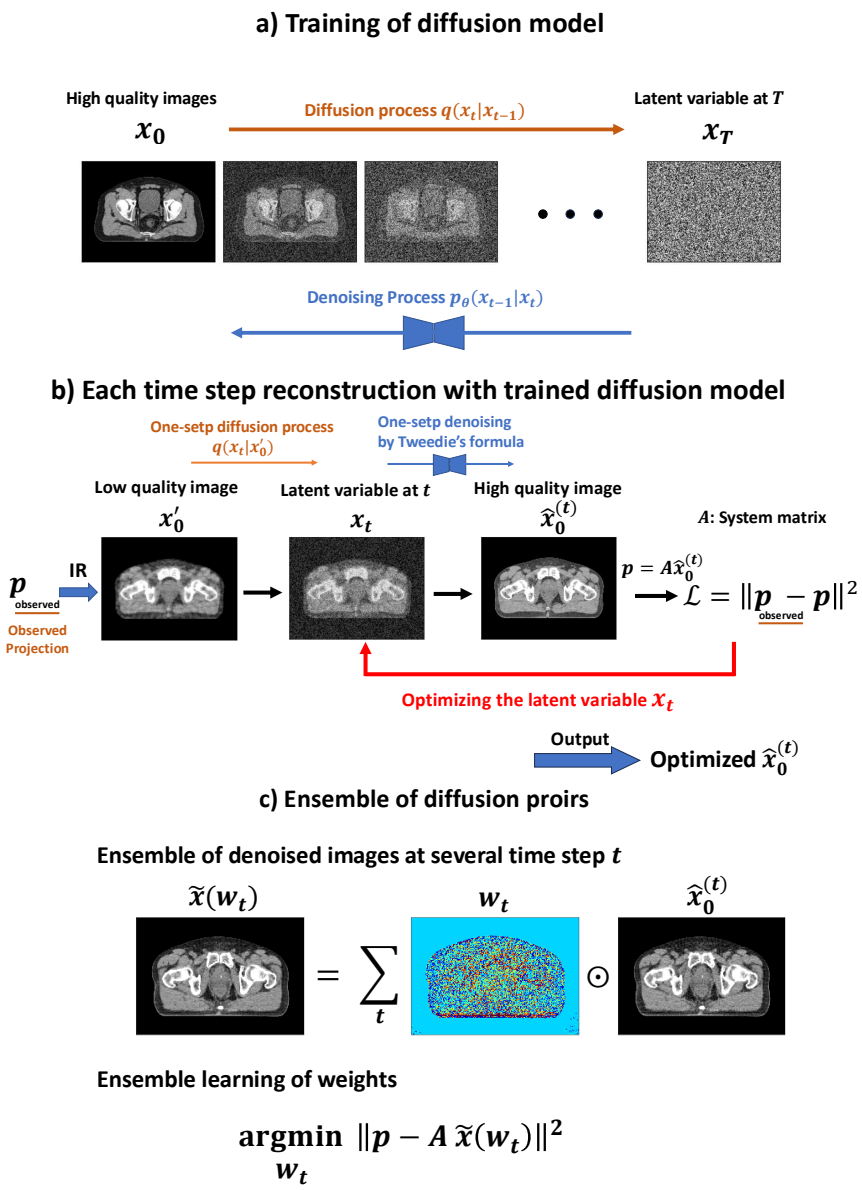


Fig. 1. Schema of the proposed ensemble method. a) Training of diffusion model. b) Each time step reconstruction with the trained diffusion model. At the beginning of the reconstruction process, we reconstructed an initial image using IR and the observed projection data. Subsequently, we conduct a one-step diffusion sampling defined by (1) using the initial image and time step t to obtain the latent variable x_t. By using the Tweedie's formula, we conduct one-step denoising of x_t and obtain a high-quality image $\hat{x}_0^{(t)}$. Reconstruction loss is computed using the output image and observed projection data, and the loss function is minimized by optimizing the latent variable x_t. The $\hat{x}_0^{(t)}$ at the end of iteration is the optimized image. c) Ensemble learning of the optimized $\hat{x}_0^{(t)}$ with several time steps t. We introduce the pixel-wise weights w_t, which have the same dimensions as the image $\hat{x}_0^{(t)}$. In the ensemble learning, the reconstruction loss is minimized by optimizing the weights w_t. The proposed ensemble method leverages the advantages of different time step diffusion priors, while suppressing their disadvantages, including artifacts and changes of anatomical structures.

with the trained denoising model $\epsilon_{\theta^*}(x_t, t)$. $\hat{x}_0^{(t)}$ is a predicted clean image at time step t. Then, we minimize the reconstruction loss $||y - A\hat{x}_0^{(t)}||^2$ by optimizing the noised image x_t:

$$x_t = x_t - \gamma \nabla_{x_t} ||y - A\hat{x}_0^{(t)}||^2, \tag{3}$$

where A is the system matrix representing the forward CT imaging process. At the end of the reconstruction iterations, we obtain the reconstructed image $\hat{x}_0^{(t)}$. In this way, we combine IR with diffusion priors. It is worth mentioning that $\hat{x}_0^{(t)}$ at a small t corresponds to a reconstructed image obtained with a small noise prior, which preserves fine-grained details. Conversely, $\hat{x}_0^{(t)}$ at a relatively large t corresponds to a reconstructed image obtained with a large noise prior, which captures coarse-grained information. Small noise prior preserves anatomical structures but remains artifacts, while large noise prior reduces artifacts but changes the structures.

To effectively leverage advantages of both priors while suppressing the disadvantages, we propose an ensemble method

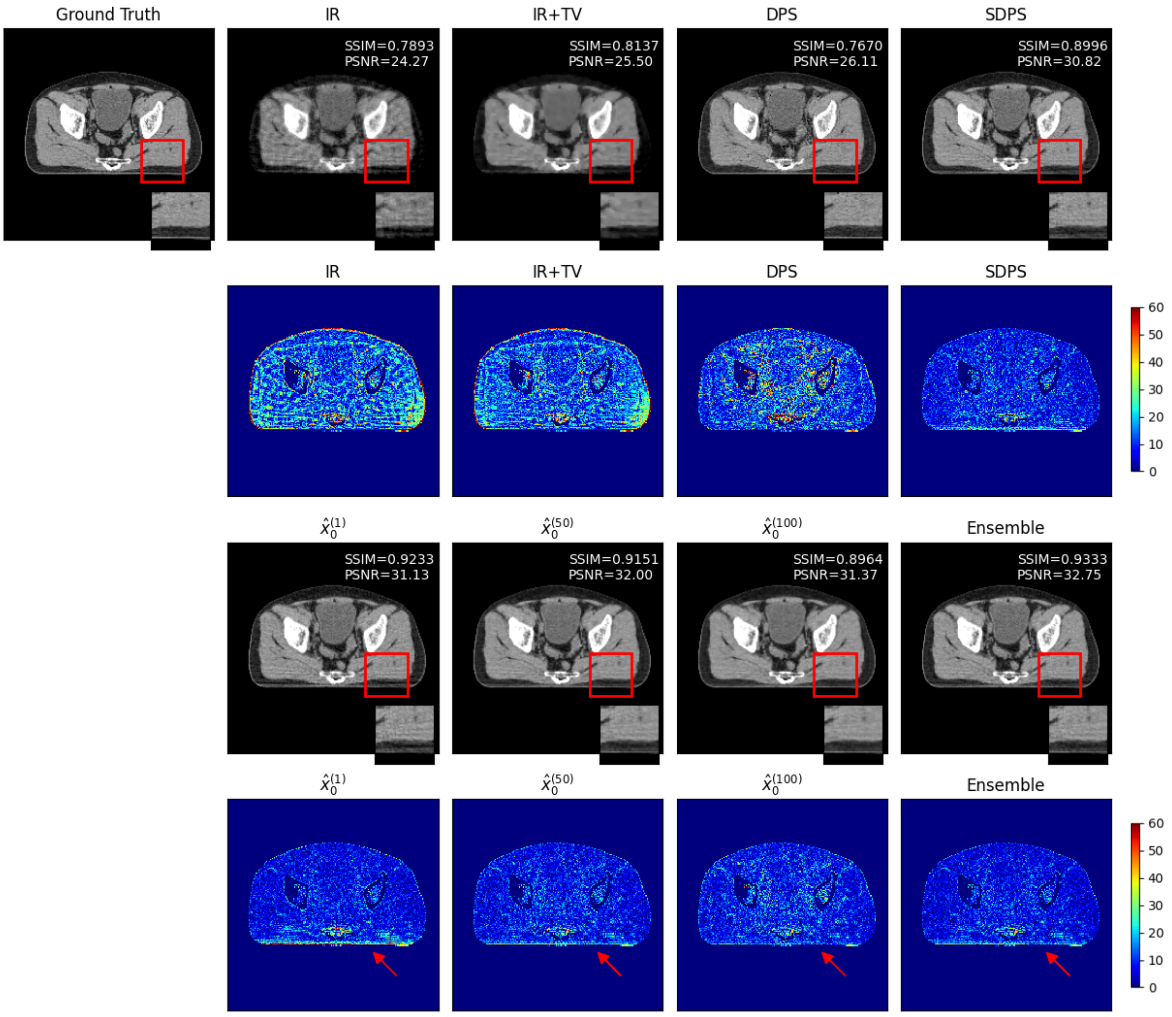


Fig. 2. Visual comparison of images reconstructed using different methods, with corresponding SSIM and PSNR values calculated against the ground truth. The images are reconstructed using 80 view projection data. A display window of [-150, 200] HU is used for all images. The color maps represent the absolute difference between each reconstructed image and the ground truth, with a display window of [0, 100] HU.

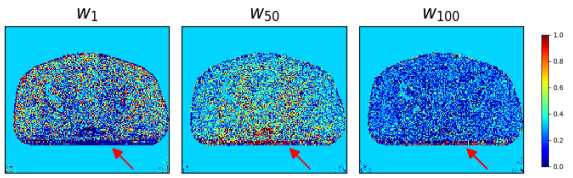


Fig. 3. Optimized pixel-wise weights w_t for $t = 1, 50, 100$. The same anatomical slice as shown in Fig. 2 is presented.

that combines multiple images $\hat{x}_0^{(t)}$ reconstructed from different time steps $\{t\}$ (different diffusion priors). Introducing the weights w_t, which have the same dimensions and size as the image $\hat{x}_0^{(t)}$ and are normalized using the softmax function: $w_t = \mathrm{softmax}(v_t)$, we obtain the ensemble image as

$$\tilde{x}_0 = \sum_{t=\{t\}} w_t \odot \hat{x}_0^{(t)}. \tag{4}$$

Here, we optimize v_t using the reconstruction loss:

$$v_{\{t\}} = v_{\{t\}} - \gamma \nabla_{v_{\{t\}}} ||y - A\tilde{x}_0||^2. \tag{5}$$

The pixel-wise weights enable local adaptation, selecting regions where each prior performs best while suppressing regions where it fails. The schema of the proposed ensemble method is presented in Figure 1.

III. Results

We evaluate the proposed methods on sparse-view CT reconstruction of the lower abdomen. The training dataset comprises 3,615 high-quality planning CT slices from 20 prostate cancer patients who underwent stereotactic radiotherapy. Images were acquired with a helical CT scanner (Aquilion LB, Canon Medical Systems, Japan) at 120 kV and 350 mA. Each image has an in-plane pixel size of 1.074 mm $\times$ 1.074 mm and a slice thickness of 1.0 mm. All images were downsampled from 512×512 at 1.074 mm spacing to 256×256 at 2.148 mm spacing and used to train the diffusion model. In this study, we employ DDPMs [7] as the diffusion prior. For testing, we selected 20 slices from 10 additional patients (two slices per patient, at least 40 slices apart) and downsampled them to 256×256 in the same manner as the

TABLE I
AVERAGE VALUES OF SSIM AND PSNR.

Indexes	IR	DPS	SDPS	Ensemble
SSIM	$0.8403_{\pm 0.0285}$	$0.8217_{\pm 0.0326}$	$0.9325_{\pm 0.0191}$	$\mathbf{0.9581}_{\pm 0.0141}$
PSNR	$25.33_{\pm 0.80}$	$26.38_{\pm 0.69}$	$31.91_{\pm 1.13}$	$\mathbf{34.16}_{\pm 1.31}$

training data. Forward projections were simulated using a fan-beam geometry with a source-to-isocenter distance of 115.0 cm and a source-to-detector distance of 177.2 cm. The arc detector comprised 528 elements with a detector spacing of 1.250 mm, and 800 samples were acquired per gantry rotation. Using the 20 test slices, we reconstructed full-scan CT with IR and sparse-view CT (80 views) with IR, IR+TV, DPS, SDPS, and the proposed methods. Quantitative image quality was assessed using the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR), where the full-scan IR reconstruction served as the ground truth (GT). All training was performed on a single NVIDIA H100 Tensor Core GPU, and reconstructions with the trained models were run on NVIDIA RTX 5000 Ada GPUs.

Figure 2 compares reconstructions at $t = 1, 50, 100$ and their ensemble to IR, IR+TV, DPS, and SDPS; the GT image is also shown. Except for GT, all images are reconstructed from 80-view projections. The image $\hat{x}_0^{(t)}$ reconstructed using the denoising model with $t = 1$ with (3) preserves the anatomical structure well. However, artifacts similar to those observed in the IR-reconstructed image remain around the prostate, rectum, and dorsal side. In contrast, the image reconstructed with $t = 100$ effectively suppresses these artifacts, but subtle structural changes can be observed. These structural changes are evident in the color maps indicating the difference from the GT image. The advantages and disadvantages of small and large time steps manifest locally, as clearly seen in the color maps in Fig. 2. The image reconstructed using the proposed ensemble method successfully reduces artifacts while preserving the patient's anatomical structure. DPS shows pronounced structural changes and consequently lower metrics; SDPS suppresses some distortions but remains inferior to the ensemble.

Figure 3 illustrates the optimized weights w_t for $t = 1, 50, 100$ in the ensemble method. We observe that the weights for $t = 1$ and $t = 50$ provide large contributions in most regions of the ensemble, while the weights for $t = 1$ exhibit very small values in the dorsal region. Conversely, the weights for $t = 100$ are relatively large in the dorsal region but small in other areas. These results indicate that the proposed ensemble method successfully learns spatially adaptive contribution ratios from different diffusion priors.

Table I presents mean values and standard deviations of SSIM and PSNR across all test images. The proposed ensemble method demonstrates the best performance against existing methods, by significantly reducing artifacts and structural changes while achieving the highest SSIM and PSNR values.

IV. DISCUSSION

In this study, we compared the proposed ensemble method with existing approaches, including IR, IR+TV, DPS, and SDPS. Our analysis revealed several key findings. DPS produces noticeable alterations of patient anatomy, yielding low SSIM and PSNR values. By contrast, SDPS better preserves anatomical structures and achieves rapid reconstruction. The proposed ensemble method surpasses SDPS in image quality (SSIM and PSNR).

The behavior across time steps t clarifies these trends. At small t, $\hat{x}_0^{(t)}$ tends to retain artifacts because the model emphasizes concrete image structures, including artifact patterns. Since the denoiser is trained on high-quality images only, Tweedies formula does not fully remove such artifacts. At larger t, strong Gaussian noise obscures fine structures (including artifacts) in x_t, so the denoised output reduces artifacts but may exhibit subtle anatomical changes. As shown in the experiments, the proposed ensemble leverages the complementary strengths of different noise levels while suppressing their drawbacks, thereby reducing artifacts and maintaining structural fidelity.

A practical consideration is computational cost, which scales with the number of time steps $\{t\}$. This cost is readily mitigated by parallelization because reconstructions for distinct t are independent. On a single RTX 5000 Ada GPU, the ensemble took 87 seconds per slice for 1,000 iterations. With three RTX 5000 Ada GPUs, assigning one time step to each GPU reduced the time to 30 seconds per slice for the same 1,000 iterationsan almost threefold speedup. We also note a limitation of the present work: reconstructed images were 256×256 due to computational constraints, whereas clinical CT typically uses 512×512 resolution.

V. CONCLUSION

In summary, the proposed ensemble framework combines diffusion priors across multiple noise levels to achieve high-fidelity CT reconstructions. By balancing fine-detail preservation and artifact suppression, it surpasses conventional and diffusion-based reconstruction methods, including IR, IR+TV, DPS, and SDPS.

REFERENCES

[1] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving Inverse Problems in Medical Imaging With Score-Based Generative Models," in *The 10th International Conference on Learning Representations*, 2022.

[2] H. Chung, B. Sim, D. Ryu, and J. C. Ye, "Improving Diffusion Models for Inverse Problems using Manifold Constraints," in *Advances in Neural Information Processing Systems*, vol. 35, pp. 25683–25696, 2022.

[3] J. Xie, H.-C. Shao, Y. Li, and Y. Zhang, "Prior frequency guided diffusion model for limited angle (LA)-CBCT reconstruction," *Phys. Med. Biol.*, vol. 69, no. 13, p. 135008, 2024.

[4] A. Lopez-Montes, T. McSkimming, A. Skeats, C. Delnooz, B. Gonzales, W. Zbijewski, and A. Sisniega, "Stationary CT Imaging of Intracranial Hemorrhage with Diffusion Posterior Sampling Reconstruction," *arXiv preprint arXiv: 2407.11196*, 2024.

[5] H. Chung, J. Kim, M. T. McCann, M. L. Klasky, and J. C. Ye, "Diffusion Posterior Sampling for General Noisy Inverse Problems," *11th Int. Conf. Learn. Represent. ICLR 2023*, pp. 1–30, 2023.

[6] X. Jiang, S. Li, P. Teng, G. Gang, and J. Webster Stayman, "Strategies for CT Reconstruction using Diffusion Posterior Sampling with a Nonlinear Model," *arXiv preprint arXiv: 2407.12956*, 2024.

[7] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

Addressing noise and resolution non-uniformities in fully 3D axial CT reconstruction

Stanislav Žabić, Kevin M. Brown

***Abstract*—We examine an artifact that arises in classical 3D axial reconstruction when axial sampling conditions are not satisfied. Although a mechanical solution, such as employing a flying spot in the axial direction, is known to mitigate this issue, it remains important to characterize and address the problem in scenarios where such mechanical solutions are not feasible. Our analysis shows that the observed non-uniformity is closely tied to the noise properties of the acquired data. Leveraging this insight, we derive an upsampling strategy that conditions the data prior to backprojection, enabling the formation of thin-slice images with substantially improved uniformity in both noise and resolution.**

I. Introduction

Modern multi-detector CT scanners can acquire and reconstruct data in helical or axial scanning modes. Each mode has certain advantages and disadvantages, but axial scanning has become especially popular for cardiac imaging as detector axial coverage has increased, with 16cm scanners now offered by several manufacturers. Yet, images reconstructed from axial scans have certain properties that are less ideal than images reconstructed from helical scans, in the sense that they can produce nonuniform noise and nonuniform resolution characteristics at different locations in the volume when reconstructed using the standard FDK-based FBP reconstruction [1], [2]. This paper aims to address both of these problems with a novel approach to data filtering.

II. Non-uniformities in axial 3D reconstuction

In the further text, we will refer to the direction perpendicular to the acquisition plane in the axial setting as the z-direction.

Imaging of coronary arteries requires high resolution in all three dimensions, including the z-direction. In systems in which the focal spot is always in the acquisition plane, FDK-based algorithms typically use linear interpolation in the z-direction through the acquired volume during the backprojection.

Using linear interpolation in backprojection yields an image thickness with a broadening factor of about 1.3, where broadening factor is ratio between the attained average image thickness vs. detector row sampling size. According to the Nyquist sampling criteria, in systems without focal spot deflection in the z-direction, only at the broadening factor of 2 can we expect images without aliasing artifacts characteristic of reconstruction without sufficient sampling conditions. Yet, there is often a desire for even thinner images (often advertised as higher z-resolution), and in this case the interpolated rows are typically generated by high-order upsampling or sharpening filtering techniques. In either case, the noise levels and resolution in the original and interpolated detector rows can be quite different. In the following sections, we illustrate and provide solutions for two distinct scenarios: 1) the case when only linear interpolation is used, which yields a broadening factor is 1.3 (at least in Helical scans), and 2) the case when an upsampling method which gives a broadening factor is 1.1 is used (again, in Helical scans). When applied to Axial reconstructions, both of these methods yield quite non-uniform noise and resolution patterns throughout the volume, as we will show, and then demonstrate two novel solutions - one which solves simply the noise nonuniformity, and one which improves both the noise and the resolution non-uniformities.

Both authors are with Philips Healthcare in Cleveland, Ohio USA

A. Noise non-uniformity

Let us look at a simplified schematic of a 3D backprojection. Images may be reconstructed at a z-position directly aligned with an original detector row, or they may be reconstructed from rays that are passing through multiple rows. For example, in the image plane located at z_1 on Fig. 1, voxels close to the isocenter will be reconstructed mostly from data that come from an originally acquired detector row. However, images at a radius of r in the same image plane will be reconstructed primarily from data that come from positions in between two original detector rows. Other image positions, such as the image located at z_2 (Fig. 1), can have opposite properties: voxels near the isocenter will be reconstructed mostly from an interpolated detector row, and voxels at some radius r will come primarily from an original detector row.

If the noise levels of the data in the original and interpolated rows are not similar, the rapid transition of image regions that come from original or interpolated rows can cause artifacts in the reconstructed images. They appear as distracting noise non-uniformity that varies rapidly as the user scrolls through the image volume. These artifacts are colloquially known as *pinwheel artifacts*. This image artifact is illustrated in Fig. 3 (a) and (b). The artifacts are denoted by the red arrows.

Algorithms [3], [4] have been proposed previously to solve somewhat similar-looking artifacts in helical data resulting from detector undersampling problems. These artifacts are sometimes referred to as *windmill* or *splay*. However, they are usually not effective in the removal of the pinwheel. Firstly, images reconstructed from helical scans do not exhibit this noise uniformity problem, because even voxels at the isocenter are reconstructed sometimes from the original rows

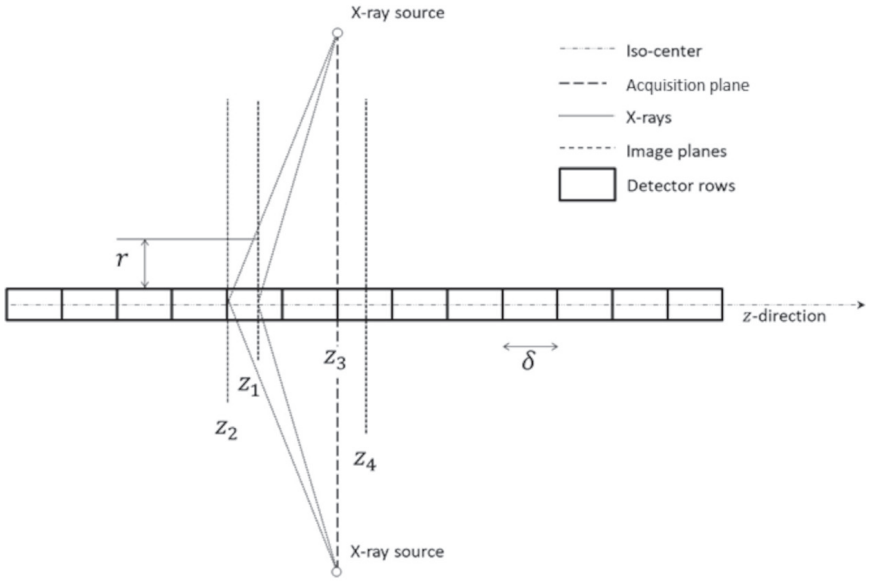


Fig. 1. Not-to-scale illustration of the backprojection geometry for a 3D axial scan and reconstruction. Iso-center, rotation axis, various image planes and detector rows are as labeled in the images. Detector row sample size is denoted by δ. Explanation for the image locations z_1, z_2 and radius r is in sub-section II-A. Explanation for the image locations z_3 and z_4 and is in sub-section II-B.

and sometimes from interpolated rows, and so the noise non-uniformity problem shown above averages out during the backprojection step. Secondly, pinwheel artifacts can appear even in objects that do not change in the z-direction, for example in a water cylinder positioned along the z-axis. This is because they do not necessarily arise from missing samples of the structures being scanned, but from non-uniform noise and properties of the axial 3D reconstruction.

B. Resolution non-uniformty

Noise non-uniformity is not the only distracting artifact in thin image volumes reconstructed with the classic 3D axial reconstruction. Non-uniformities in resolution follow similar patterns to the ones described in II-A. Let us go back to Fig. 1. Suppose that the detector increment in the z-direction is δ. Then, depending on the position of the object, its resolution in the image may vary in the z-direction. For example, an object at $z_3 = 0$ is going to be reconstructed from an average of two acquired sinograms right above, at locations $z_4 = \pm\delta/2$, and therefore the object will appear two times thicker at location $z_3 = 0$ than at the location $z_4 = \pm\delta/2$. We will show an example in Section V of how the resolution of a rapidly changing object in the z-direction can vary with a relatively high amplitude of oscillation. We will first illustrate that the problem of noise non-uniformity can be solved independently from the resolution non-uniformity in Section III, and then we will improve both of these phenomena together with a newly derived filtering technique in Section IV.

III. Synthetic noise adding

Consider acquired projections, which could be either the cone-beam projections or even those converted to parallel wedges as in [1]. Let us first upsample them in the z-direction.

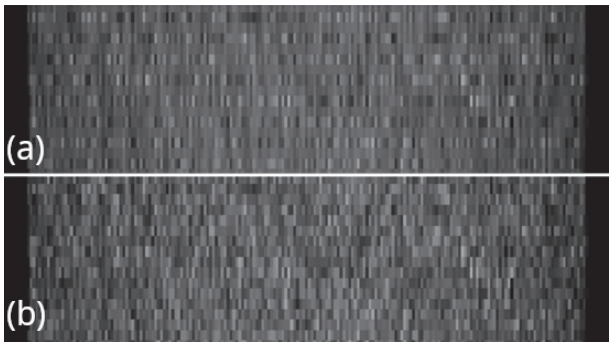


Fig. 2. Noise non-uniformity illustrated in projection domain. Image (a) shows original upsampled projection with non-uniform noise between detector rows and image (b) is the same projection after noise interleaving, showing uniform noise across detector rows.

Fig. 2 (a) show a simulated projection (with noise) of a homogeneous phantom after upsampling for a targeted slice thickness achieved with the nominal processing. The non-uniform noise pattern is clearly visible even in the projection data prior to backprojection: every other detector row is either a low-noise row or a high-noise row. Now, consider adding noise only to the upsampled rows by known methods, such as the *low dose simulation method* described in [5], with the condition that all rows in the final upsampled projection have similar noise values. Following the noise-interleaving step, the backprojection step can continue as normal, and the result is an improved image with reduced or removed pinwheel artifacts.

The impact of the proposed noise-equalizing method on projections is illustrated in Fig. 2 (b), which shows uniform noise across detector rows. Fig. 3 (c) and (d) shows images reconstructed using the noise adding method that we have just described. Reduction of the artifacts using the proposed

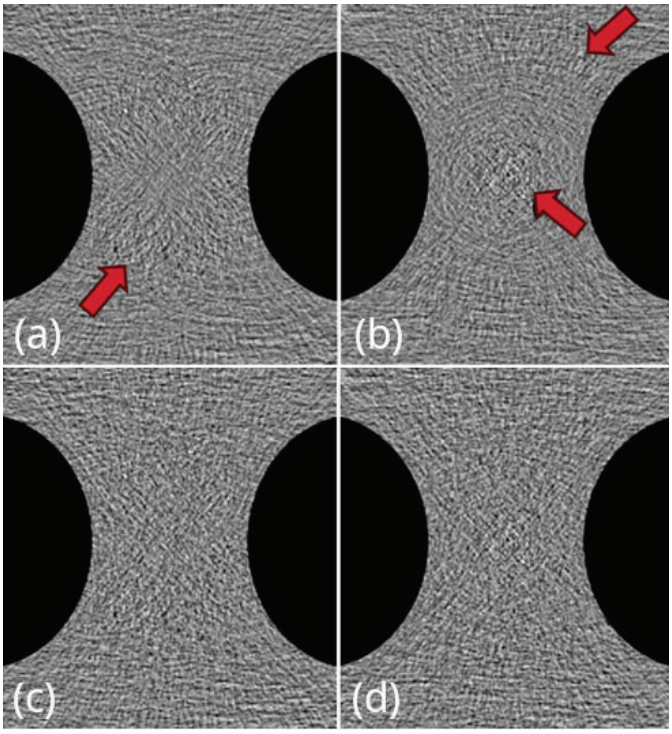


Fig. 3. Results of a noise equilizing method on axial images. Images (a) and (b) are images at two different z-positions reconstructed with the nominal processing method. Images (c) and (c) show the same images reconstructed with noise equilization. Pinwheel artifacts (red arrows) are reduced or removed in the new images (window / level = 225 / 1000 HU).

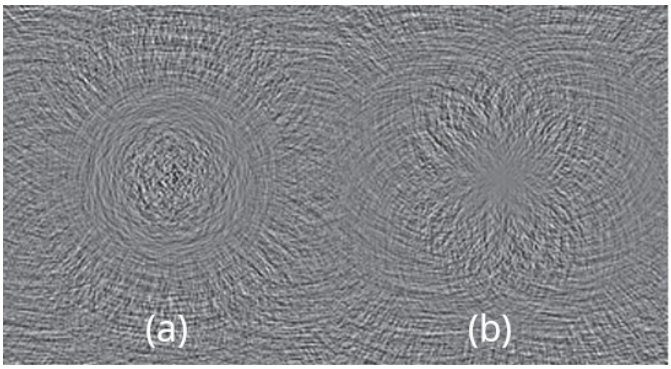


Fig. 4. Difference between nominal and noise equilizing methods. Image (a) shows the difference between images Fig. 3 (c) and Fig 3 (a). Image (b) shows the difference between images Fig. 3 (d) and Fig 3 (b) (window / level = 150 / 0 HU.

invention can be clearly seen.

Fig. 4 shows difference images between the proposed noise adding and the nominal reconstruction. Two comments are important here. First, notice that artifacts in the difference image in (a) of Fig. 4 look like the visible artifacts in Fig. 3 (b), and likewise the artifacts in b) of Fig. 4 look like the visible artifacts in Fig. 3 (a). Intuitively, adding noise provides the *missing* features that are needed to achieve a uniform noise pattern in the reconstructed images. Second, notice that the borders of the black ellipses (which resemble lungs in the phantom) are completely missing from the difference images in Fig. 4. This helps to show that the method has no impact on the resolution of structures in the images; it only addresses the noise problem.

IV. Noise equilizing high-order upsampling

Section III proves that the pinwheel itself is coming from the noise properties. But that still may not be the full solution, as the distracting flicker from the non-uniformities in resolution may remain as the user scrolls through the volume. In this section, we introduce a novel noise-equilizing upsampling that solves both uniformity issues. Let us denote this collection of detector rows as $\{r_i \mid i = 0, 1, \ldots, N-1\}$. Instead of linear upsampling, as we did in Section III, consider the following high-order upsampling that uses four neighbors instead of the usual two neighbors for all rows except for the top and bottom rows, where four neighbors are not available. For those two rows we resort to simple linear upsampling. We get an upsampled set of rows $\{\hat{r}_k \mid k = 0, 1, \ldots, N_2 - 1\}$, parameterized by α, where $N_2 = 2N - 1$. More precisely,

- when $k \in \{3, 5, 7, \ldots, N_2 - 3\}$,

$$\hat{r}_k = \frac{-\alpha \cdot r_{(k-1)/2-1} + (1+2\alpha) \cdot r_{(k-1)/2}}{2+2\alpha} + \\ + \frac{(1+2\alpha) \cdot r_{(k-1)/2+1} - \alpha \cdot r_{(k-1)/2+2}}{2+2\alpha} \tag{1}$$

- when $k = 1$ or $k = N_2 - 1$,

$$\hat{r}_k = \frac{r_{(k-1)/2} + r_{(k-1)/2+1}}{2},$$

- when $k \in \{0, 2, 4, \ldots, N_2 - 1\}$,

$$\hat{r}_k = r_{k/2}.$$

For a water cylinder, noise variance in every row r_i is equal for all rows $i = 0, 1, \ldots, N_2 - 1$. We derive the noise-equalizing parameter α by requiring that the noise variance remains equal. That is, for a fixed index $i \in \{0, 1, \ldots, N-1\}$ and all $k \in \{3, 5, 7, \ldots, N_2 - 3\}$,

$$\mathrm{var}\,(r_i) = \mathrm{var}(\hat{r}_k). \tag{2}$$

After subtituting equation (1) and applying variance properties [6] to equation (2), we get

$$1 = \frac{\alpha^2 \ + \ 2(1+2\alpha)^2 \ + \alpha^2}{(2+2\alpha)^2}. \tag{3}$$

It is trivial to see that $\alpha = 1/\sqrt{3}$ satisfies the equation 3.

V. Measurements

We find that the new method eliminates the pinwheel artifacts in a very similar way as the noise adding described in Section III and illustrated in Fig. 3 and Fig. 4, which is trivial and expected from our noise-equalizing condition. What may be surprising, though, is that in addition to the reduction of noise non-uniformity, resolution measurements also become more stable. To illustrate that, let us simulate a tilted cylinder with 1mm thickness using a slant of angle γ, so that $\tan(\gamma) = 2$. The intersection of the cylinder wall with a plane parallel and normal to the z-axis is then

$$g_{\mathrm{t}} = \frac{1}{\cos\left(\mathrm{atan}(1/2)\right)} \approx 1.118\mathrm{mm}.$$

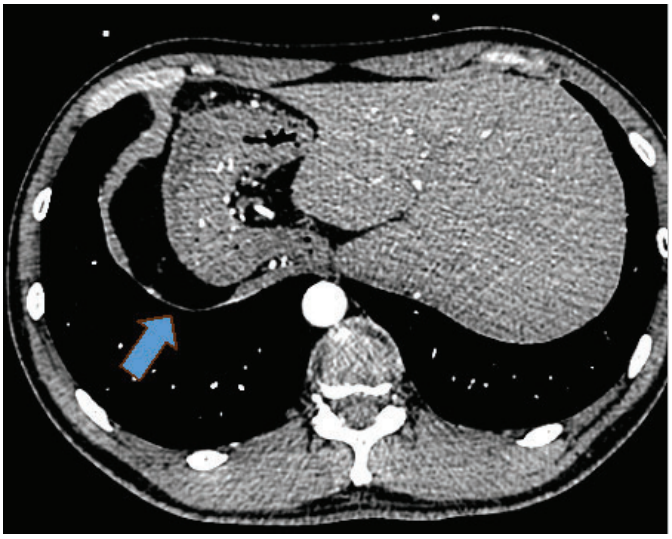


Fig. 5. Newly proposed noise-equilizing high-order upsampling method has more uniform noise and blue arrow is pointing at a more accurate representation of the heart boundary.



Fig. 6. Red arrow is pointing at pinwheel artifacts in nominal reconstruction.

Fig. 7 shows the measured thickness of the cylinder wall at different z-positions within a reconstructed volume from an axial scan. In the nominal reconstruction, we can see that the measured width of the cylinder wall varies with mean full-width-half-max 1.2696mm and standard deviation 0.0794mm. In the newly proposed reconstruction with the same row filter, but with the four-point upsampling, the measured width of the cylinder wall varies with mean full-width-half-max 1.1884mm and standard deviation 0.0267mm. Raw measurements are presented in the graph, illustrating that the newly proposed method gives a measurement closer to the calculated ground truth $g_{\mathtt{t}} = 1.118\mathtt{mm}$ and with significantly lower thicnkness variability through the volume.

Similar improvements in the reduction of pinwheel noise and better z-resolution properties are also observed in a clinical scan (compare Fig. 6 to Fig. 5).

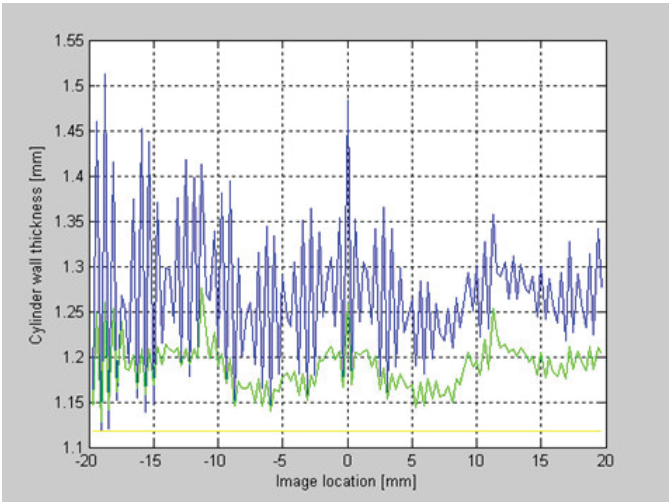


Fig. 7. Measured slanted cylinder width with the new high-order upsampling method (green curve) is more uniform than the nominal method (highly oscilating blue line) and closer to the ground truth value (yellow line).

VI. Conclusion

Using a noise-adding methodology, we demonstrated that an undersampling artifact in 3D axial reconstruction, referred to as the pinwheel, is fundamentally a manifestation of noise rather than a consequence of imagining small structures with rapid variation along the z-direction. Building on this insight, we introduced a noise-equalizing high-order upsampling approach designed to mitigate both noise and resolution non-uniformities under low-sampling conditions, particularly when thin slices are required in 3D axial CT reconstruction. Although the inherent violation of Nyquist sampling criteria for the given imaging task precludes complete elimination of all undersampling artifacts, the proposed method offers a clear and measurable improvement over the conventional strategy of applying a high-pass filter with linear interpolation during backprojection.

References

[1] M. Grass, T. Köhler, R. Proksa, *3D cone-beam CT reconstruction for circular trajectories*, Phys. Med. Biol. **45** 329, 2000

[2] L. A. Feldkamp, L. C. Davis, J. W. Krass, *Practical cone-beam algorithm*, J.Opt.Soc.Am.A/Vol.1, No.6/June 1984.

[3] J. Sunnegårdh, M. Grasruck, *Nonlinear regularization of iterative weighted filtered backprojection for helical cone-beam CT*, IEEE NSSC Record, 5090-5095, 2008

[4] K. Brown, S. Žabić, *Method for reducing windmill artifacts in multislice CT images*, Proc. SPIE 7961, Medical Imaging: Physics of Medical Imaging, 79611P, 2011

[5] S. Žabić, Q. Wang, T. Morton, K. Brown, *A low dose simulation tool for CT systems with energy integrating detectors*, Med Phys. 2013 Mar;40(3):031102. doi: 10.1118/1.4789628. PMID: 23464282, 2013

[6] G. Casella, R. Berger, *Statistical Inference*, Duxbury Resource Center, Pacific Grove, CA, 2001

MeanFlow Posterior Sampling for Low-Dose Head CT Reconstruction

Dufan Wu, Yuang Wang, Matthew Tivnan, Quanzheng Li

***Abstract*—Low-dose head computed tomography (CT) is used in portable systems due to tube current constraints and pediatrics due to dose concerns. Diffusion posterior sampling (DPS) is an advanced reconstruction algorithm that combines unsupervised diffusion models and iterative reconstruction and is promising in reducing noise in low-dose head CT. However, DPS suffers from a long inference time because the underlying diffusion model needs hundreds of sampling steps. In this work, we proposed a novel MeanFlow Posterior Sampling (MPS) method for CT reconstruction, where data fidelity was built into the sampling process of MeanFlow, a recently proposed one-step flow matching generative model. Given a pretrained unconditional MeanFlow model, we established the relationship between the flow and score, then built the data fidelity into the score. To validate the proposed method, an unconditional MeanFlow model was trained from photon-counting CT (PCCT) images for head, and applied to simulated low-dose PCCT, energy integrated detector (EID) CT, and portable CT for head. With only 5 sampling steps, the proposed method achieved superior image quality compared to DPS on the PCCT and EID-CT datasets.**

***Index Terms*—Computed tomography, deep learning, diffusion model, flow matching, acceleration**

I. Introduction

COMPUTED tomography (CT) imaging for head often requires high radiation dose to separate low contrast anatomy such as gray and white matter, hemorrhage, infarction, etc. Low-dose head CT protocols are used on portable head CT scanners due to limited X-ray tube power or pediatric imaging to reduce dose, and leads to increased noise.

Recently, diffusion model-based deep learning reconstruction has been of great research interest in low-dose CT thanks to its capability in generating high-quality images and preserving texture. Diffusion Posterior Sampling (DPS) [1] further combines a pretrained unconditional diffusion model with data fidelity. Thus it can be used for image reconstruction under various data conditions using the same pretrained model. However, DPS often suffers from long inference time because the underlying diffusion model usually needs hundreds of sampling steps to acquire high quality samples.

This work was supported in part by the National Institute Of Biomedical Imaging And Bioengineering of the National Institutes of Health under Award Number R01EB035394. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

D. Wu is with the Department of Radiology, The Ohio State University, Columbus, OH 43221 USA. E-mail: dufan.wu@osumc.edu.

Y. Wang was with Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston MA 02114 USA. He is currently with the Department of Engineering Physics, Tsinghua University, Beijing 100084 China.

M. Tivnan and Q. Li are with the Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston MA 02114 USA.

Flow matching [2] is another generative model that estimates the sample's flow from Gaussian to the desired distribution. It predicts the sample's velocity so small step size is still needed during inference to reduce discretization error. To tackle this challenge, the MeanFlow model was recently proposed to learn the integral velocity instead of the instantaneous velocity [3]. It demonstrated high-quality sampling with very few sampling steps as low as one.

In this work, we proposed a novel MeanFlow Posterior Sampling (MPS) method for CT image reconstruction that combines MeanFlow with data fidelity. Specifically, a novel theory was derived to incorporate data fidelity and unconditional MeanFlow model through the relationship between the mean velocity and score. Only five sampling steps were needed for high quality image reconstruction. The proposed model was validated and compared to DPS on three datasets: simulated low-dose photon counting CT (PCCT), real energy integrated detector (EID) CT, and real portable PCCT.

II. Preliminary

A. Diffusion Posterior Sampling

Given measurement $\mathbf{y}$, the reverse sampling of a score-based diffusion model can be written as

$$d\mathbf{x} = [f(\mathbf{x}_t, t) - g(t)^2 \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}|\mathbf{y})]dt + g(t)d\mathbf{w}, \quad (1)$$

where $\mathbf{x}_t$ is the sample at time t, $f(\mathbf{x}, t)$ and $g(t)$ are predefined functions, and $\mathbf{w}$ is a standard Wiener process.

Using Bayesian rules,

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y}) = \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{x}_t), \quad (2)$$

where $\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) = s_\theta(\mathbf{x}_t, t)$ is a pretrained unconditional score estimator, and

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{y}|\mathbf{x}_t) \approx \nabla_{\mathbf{x}_t} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0) \approx \alpha \nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0), \quad (3)$$

where $\hat{\mathbf{x}}_0 = \mathbb{E}(\mathbf{x}_0|\mathbf{x}_t)$ and can be calculated using $s_\theta(\mathbf{x}_t, t)$. α is a scaler to approximate the Jacobian of $\mathbb{E}(\mathbf{x}_0|\mathbf{x}_t)$ [4]. Substitute (2) and (3) into (1), and the DPS process is:

$$\begin{aligned} d\mathbf{x} \approx &[f(\mathbf{x}_t, t) - g(t)^2 s_\theta(\mathbf{x}_t, t)]dt + g(t)d\mathbf{w} \\ &- g(t)^2 \alpha \nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0)dt, \end{aligned} \quad (4)$$

which consists of disentangled unconditional diffusion and data fidelity terms.

B. Flow Matching

Given images $\mathbf{x}_0 \sim p_0(\mathbf{x})$ and $\epsilon \sim p_1(\epsilon)$ where $p_1(\epsilon) = N(\epsilon|\mathbf{0}, \mathbf{I})$ is Gaussian, a conditional path is defined as

$$\mathbf{x}_t = (1-t)\mathbf{x}_0 + t\epsilon, 0 \le t \le 1 \tag{5}$$

The conditional velocity $\mathbf{v}(\mathbf{x}_t|\mathbf{x}_0, t) = \epsilon - \mathbf{x}_0$ and the marginal velocity $\mathbf{v}(\mathbf{x}_t, t) = \mathbb{E}_{p(\mathbf{x}_0|\mathbf{x}_t)}\mathbf{v}(\mathbf{x}_t|\mathbf{x}_0, t)$. $\mathbf{v}(\mathbf{x}_t, t)$ is represented by model $\mathbf{v}_\theta(\mathbf{x}_t, t)$ that is trained by:

$$\theta^* = \arg\min_\theta \mathbb{E}_{\mathbf{x}_0,\epsilon}\|\mathbf{v}_\theta(\mathbf{x}_t, t) - (\epsilon - \mathbf{x}_0)\|_2^2 \tag{6}$$

Starting from $\mathbf{x} \sim N(\mathbf{x}|\mathbf{0}, \mathbf{I})$, images $\mathbf{x}_0 \sim p_0(\mathbf{x})$ can be generated by the following process from $t = 1$ to $t = 0$:

$$d\mathbf{x} = \mathbf{v}_{\theta^*}(\mathbf{x}_t, t)dt \tag{7}$$

C. MeanFlow

To eliminate the discretization error in (7), MeanFlow learns the mean of $\mathbf{v}(\mathbf{x}_t, t)$ between time t and r:

$$u(\mathbf{x}_t, r, t) = \frac{1}{t-r}\int_r^t \mathbf{v}(\mathbf{x}_\tau, \tau)d\tau, r < t \tag{8}$$

The model $u_\theta(\mathbf{x}_t, r, t)$ can be trained as:

$$\begin{aligned} \theta^* &= \arg\min_\theta \mathbb{E}\|u_\theta(\mathbf{x}_t, r, t) - \mathrm{sg}(u_{tgt})\|_2^2 \\ u_{tgt} &= \mathbf{v}(\mathbf{x}_t, t) - (t-r)(\mathbf{v}(\mathbf{x}_t, t)\partial_{\mathbf{x}_t}u_\theta + \partial_t u_\theta), \end{aligned} \tag{9}$$

where $\mathrm{sg}(\cdot)$ is the stop gradient operation to avoid high order gradient during the training. The first $\mathbf{v}(\mathbf{x}_t, t)$ can be replaced with the conditional velocity $\epsilon - \mathbf{x}_0$ but the second $\mathbf{v}(\mathbf{x}_t, t)$ should remain marginal. In practice, one can either use $\epsilon - \mathbf{x}_0$ or $u_\theta(\mathbf{x}_t, t, t)$ for the second $\mathbf{v}$ and the latter choice leads to faster convergence and slightly better performance [5].

After u_θ is trained, the sampling is:

$$\mathbf{x}_r = \mathbf{x}_t + (r-t)u_\theta(\mathbf{x}_t, r, t), r < t \tag{10}$$

Note that we are sampling from $t = 1$ to $t = 0$, which is consistent with the diffusion models.

III. Methodology

A. Relationship between Flow and Score

To incorporate data fidelity into the flow matching, we need to build connection between the marginal velocity $\mathbf{v}(\mathbf{x}_t, t)$ and the score $\nabla_{\mathbf{x}} \log p(\mathbf{x})$ so that we can replace $\nabla_{\mathbf{x}} \log p(\mathbf{x})$ with $\nabla_{\mathbf{x}} \log p(\mathbf{x}|\mathbf{y})$ and apply the Bayesian rules (2).

Given the conditional path (5),

$$p_t(\mathbf{x}_t|\mathbf{x}_0) = N(\mathbf{x}_t|(1-t)\mathbf{x}_0, t^2) \tag{11}$$

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0) = -\frac{\mathbf{x}_t - (1-t)\mathbf{x}_0}{t^2} \tag{12}$$

Meanwhile, we have $\epsilon = (\mathbf{x}_t - (1-t)\mathbf{x}_0)/t$ from (5) so

$$\mathbf{v}(\mathbf{x}_t|\mathbf{x}_0, t) = \epsilon - \mathbf{x}_0 = \frac{\mathbf{x}_t - \mathbf{x}_0}{t} \tag{13}$$

Combine (12) and (13) to cancel out $\mathbf{x}_0$:

$$\mathbf{v}(\mathbf{x}_t|\mathbf{x}_0, t) = -\frac{1}{1-t}\mathbf{x}_t - \frac{t}{1-t}\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{x}_0) \tag{14}$$

By taking expectation $\mathbb{E}_{p(\mathbf{x}_0|\mathbf{x}_t)}$ on both sides of (14):

$$\mathbf{v}(\mathbf{x}_t, t) = -\frac{1}{1-t}\mathbf{x}_t - \frac{t}{1-t}\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) \tag{15}$$

B. MeanFlow Posterior Sampling

Given measurement $\mathbf{y}$, the velocity becomes:

$$\mathbf{v}(\mathbf{x}_t|\mathbf{y}, t) = -\frac{1}{1-t}\mathbf{x}_t - \frac{t}{1-t}\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y}) \tag{16}$$

From (2) and (3),

$$\nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t|\mathbf{y}) \approx \nabla_{\mathbf{x}_t} \log p_t(\mathbf{x}_t) + \alpha\nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0) \tag{17}$$

Hence,

$$\mathbf{v}(\mathbf{x}_t|\mathbf{y}, t) \approx \mathbf{v}(\mathbf{x}_t, t) - \frac{t}{1-t}\alpha\nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0) \tag{18}$$

Substitute (18) into (8) and we have u given $\mathbf{y}$:

$$\begin{aligned} u(\mathbf{x}_t|\mathbf{y}, r, t) \approx& \frac{1}{t-r}\int_r^t \mathbf{v}(\mathbf{x}_\tau, \tau)d\tau \\ &- \frac{1}{t-r}\int_r^t \frac{\tau}{1-\tau}\alpha\nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0)d\tau \\ \approx& u(\mathbf{x}_t, r, t) - c(r, t)\alpha\nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0), \end{aligned} \tag{19}$$

where

$$c(r, t) = \frac{1}{t-r}\int_r^t \frac{\tau}{1-\tau}d\tau = \frac{1}{t-r}\ln\frac{1-r}{1-t} - 1 \tag{20}$$

Several approximations are used in the last step of (19). The first term does not strictly equal to $u(\mathbf{x}_t, r, t)$ because the path of $\mathbf{x}_\tau$ given $\mathbf{y}$ is different from the unconditional path and will lead to different integral results. In the second term, $\hat{\mathbf{x}}_0 = \hat{\mathbf{x}}_0(\mathbf{x}_\tau)$ but is approximated by $\hat{\mathbf{x}}_0(\mathbf{x}_t)$ so that the data fidelity can move out of the integral. $\hat{\mathbf{x}}_0$ is estimated by:

$$\hat{\mathbf{x}}_0 = \mathbf{x}_t - tu_\theta(\mathbf{x}_t, 0, t) \tag{21}$$

And the MPS step from t to r ($r < t$) is:

$$\begin{aligned} \mathbf{x}_r =& \mathbf{x}_t + (r-t)u_\theta(\mathbf{x}_t, r, t) \\ &+ c(r, t)(t-r)\alpha\nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0) \end{aligned} \tag{22}$$

C. Algorithm

Several considerations are needed to convert (21) and (22) into practical MPS algorithms.

1) Jump Start: Instead of starting from $t = 1$, one can start from $\mathbf{x}_s = (1-s)\mathbf{x}' + s\epsilon$, where $\mathbf{x}'$ is the filtered backprojection (FBP) reconstruction. With large enough s, $\mathbf{x}_s \approx (1-s)\mathbf{x}_0 + s\epsilon$, serving as a good starting point.

2) Data Fidelity: The gradient $\nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0)$ has slow convergence for image reconstruction and will not be enough to impose data fidelity during limited sampling steps. Following [4], we replace $(t-r)\alpha\nabla_{\hat{\mathbf{x}}_0} \log p_0(\mathbf{y}|\hat{\mathbf{x}}_0)$, the gradient times step size, with a few steps of iterative reconstruction for the following cost function starting from $\mathbf{x} = \hat{\mathbf{x}}_0$:

$$L(\mathbf{x}) = \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 + \beta\sum_{j\in N_i}(x_j - x_i)^2, \tag{23}$$

where $\mathbf{A}$ is the system matrix, β is the hyperparameter for Gaussian penalty, and N_i is the neighborhood of pixel i.

The MPS is summarized in algorithm 1. $\mathbf{y}$ is the measurement, $\mathbf{x}'$ is the FBP image, N is the number of sampling steps, K is the number of steps for iterative reconstruction, S is the number of ordered subsets, β is the penalty strength during

Algorithm 1 MeanFlow Posterior Sampling

Require: $\mathbf{y}, \mathbf{x}', N, K, S, \beta, \{t\}_{i=1}^{N}, t_0 = 0$
1: $\mathbf{x}_t \leftarrow (1 - t_N)\mathbf{x}' + t_N\epsilon, \epsilon \sim N(\mathbf{0}, \mathbf{I})$
2: for $i = N$ to 1 do
3: $t \leftarrow t_i, r \leftarrow t_{i-1}$;
4: $\hat{\mathbf{x}}_0 \leftarrow \mathbf{x}_t - tu_\theta(\mathbf{x}_t, 0, t)$;
5: $\hat{\mathbf{x}}_0' \leftarrow \text{OSSQS}(\hat{\mathbf{x}}_0; \mathbf{A}, \mathbf{y}, \beta, K, S)$;
6: if using MPS:
7: $\mathbf{x}_t \leftarrow \mathbf{x}_t + (r - t)u_\theta(\mathbf{x}_t, r, t)$;
8: $\mathbf{x}_t \leftarrow \mathbf{x}_t + c(r, t)(\hat{\mathbf{x}}_0' - \hat{\mathbf{x}}_0)$
9: if using MPS+:
10: $\mathbf{x}_t \leftarrow \mathbf{x}_t + c(r, t)(\hat{\mathbf{x}}_0' - \hat{\mathbf{x}}_0)$
11: $\mathbf{x}_t \leftarrow \mathbf{x}_t + (r - t)u_\theta(\mathbf{x}_t, r, t)$;
12: end for
13: **return** $\mathbf{x}_t$

reconstruction, and $\{t\}_{i=1}^{N}$ are the sampling time points. Step 1 is the jump start. Step 4 estimates $\hat{\mathbf{x}}_0$ using (21). Step 5 uses K steps of separable quadratic surrogate with S ordered subsets to solve (23) starting from $\hat{\mathbf{x}}_0$. Steps 7 and 8 are the MPS sampling according to (22).

3) MPS+: Note that $u_\theta(\mathbf{x}_t, r, t)$ in step 7 assumes $\mathbf{x}_t$ follows the unconditional path from t to r and is inaccurate. Hence, $\mathbf{x}_t$ after the posterior correction may lead to better path prediction. It can be simply done by flipping the order of steps 7 and 8 in algorithm 1, leading to steps 10 and 11.

IV. Experiments

A. Simulation Study

Under institutional review board approval, 25,369 slices of non-contrast brain CT from 70 patients were used for training and 3,285 slices from another 9 patients were used for testing. All images were acquired with the NAEOTOM Alpha (Siemens Healthineers, Malvern PA), and reconstructed into 63 keV and 0.5 mm slice thickness. The low-dose simulation was done by forward projection with the OmniTom Elite PCD (Neurologica, Danvers MA) geometry and adding Poisson noise. Three levels of noise were simulated using 2×10^5, 1×10^5, and 5×10^4 initial photons per ray.

An unconditional MeanFlow model was trained using a 2D UNet for 800k iterations with a batch size of 32. A learning rate of 10^{-4} was used with an exponential moving average (EMA) of 0.9999. The UNet was a simplified version of that in DDPM++ [6] and had 9.7M parameters. The model was only trained for images clamped to [-100, 200] HU and all pixels outside the range were unchanged.

During testing, $\mathbf{x}'$ was FBP with Hann filter, $N = 5$, $K = 10$, $t_N = 0.5$. S was set to 12 or 24 and β was chosen from 0.0 to 0.5 with 0.1 interval by minimizing the RMSE inside brain on one of the testing patients for each noise level.

B. Real Data Study

The trained model was applied to 5 patients from SOMATOM Force (Siemens Healthineers, Malvern PA) and another 5 patients from OmniTom Elite PCD. Measurement $\mathbf{y}$ was acquired by forward projection without adding noise.

TABLE I
Quantitative results of the simulation study.

Method	RMSE (HU)↓	SSIM↑	LPIPS↓	KID↓[1]
$N_0 = 5 \times 10^4$				
FBP	23.08	0.8148	0.1944	0.0702
DPS (NFE=50)	**12.83** [2]	**0.8589**	**0.0671**	0.0213
DPS (NFE=10)	13.52	<u>0.8557</u>	0.0727	0.0249
MPS (NFE=10)	13.87	0.8518	0.0771	<u>0.0185</u>
MPS+ (NFE=10)	<u>13.29</u>	0.8539	<u>0.0683</u>	**0.0127**
$N_0 = 1 \times 10^5$				
FBP	17.73	0.8372	0.1535	0.0446
DPS (NFE=50)	11.46	<u>0.8741</u>	0.0614	0.0189
DPS (NFE=10)	**10.88**	**0.8752**	0.0570	0.0153
MPS (NFE=10)	11.53	0.8688	<u>0.0555</u>	<u>0.0057</u>
MPS+ (NFE=10)	<u>11.37</u>	0.8674	**0.0529**	**0.0036**
$N_0 = 2 \times 10^5$				
FBP	14.32	0.8615	0.1098	0.0268
DPS (NFE=50)	9.89	**0.8895**	0.0492	0.0114
DPS (NFE=10)	**9.52**	<u>0.8887</u>	0.0509	0.0112
MPS (NFE=10)	9.75	0.8868	<u>0.0471</u>	<u>0.0035</u>
MPS+ (NFE=10)	<u>9.64</u>	0.8845	**0.0445**	**0.0019**

[1] KID used the central 299×299 pixels to avoid resampling.
[2] For each category, bold is the best and underline is the second best.

For SOMATOM, $\beta = 0.0$ and $S = 24$, and we further set $c(r, t) = 1$ because the images had low noise level. For OmniTom, $\beta = 0.1$ and $S = 12$ because the noise level matched that of $N_0 = 1 \times 10^5$ in the simulation.

C. Implementation of DPS

DPS was trained on the same dataset for 800k iterations, a batch size of 32, and EMA of 0.995. A larger UNet with 35.7M parameters were used. For testing, we implemented both fast ($N = 10$, $K = 10$) and slow sampling ($N = 50$, $K = 1$). Other implementations, including image clamping, jump start, OSSQS, and selection of β and S were the same with MPS.

V. Results

A. Simulation Study

Table I gives the quantitative results for simulation. Note that each sampling step in MPS has two number of function evaluations (NFE). RMSE was calculated for pixels within [0, 80] HU in the ground truth. SSIM, LPIPS (perceptual distance), and KID (kernel inception distance) were calculated after clamping the images to [0, 80] HU. MPS and MPS+ demonstrated the best LPIPS and KID and had similar RMSE and SSIM compared to DPS. As shown in figure 1, DPS images were smoother compared to MPS+, which potentially led to the better RMSE and SSIM. DPS also had hemorrhage-like speckle noises where MPS+ was much cleaner.

B. Real Data Study

We calculated the contrast between manually placed regions of interests (ROI) inside the frontal lobe white matter and the caudate nuclei gray matter. We also calculated the KID between the reconstructed images and the PCCT datasets. The results are summarized in table II. For both scanners, DPS and MPS achieved better contrast-to-noise ratio (CNR) and KID compared to the original FBP images. For SOMATOM, MPS and MPS+ kept the contrast between gray and white

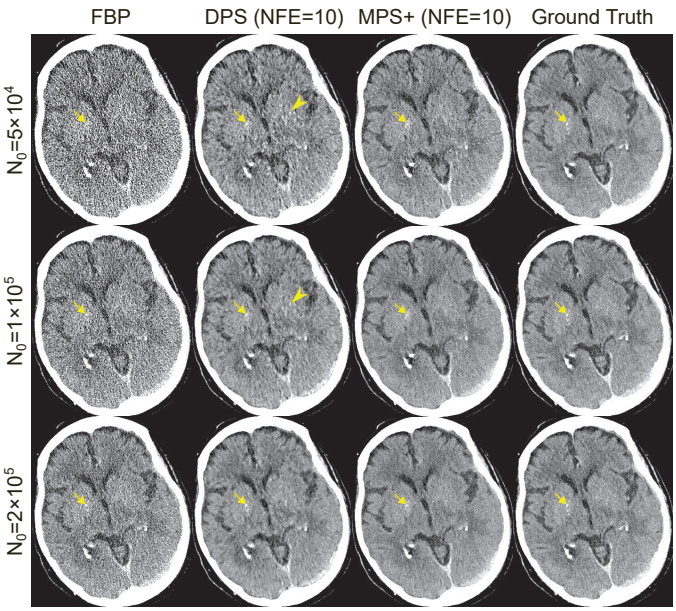


Fig. 1. Reconstruction results of simulation. Both DPS and MPS+ better depicted the hemorrhage (yellow arrow) compared to FBP. However, DPS led to hemorrhage-like speckles (arrow head).

TABLE II
QUANTITATIVE RESULTS OF THE REAL DATA STUDY.

Method	Contrast (HU)[1]	CNR↑	KID↓[2]
SOMATOM Force			
FBP	7.685	1.150	0.0200
DPS (NFE=50)	7.721	1.256	**0.0134**
DPS (NFE=10)	7.742	**1.371**	0.0196
MPS (NFE=10)	7.684	1.229	0.0200
MPS+ (NFE=10)	**7.686**	1.277	0.0159
OmniTom Elite PCD			
FBP	6.276	0.485	0.0573
DPS (NFE=50)	**6.004**	0.851	0.0491
DPS (NFE=10)	5.798	**0.961**	0.0531
MPS (NFE=10)	5.648	0.856	**0.0453**
MPS+ (NFE=10)	5.569	0.860	0.0463

[1] Contrast closer to FBP is considered better (less bias).
[2] KID used the central 299×299 pixels to avoid resampling.

matter whereas DPS boosted the contrast slightly. Similar to simulation, DPS (NFE=10) had slight oversmoothing, which contributed to its best CNR. As shown in figure 2, DPS with both 10 and 50 NFEs smoothed out a small infarction inside the left caudate nuclei whereas MPS+ preserved the feature. MPS+ had the second best KID besides DPS with 50 NFEs.

MPS did not perform as well for the OmniTom Elite PCD in terms of contrast, but it still had better KID compared to DPS. It had contrast loss as demonstrated in table II and figure 2. DPS had better contrast preservation but worse spatial resolution. The main reason for the deteriorate performance of MPS in OmniTom was that the MeanFlow model did not generalize well to portable CT, possibly due to insufficient network capacity. As shown in figure 3, the MeanFlow model could predict reasonable $\hat{\mathbf{x}}_0$ from the jump start for the simulation and SOMATOM data, but the prediction totally lost the brain structure for OmniTom. And DPS generated reasonable but blurred brain structure in $\hat{\mathbf{x}}_0$.

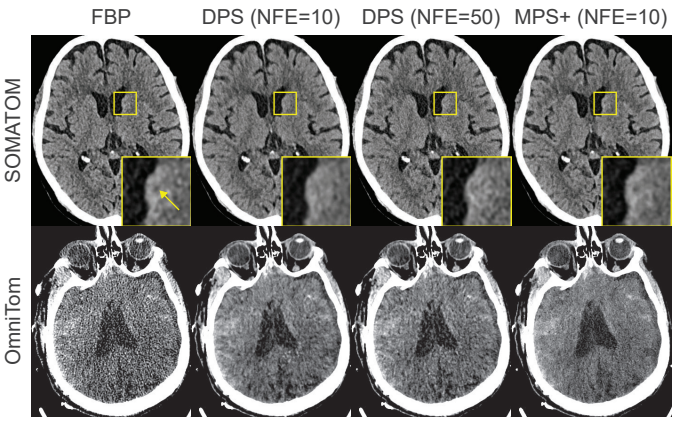


Fig. 2. Reconstruction results of real data. In the SOMATOM results, DPS smoothed out the infarction inside the caudate nuclei whereas MPS+ preserved this subtle lesion (yellow arrow).

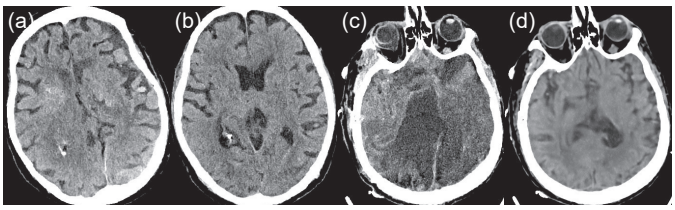


Fig. 3. Predicted $\hat{\mathbf{x}}_0$ from jump start. (a) MPS for simulation data with $N_0 = 5 \times 10^4$; (b) MPS for SOMATOM; (c) MPS for OmniTom; (d) DPS for OmniTom.

VI. CONCLUSION

In this work, we proposed a novel MeanFlow Posterior Sampling (MPS) method for low-dose brain CT reconstruction. By incorporating a pretrained unconditional MeanFlow model and data fidelity, the model could perform fast high-quality sampling. It demonstrated better spatial resolution and texture compared to DPS on the PCCT simulation and EID-CT data. The MPS's performance still needs improvement on the portable CT, due to deficiency of MeanFlow's generalization caused by limited model capacity. We plan to improve the MeanFlow's generalization to the portable CT by increasing model size and using latent models.

REFERENCES

[1] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," in *The Eleventh International Conference on Learning Representations*, 2023.
[2] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow matching for generative modeling," in *The Eleventh International Conference on Learning Representations*, 2023.
[3] Z. Geng, M. Deng, X. Bai, J. Z. Kolter, and K. He, "Mean flows for one-step generative modeling," in *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.
[4] X. Jiang, G. J. Gang, and J. W. Stayman, "Multi-material decomposition using spectral diffusion posterior sampling," *IEEE Transactions on Biomedical Engineering*, vol. 72, no. 8, pp. 2447–2461, 2025.
[5] Z. Geng, Y. Lu, Z. Wu, E. Shechtman, J. Z. Kolter, and K. He, "Improved mean flows: On the challenges of fastforward generative models," *arXiv preprint arXiv:2512.02012*, 2025.
[6] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," in *International Conference on Learning Representations*, 2021.

Median2Median: Zero-shot Suppression of Structured Noise in CT Images

Jianxu Wang, Wenjun Xia, Mannudeep K. Kalra, and Ge Wang, *Fellow, IEEE*

***Abstract*—Image denoising is a fundamental problem in medical imaging. However, clinical CT images are often degraded by structured noise with strong anisotropic correlations, which remains challenging for existing methods. Most data-driven approaches rely on large-scale datasets with high-quality labels, yet they often suffer from limited generalization. Existing zero-shot methods avoid this data dependency, but they are primarily effective under independent and identically distributed (i.i.d.) noise assumptions. To address this gap, we propose Median2Median (M2M), a zero-shot denoising framework designed for structured noise. M2M introduces a novel sampling pattern that constructs pseudo-independent image pairs from a single noisy input via local directional interpolation and generalized median filtering. Coupled with a phase-aware denoising pipeline and a tailored optimization strategy, M2M enables effective Noise2Noise-style training and robust suppression of structured noise. In our studies on clinical CT datasets, M2M consistently outperforms state-of-the-art zero-shot denoising methods, effectively suppressing structured noise across multiple noise levels while preserving fine anatomical details. These results establish M2M as an efficient, data-free solution for structured noise suppression. We hope that M2M opens up new opportunities for practical zero-shot denoising beyond the strict i.i.d. noise assumption.**

***Index Terms*—Image Denoising, Low-dose CT, Structured Noise, Zero-shot, Noise2Noise, Median Filtering.**

I. Introduction

COMPUTED tomography (CT) is a widely used medical imaging modality that provides cross-sectional visualization of anatomical structures. To reduce the health risks associated with radiation exposure, low-dose CT protocols are increasingly adopted in clinical practice. However, reducing the radiation dose inevitably increases noise in CT images, as fewer detected photons lead to stronger statistical fluctuations. In low-dose CT, such noise is often structured and spatially correlated, posing substantial challenges for clinical diagnosis and quantitative analysis. Consequently, image denoising remains a foundational and indispensable task in clinical CT imaging pipelines.

Conventional denoising methods [1] typically assume that noise follows a known statistical distribution, whereas real-world noise is often complex and difficult to characterize with explicit analytical models. Among learning-based methods, early supervised methods [2] learn end-to-end mappings from noisy inputs to clean targets. Noise2Noise (N2N) [3] removes the need for clean references by training on paired noisy observations with independent and identically distributed (i.i.d.) zero-mean noise. Subsequent self-supervised approaches [4], [5] further relax this requirement by eliminating the need for paired noisy observations, instead constructing training pairs directly from individual noisy images within a dataset. Despite these advances, dataset-driven approaches still depend on large curated datasets and often struggle to generalize to unseen noise distributions. Zero-shot denoising addresses this limitation by training directly on a single noisy image without external data. Many zero-shot denoising methods have shown promising performance [6]–[8]. However, such methods still rely on i.i.d. and zero-mean noise assumptions, limiting their effectiveness on clinical images degraded by structured noise with strong spatial correlations.

To robustly suppress structured noise in a zero-shot setting, we propose Median2Median (M2M), a blind denoising framework that introduces a novel sampling pattern, as illustrated in Fig. 1. M2M exploits the directional characteristics of structured noise while preserving the signal statistics required for effective Noise2Noise-style training.

Specifically, the noisy image is first partitioned into non-overlapping 3×3 patches, from which nine spatial phases are defined by grouping pixels at the same relative positions, i.e., a polyphase representation. As illustrated in the left box of Fig. 1, we use the green center phase as an example to describe the phase-wise sampling process. For each sampling location associated with this phase, i.e., the center position of each patch in this example, estimates are obtained via directional interpolation within a local 3×3 neighborhood. The interpolation estimates and the original intensity value are collected into a candidate set for that sampling location. The candidate set is then filtered by generalized median filtering, which retains the three median-centered values to suppress structured-noise-induced outliers. Finally, two values are randomly selected from the filtered candidates and assigned to the corresponding locations of the two sampled sub-images. Repeating this process across all patches constructs a complete sub-image pair with de-structured noise for the given spatial phase. Applying the M2M sampling procedure to all spatial phases yields nine complementary phase-wise sub-image pairs.

During training, the phase-wise sub-image pairs are resampled at every epoch rather than kept fixed, and the network is optimized using the newly sampled pairs. Over n training epochs, all phase-wise pipelines are iteratively optimized, each with n different sampling realizations. During inference, the noisy image is first processed by one M2M sampling operation to generate sub-image pairs for all spatial phases. For each phase, the two sampled sub-images are averaged and fed into the corresponding phase-wise denoising pipeline to produce a denoised output. To reduce sampling variance, this procedure is repeated k times with newly resampled sub-image pairs, and

Jianxu Wang, Wenjun Xia, and Ge Wang are with the Biomedical Imaging Center, Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: wangj68@rpi.edu; xiaw4@rpi.edu; wangg6@rpi.edu).

Mannudeep K. Kalra is with the Department of Radiology, massachusetts general hospital and harvard medical school, Boston, MA 02114 USA (e-mail: mkalra@mgh.harvard.edu)

Code is available at: https://github.com/Jayx-Wang/M2M.

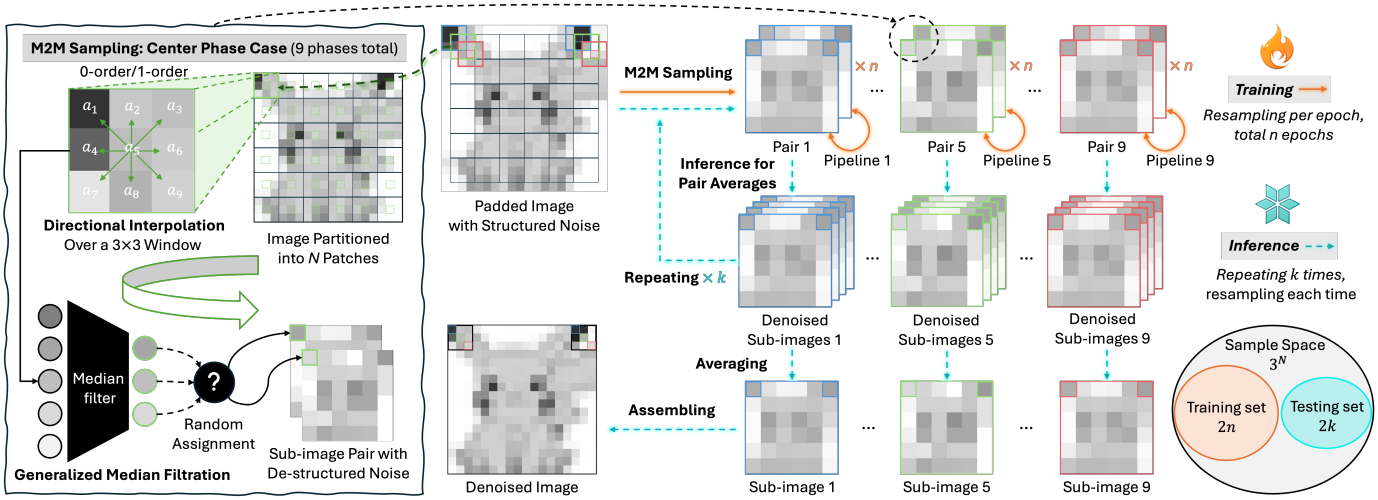


Fig. 1. Overview of the proposed M2M framework. The left box illustrates an example of M2M sampling at the center phase in the first patch. To the right of this box are the training and inference pipelines, where the solid arrows indicate the training process and the dashed arrows denote the inference process.

the resulting outputs are averaged to obtain the final denoised estimate for each phase. Finally, the phase-wise denoised outputs are refilled according to their original sampling positions to reconstruct the complete denoised image.

Our main contributions are summarized as follows:

- **A novel construction of noisy image pairs for structured noise.** We propose a new sampling strategy that leverages directional interpolation to generate multiple estimates and applies generalized median filtering to suppress outliers induced by structured noise. This design enables the construction of surrogate noisy samples that approximately satisfy the assumptions required for Noise2Noise-style learning.
- **A zero-shot denoising framework effective for structured noise suppression.** Unlike existing zero-shot methods primarily designed for i.i.d. noise, M2M explicitly targets structured noise commonly encountered in clinical imaging. Our experiments on clinical datasets show that M2M consistently outperforms state-of-the-art zero-shot methods in structured noise suppression while preserving image details.

II. Methodology

A. Sampling Pattern

Structured noise in clinical images is inherently anisotropic, often amplifying or suppressing pixel intensities along dominant directions and thereby producing coherent stripe-like artifacts. When the sampling domain contains such structured noise, applying fixed sampling patterns [6], [7] tends to propagate directional artifacts into the sampled results, preventing the network from effectively suppressing them. To address this challenge, the proposed sampling pattern generates multiple candidate estimates along different directions and incorporates a robust selection mechanism to extract reliable samples.

1) Directional Interpolation: Let $I \in \mathbb{R}^{H\times W}$ denote a 2D noisy image. Reflective padding is first applied at the image boundaries so that the resulting image $\tilde{I} \in \mathbb{R}^{\tilde{H}\times\tilde{W}}$ can be uniformly divided into non-overlapping 3×3 patches, where $\tilde{H} = \lceil H/3\rceil \times 3$ and $\tilde{W} = \lceil W/3\rceil \times 3$. Within these patches, pixels occupying the same relative position are grouped together, forming nine globally defined spatial phases. This corresponds to a 3×3 polyphase decomposition. Each phase is indexed by a relative offset $(\Delta x, \Delta y) \in \{-1, 0, 1\}^2$, where $(0, 0)$ denotes the center phase, corresponding to the center position within each patch, and the remaining offsets denote the surrounding phases.

For a given spatial phase, the sampling locations are determined by its relative offset and are shared consistently across all patches. At each such location, M2M generates local estimates using directional interpolation within a 3×3 window, which serve as direction-dependent candidates. In our sampling pattern, two efficient interpolation strategies are considered: zero-order interpolation using nearest neighbors and first-order interpolation based on directional averaging.

Zero-order estimates are obtained by selecting neighboring pixel values along predefined directions within a 3×3 window. Formally, a zero-order estimate at a sampling location (i, j) along a directional offset θ_0 is defined as

$$\hat{I}_{\theta_0}(i, j) = \tilde{I}(i + \delta i, j + \delta j), \tag{1}$$

where $\theta_0 = (\delta i, \delta j)$ is selected from either the four-neighborhood offset set $\mathcal{D}_4 = \{(\pm 1, 0), (0, \pm 1)\}$ or the eight-neighborhood offset set $\mathcal{D}_8 = \mathcal{D}_4 \cup \{(\pm 1, \pm 1)\}$.

First-order estimates are formed by averaging the pixel values of two symmetric neighboring pixels along predefined directions within a 3×3 window. Formally, a first-order estimate at a sampling location (i, j) along a directional offset θ_1 is defined as

$$\hat{I}_{\theta_1}(i, j) = \frac{1}{2}\left[\tilde{I}(i + \delta i_\theta, j + \delta j_\theta) + \tilde{I}(i - \delta i_\theta, j - \delta j_\theta)\right], \tag{2}$$

where $\theta_1 = (\delta i_\theta, \delta j_\theta)$ takes values $(1, 0)$, $(1, 1)$, $(0, 1)$, and $(-1, 1)$ for $\theta \in \{0°, 45°, 90°, 135°\}$, respectively.

In addition to the interpolation-based estimates, the original pixel value $\tilde{I}(i, j)$ is retained as a valid candidate to preserve potentially useful local information at the sampling location.

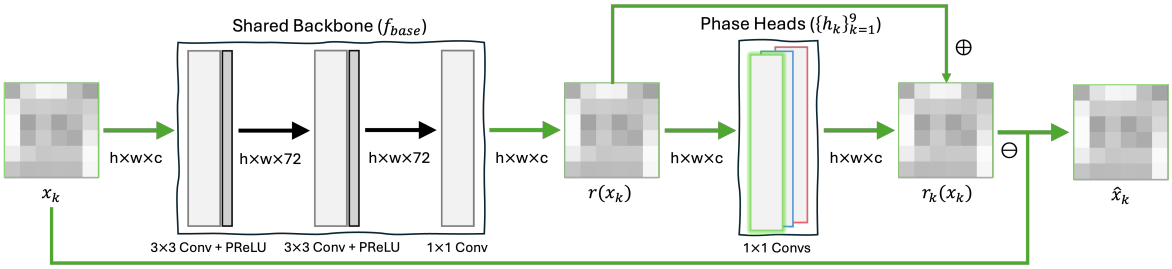


Fig. 2. Phase-aware residual denoising pipeline. The network adopts a lightweight architecture consisting of a shared convolutional backbone f_{base} and phase-specific prediction heads $\{h_k\}_{k=1}^{9}$. In this example, the sub-image pair associated with the green spatial phase activates the corresponding phase head.

Accordingly, the complete candidate set at (i, j) can be written in a unified form as

$$\mathcal{S}(i,j) = \left\{\tilde{I}(i,j)\right\} \cup \left\{\hat{I}_\alpha(i,j) \mid \alpha \in \mathcal{A}\right\}, \tag{3}$$

where the index set $\mathcal{A}$ depends on the interpolation scheme. For zero-order interpolation, $\mathcal{A}$ is set to either $\mathcal{D}_4$ or $\mathcal{D}_8$, whereas for first-order interpolation, $\mathcal{A} = \{(1,0),(1,1),(0,1),(-1,1)\}$.

2) Generalized Median Filtering: To reduce the influence of outliers in the candidate set, M2M applies a generalized median filtering step to select reliable estimates. The key intuition is that candidates strongly biased by structured noise tend to deviate from the central tendency of the candidate set, whereas median-centered candidates are more likely to provide stable local estimates. Specifically, given n candidates sorted in ascending order as $v_{(1)} \le v_{(2)} \le \cdots \le v_{(n)}$, M2M retains the three values centered around the median and defines the filtered candidate subset as

$$\mathcal{S}_f(i,j) = \left\{v_{(\lceil n/2\rceil-1)},\ v_{(\lceil n/2\rceil)},\ v_{(\lceil n/2\rceil+1)}\right\} \subset \mathcal{S}(i,j). \tag{4}$$

3) Random Assignment: After generalized median filtering, M2M randomly assigns the filtered candidates to construct a pseudo-independent sub-image pair. This random assignment serves two purposes: it avoids systematic intensity bias between the two sampled sub-images and enlarges the effective sampling space. Specifically, two distinct elements are uniformly sampled without replacement from the filtered subset $\mathcal{S}_f(i,j)$. The sampled values, denoted as (m_1, m_2) with $m_1 \le m_2$, are drawn as $(m_1, m_2) \sim \mathcal{U}_2(\mathcal{S}_f(i,j))$, where $\mathcal{U}_2(\cdot)$ denotes uniform sampling of two distinct elements from the given set. If the ordered values were always assigned as (m_1, m_2), the first sampled sub-image would have systematically lower intensity than the second one. To remove this assignment bias, a random swapping mechanism is applied. For each spatial location (r, c) in the sub-image pair, a random variable $p \sim \mathcal{U}[0,1)$ is drawn, and the assignment is defined as

$$(x_1,\ x_2)_{r,c} = \begin{cases} (m_1,\ m_2), & \text{if } p < 0.5, \\ (m_2,\ m_1), & \text{otherwise.} \end{cases} \tag{5}$$

This symmetric assignment makes x_1 and x_2 approximately identical in expectation.

For each sampling location, the three filtered candidates yield $\binom{3}{2}$ possible value pairs. The two selected values can be assigned to (x_1, x_2) in two orders, but the symmetric loss treats the two orders equivalently, resulting in $\binom{3}{2} \times \binom{2}{1} \times \frac{1}{2}$ effective sampling outcomes per location. Therefore, if $\tilde{I}$ is partitioned into N patches, the effective sampling space for each spatial phase has size 3^N.

B. Denoising Pipeline

The M2M sampling pattern produces phase-wise sub-image pairs, which may contain phase-dependent sampling bias across different spatial phases. To address this issue, we adopt a lightweight residual architecture that decouples phase-invariant image structure from phase-dependent sampling bias. As illustrated in Fig. 2, the network consists of a shared convolutional backbone followed by phase-specific prediction heads, with each spatial phase associated with a dedicated head. The shared backbone embeds the input into a 72-channel feature space using two 3×3 convolutional layers, each followed by a PReLU activation, and then produces the base residual through a final 1×1 convolution. By default, each phase-specific head is implemented as a lightweight 1×1 convolutional layer. For more complex noise patterns, we also provide a deeper phase-head variant consisting of multiple 3×3 convolutional layers followed by ReLU activations and a final 1×1 convolution.

Given a sampled noisy sub-image x_k from the k-th spatial phase, the shared backbone $f_{\text{base}}(\cdot)$ with learnable parameters θ_b first predicts a phase-invariant base residual,

$$r(x_k) = f_{\text{base}}(x_k;\, \theta_b). \tag{6}$$

To compensate for phase-dependent sampling bias, the base residual is further refined by the corresponding phase head $h_k(\cdot)$ with parameters θ_k, yielding a phase-adapted residual

$$r_k(x_k) = r(x_k) + h_k\big(r(x_k); \theta_k\big). \tag{7}$$

The denoised sub-image is then obtained by subtracting this residual from the noisy input,

$$\hat{x}_k = x_k - r_k(x_k). \tag{8}$$

This process is applied independently to all spatial phases. The resulting denoised sub-images are finally refilled to their original spatial locations to complete the reconstruction.

C. Optimizing Strategy

For each spatial phase $k \in \{1, \ldots, 9\}$, the M2M sampling pattern constructs a pseudo-independent noisy pair $(x_{k,1}, x_{k,2})$ from the noisy image I. These paired samples are destructured and approximately satisfy the statistical requirements for Noise2Noise-style training. Let $F_k(\cdot)$ denote the

phase-specific denoising operator composed of the shared backbone, the k-th phase head, and residual reconstruction, parameterized by θ_b and θ_k.

1) Phase-wise symmetric loss: To exploit the mutual information within each pseudo-independent pair, we employ a symmetric loss [9], where the two samples alternately serve as the network input and target:

$$\mathcal{L}_{\text{sym}}^{(k)} = \frac{1}{2}\,\mathbb{E}_{(x_{k,1},x_{k,2})\sim I}\Big[\,\|F_k(x_{k,1};\theta_b,\theta_k)-x_{k,2}\|_2^2 + \|F_k(x_{k,2};\theta_b,\theta_k)-x_{k,1}\|_2^2\Big]. \tag{9}$$

2) Phase-wise consistency loss: We further incorporate the consistency loss [7] into our optimization strategy. The intuition is that denoising followed by sampling should be consistent with sampling followed by denoising. To make this constraint applicable to structured noise, M2M first reconstructs a de-structured image from the phase-wise noisy pairs.

Specifically, the two samples in each phase-wise pair are averaged to form a de-structured sub-image,

$$\bar{x}_k = \frac{1}{2}\,(x_{k,1}+x_{k,2}), \quad k\in\{1,\dots,9\}. \tag{10}$$

The resulting sub-images $\{\bar{x}_k\}_{k=1}^{9}$ are refilled according to their sampling positions to reconstruct a de-structured image,

$$I' = \mathcal{R}\big(\{\bar{x}_k\}_{k=1}^{9}\big), \tag{11}$$

where $\mathcal{R}(\cdot)$ denotes the phase-wise refilling operator.

For a given phase k, the denoising operator $F_k(\cdot)$ is applied to I' to obtain a phase-wise full-image denoising estimate,

$$\hat{I}'_k = F_k(I';\theta_b,\theta_k). \tag{12}$$

This estimate is then re-sampled at the k-th phase using the same M2M sampling pattern to produce a denoised sub-image pair $(\hat{y}_{k,1},\hat{y}_{k,2})$. The phase-wise consistency loss is defined as

$$\mathcal{L}_{\text{cons}}^{(k)} = \frac{1}{2}\,\mathbb{E}_{(x_{k,1},x_{k,2})\sim I}\Big[\,\|\hat{y}_{k,1}-F_k(x_{k,1};\theta_b,\theta_k)\|_2^2 + \|\hat{y}_{k,2}-F_k(x_{k,2};\theta_b,\theta_k)\|_2^2\Big]. \tag{13}$$

Finally, the overall training objective is formulated as a weighted sum over all spatial phases,

$$\mathcal{L}_{\text{total}} = \sum_{k=1}^{9}\Big(\mathcal{L}_{\text{sym}}^{(k)} + \lambda\,\mathcal{L}_{\text{cons}}^{(k)}\Big), \tag{14}$$

where λ controls the contribution of the consistency loss.

III. Experiments

Experiments were conducted on two clinical CT datasets: the Mayo Clinic dataset from the 2016 NIH–AAPM–Mayo Clinic Low-Dose CT Grand Challenge[1] and a real clinical CT dataset collected at Massachusetts General Hospital (MGH). We compared M2M with several representative baselines, including blind zero-shot methods N2F [6], ZS-N2N [7], and ZS-KAN [8], as well as the classical non-blind method BM3D [1]. Two M2M variants were evaluated: M2M-0 (8), the zero-order interpolation variant with an 8-neighborhood, and M2M-1, the first-order interpolation variant. All zero-shot methods were trained and evaluated independently for each LDCT slice without external training data. Denoising performance was evaluated using peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and Fréchet inception distance (FID).

[1]https://www.aapm.org/grandchallenge/lowdosect/

A. Qualitative Comparison

We visually compared M2M with competing baselines using representative quarter-dose CT slices from the Mayo dataset. The selected slices contain lesions, enabling evaluation of structure preservation in clinically relevant regions. To assess robustness under different denoising difficulties, we selected three cases with varying slice thicknesses and reconstruction kernels: a mild case with 3-mm slice thickness and D45 kernel, a moderate case with 1-mm slice thickness and B30 kernel, and a challenging case with 1-mm slice thickness and D45 kernel. Fig. 3 presents the visual denoising results of M2M and all the aforementioned methods, with zoomed-in views of the highlighted regions of interest (ROIs). The corresponding NDCT images were used as references for computing PSNR and SSIM over the full HU range. Since these NDCT references still contain noticeable residual noise, the reported metrics should be interpreted as reference values rather than absolute measures of denoising quality.

As shown in Fig. 3, BM3D consistently oversmooths fine structures and introduces stripe-like artifacts across all three settings. ZS-N2N suppresses noise effectively only in the mild case but still retains pronounced streaking artifacts; in the moderate and challenging cases, substantial residual noise obscures the lesion regions. N2F and ZS-KAN produce relatively competitive results in the mild and challenging cases, but visible residual artifacts and granular noise remain. In the moderate case, both methods struggle to effectively suppress noise. In contrast, M2M demonstrates strong and consistent performance across all three settings. The proposed median-based filtering effectively suppresses outliers induced by structured noise, leading to cleaner reconstructions with fewer artifacts than N2F and ZS-KAN, particularly in the mild and challenging cases. In the moderate case, M2M is the only method that successfully separates useful anatomical information from structured noise, producing clear lesion regions and well-preserved tissue textures. These visual results demonstrate the superior ability of M2M to suppress structured noise under different imaging conditions while preserving important anatomical details, compared with the baseline methods.

B. Quantitative Comparison

To obtain objective quantitative evaluations, we conducted dataset-level experiments on two clinical low-dose CT benchmarks: the Mayo dataset and the MGH dataset. For the Mayo dataset, we used the 1-mm B30 reconstruction setting and selected 100 paired slices from patient L506, corresponding to slice indices 50–149. For the MGH dataset, we evaluated the methods under three noise levels. Specifically, we used four 120-kVp CT settings reconstructed with filtered back projection (FBP), corresponding to noise index levels of 10,

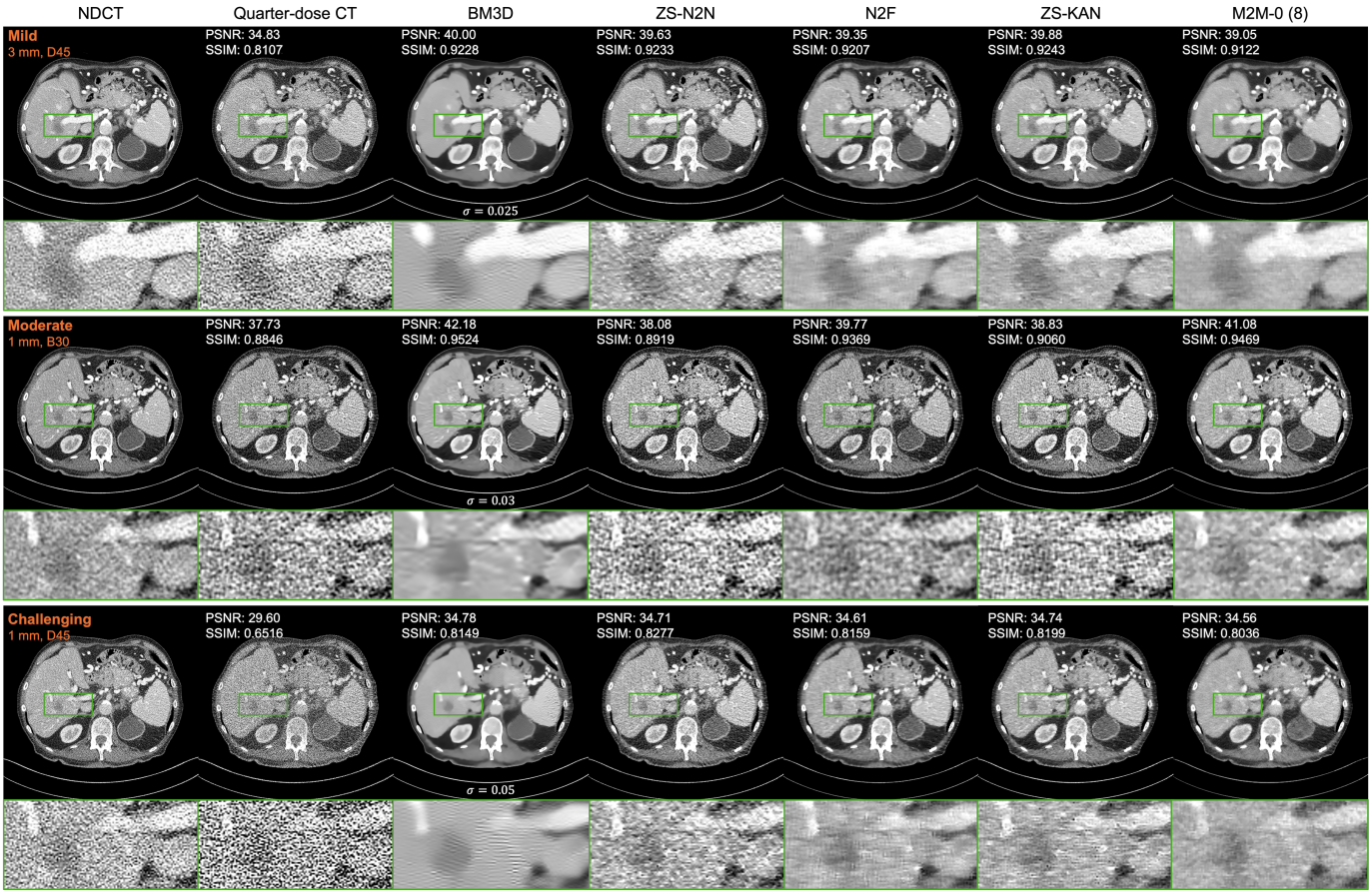


Fig. 3. Denoising results of different methods on representative slices from the Mayo dataset. From top to bottom, the examples correspond to increasing denoising difficulty from mild to challenging cases. Green boxes highlight lesion regions, with corresponding zoomed-in views shown below each image. The display window is $[-160, 240]$ HU. "M2M-0 (8)" denotes the zero-order interpolation variant with an 8-neighborhood.

20, 30, and 40, denoted as NI10, NI20, NI30, and NI40, respectively. A larger noise index indicates a higher noise level. The NI10 images served as high-quality references, whereas NI20, NI30, and NI40 were used as progressively noisier low-dose inputs. For each low-dose setting, 100 central slices were paired with the corresponding NI10 reference slices according to slice location. For both datasets, PSNR and SSIM were computed in the full HU range and reported as mean ± standard deviation over 100 slices. In addition, FID was computed between the denoised results and the corresponding reference distribution.

Table I reports the results on the Mayo dataset. For BM3D, we used $\sigma = 0.03$ as the noise-level parameter. Although BM3D achieves the highest PSNR and SSIM, it relies on prior knowledge of the noise level and tends to oversmooth structural details while introducing stripe-like artifacts. Among the blind zero-shot methods, M2M achieves the best overall performance. M2M-0 (8) obtains the highest PSNR and SSIM, outperforming the strongest non-M2M zero-shot baseline, N2F, by approximately 1 dB in PSNR and 0.01 in SSIM. It also achieves a substantially lower FID than competing zero-shot methods, indicating better distributional similarity to the reference images. In addition, M2M-1 achieves the best FID, suggesting that its reconstructions are most closely aligned with the reference distribution. Although its PSNR is comparable to that of N2F, M2M-1 yields a substantially higher SSIM, indicating better structural preservation. Notably, both M2M variants exhibit the smallest standard deviations in PSNR and SSIM among all blind zero-shot methods, demonstrating strong stability across slices.

TABLE I
QUANTITATIVE COMPARISON ON THE MAYO DATASET. PSNR/SSIM ARE REPORTED AS MEAN ± STANDARD DEVIATION OVER 100 SLICES. BEST AND SECOND-BEST BLIND ZERO-SHOT RESULTS ARE HIGHLIGHTED IN **BOLD** AND UNDERLINE.

Method	Blind	PSNR ↑	SSIM ↑	FID ↓
LDCT	–	39.16 ± 1.15	0.9088 ± 0.0183	88.89
BM3D	No	43.27 ± 0.80	0.9608 ± 0.0058	90.04
ZS-N2N	Yes	39.52 ± 1.11	0.9150 ± 0.0165	85.56
N2F	Yes	40.82 ± 0.96	0.9474 ± 0.0143	60.61
ZS-KAN	Yes	39.54 ± 1.10	0.9154 ± 0.0161	85.10
M2M-0 (8)	Yes	**41.84 ± 0.74**	**0.9565 ± 0.0073**	39.30
M2M-1	Yes	40.85 ± 0.61	0.9559 ± 0.0071	**38.89**

Table II summarizes the results on the MGH dataset. For BM3D, we used $\sigma = 0.025$, 0.03, and 0.05 for NI20, NI30, and NI40, respectively. Although BM3D again achieves the highest PSNR and SSIM, its relatively high FID suggests that the oversmoothed reconstructions deviate from the reference distribution and may not be optimal for clinical use. Among blind zero-shot methods, M2M-0 (8) achieves the best SSIM and FID at the low-noise level. At moderate and high noise levels, M2M-0 (8) consistently obtains the best PSNR, SSIM,

TABLE II
QUANTITATIVE COMPARISON ON THE MGH CLINICAL CT DATASET UNDER DIFFERENT NOISE LEVELS. PSNR/SSIM ARE REPORTED AS MEAN ± STANDARD DEVIATION OVER 100 SLICES. BEST AND SECOND-BEST BLIND ZERO-SHOT RESULTS ARE HIGHLIGHTED IN **BOLD** AND UNDERLINE.

		NI20 (Low noise)			NI30 (Moderate noise)			NI40 (High noise)		
Method	Blind	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓	PSNR ↑	SSIM ↑	FID ↓
LDCT	–	44.39 ± 1.81	0.9665 ± 0.0111	33.93	42.10 ± 1.73	0.9467 ± 0.0132	82.42	39.67 ± 2.08	0.9145 ± 0.0216	116.54
BM3D	No	45.34 ± 2.05	0.9739 ± 0.0106	86.57	44.66 ± 2.55	0.9713 ± 0.0133	81.00	42.56 ± 3.17	0.9573 ± 0.0243	97.30
ZS-N2N	Yes	**44.32 ± 1.79**	0.9663 ± 0.0110	33.26	42.17 ± 1.75	0.9477 ± 0.0132	80.90	39.84 ± 2.14	0.9175 ± 0.0219	114.85
N2F	Yes	42.11 ± 2.11	0.9699 ± 0.0116	26.87	41.30 ± 2.13	0.9594 ± 0.0143	60.47	39.85 ± 2.45	0.9386 ± 0.0247	93.01
ZS-KAN	Yes	44.18 ± 1.78	0.9661 ± 0.0110	33.25	42.10 ± 1.74	0.9474 ± 0.0133	81.16	39.80 ± 2.14	0.9173 ± 0.0221	114.92
M2M-0 (8)	Yes	43.14 ± 1.69	**0.9723 ± 0.0114**	**24.76**	**42.50 ± 2.01**	**0.9665 ± 0.0140**	**44.94**	**41.21 ± 2.70**	**0.9509 ± 0.0251**	**68.50**
M2M-1	Yes	41.15 ± 1.32	0.9683 ± 0.0113	31.83	40.79 ± 1.59	0.9632 ± 0.0140	48.93	39.89 ± 2.23	0.9483 ± 0.0251	72.74

and FID, indicating effective structured-noise suppression and better alignment with the reference distribution. M2M-1 also achieves second-best performance in most settings, further supporting the robustness of the proposed framework.

Overall, these quantitative results demonstrate that M2M achieves superior and stable denoising performance across diverse structured noise conditions in clinical LDCT images.

C. Training Stability

To further assess optimization stability, we compared the training loss curves of learning-based zero-shot methods on the same Mayo test slice. Since the compared methods use distinct objective functions and loss scales, each loss curve was independently normalized to $[0, 1]$ for visualization, allowing the comparison to focus on convergence behavior. As shown in Fig. 4, M2M-0 and M2M-1 exhibit rapid loss reduction during the early training stage and quickly enter a stable low-loss regime. In contrast, ZS-N2N and ZS-KAN converge more slowly, while N2F shows noticeable oscillations after the initial decrease. The zoomed-in view of the first 200 epochs further shows that the M2M variants reach a stable optimization state earlier and maintain smaller fluctuations than the competing zero-shot methods. These results suggest that the M2M framework provides a more stable training paradigm for zero-shot denoising.

IV. CONCLUSION

In this work, we presented M2M, a zero-shot denoising framework for structured noise suppression in CT images. M2M introduces an adaptive sampling pattern that constructs reliable pseudo-independent pairs from a single noisy input, enabling Noise2Noise-style training beyond the strict i.i.d. noise assumption. It further incorporates a phase-aware residual denoising pipeline to handle phase-dependent sampling bias and a phase-wise optimization strategy to enable efficient and stable training. Experiments on clinical CT datasets demonstrate that M2M achieves robust denoising performance and consistently outperforms existing zero-shot methods in suppressing structured noise. Owing to its data-free training paradigm and adaptability to varying noise characteristics, M2M provides a practical and generalizable solution for challenging real-world imaging scenarios with scarce training data, and offers a promising foundation for future extensions to higher-dimensional and more complex imaging modalities.

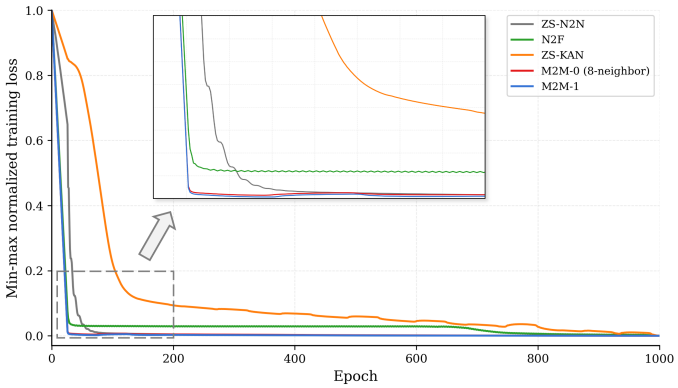


Fig. 4. Training stability comparison of learning-based zero-shot denoising methods on a Mayo slice. The inset zooms in on the first 200 epochs.

ACKNOWLEDGMENTS

This project is funded by the National Institute of Health R01 CA233888. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

REFERENCES

[1] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Transactions on image processing*, vol. 16, no. 8, pp. 2080–2095, 2007.

[2] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.

[3] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[4] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *International conference on machine learning*. PMLR, 2019, pp. 524–533.

[5] C. Niu, M. Li, F. Fan, W. Wu, X. Guo, Q. Lyu, and G. Wang, "Noise suppression with similarity-based self-supervised deep learning," *IEEE transactions on medical imaging*, vol. 42, no. 6, pp. 1590–1602, 2022.

[6] J. Lequyer, R. Philip, A. Sharma, W.-H. Hsu, and L. Pelletier, "A fast blind zero-shot denoiser," *Nature Machine Intelligence*, vol. 4, no. 11, pp. 953–963, 2022.

[7] Y. Mansour and R. Heckel, "Zero-shot noise2noise: Efficient image denoising without any data," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14 018–14 027.

[8] J. Wang and G. Wang, "Zs-kan: Zero-shot image denoising with lightweight kolmogorov-arnold networks," *IEEE Transactions on Biomedical Engineering*, 2025.

[9] X. Chen and K. He, "Exploring simple siamese representation learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 15 750–15 758.

Joint Spectrum Estimation and Material Decomposition for One-kVp-switching Dual-energy CBCT: A Pre-clinical Study

Yu Gao, Tingliang Zhuang, and You Zhang

***Abstract*—One-kVp-switching protocols enable convenient and cost-effective Dual-Energy computed tomography (DECT) on standard Cone-Beam CT (CBCT) scanners but inherently suffer from geometry-inconsistent data acquisition and spectral uncertainty. To address these challenges, we propose a novel framework that jointly estimates energy spectra and basis material decomposition directly from geometry-mismatched, limited-angle projections. Our approach leverages implicit neural representation (INR) combined with physically-informed constraints to robustly handle the ill-posed nature of the problem. Experiments on 2D and 3D phantoms demonstrate that the proposed method achieves high-fidelity material decomposition and accurate spectral recovery, significantly outperforming a conventional image domain method.**

***Index Terms*—Dual-energy CT, basis material decomposition, spectra estimation, implicit neural representation, Cone-beam CT**

I. Introduction

DUAL-energy computed tomography (DECT) has emerged as a powerful clinical modality, enabling quantitative material decomposition and tissue characterization beyond the attenuation-based imaging of conventional single-energy CT. By exploiting the energy-dependent attenuation properties of matter, DECT enables virtual monochromatic imaging and metal artifact reduction to facilitate critical clinical tasks such as disease diagnosis, lesion differentiation, and organ segmentation.

However, the requirement for specialized high-end hardware (e.g., dual-source or fast-kVp-switching systems) limits its widespread accessibility. A cost-effective and hardware-friendly solution is by using standard Cone-Beam CT (CBCT) scanners through a simple "one-kVp-switching" protocol (i.e., changing tube voltage mid-scan to acquire two limited-angle sub-scans with different energy levels).

Nevertheless, this approach introduces substantial challenges. First, standard X-ray beams are inherently polychromatic, and exact energy spectra are often unknown before cumbersome calibration procedures. Inaccurate spectral modeling inevitably leads to beam hardening artifacts and compromises quantitative accuracy. Second, and more critically, one-kVp-switching protocols on continuously rotating gantries result in geometry-inconsistent data acquisition, where high- and low-energy projections are not angularly matched.

Y. Gao, T. Zhuang and Y. Zhang are with the Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX 75390, USA

In this work, we propose a novel framework for the joint estimation of energy spectra and basis material decomposition, specifically tailored for these geometry-inconsistent and limited-angle acquisition protocols. Unlike existing methods that largely rely on consistent geometries or full-scan data, our approach directly addresses the highly ill-posed nature of under-sampled, angularly mismatched projections.

II. Methods

A. Forward polychromatic model of DECT

To accurately model the linear attenuation coefficient $\mu(\mathbf{r}, E)$ at a spatial location $\mathbf{r}$ and energy E, we adopt the basis material decomposition model. The attenuation coefficient is expressed as a linear combination of the energy-dependent mass attenuation coefficients (MAC) of $D = 2$ distinct basis materials:

$$\mu(\mathbf{r}, E) = \sum_{d=1}^{D} \rho_d(\mathbf{r})\psi_d(E), \tag{1}$$

where $\rho_d(\mathbf{r})$ is the material-equivalent density of the d-th material, and $\psi_d(E)$ denotes the known MAC obtained from physics-based databases (e.g., NIST [1]).

Governed by the Beer–Lambert law and accounting for the polychromatic nature of X-ray beams, the discrete forward model relates the material distribution to the measured projections. Assuming negligible non-linear effects (e.g., scattering), the expected log-transformed projection g_j^i for the j-th ray ($j = 1, \ldots, J$) under the i-th energy protocol ($i \in \{L, H\}$) is formulated as:

$$g_j^i = -\ln\left(\sum_{m=1}^{M} S^i(E_m)\Delta E_m \exp\left(-\sum_{n=1}^{N} a_{jn}\sum_{d=1}^{D}\rho_{dn}\psi_d(E_m)\right)\right). \tag{2}$$

Here, a_{jn} denotes the intersection length of ray j with voxel n, and $S^i(E_m)$ represents the normalized effective energy spectrum discretized into M bins (satisfying $\sum_m S^i(E_m)\Delta E_m = 1$). Mathematically, the objective of this framework is to jointly recover the unknown material density maps $\{\rho_{dn}\}$ and energy spectra S^L, S^H by inverting the non-linear system in (2).

B. Physics-informed spectrum estimation

To reduce the dimensionality of spectral recovery, we adopt a method [2]–[4] where the unknown spectra are modeled as linear combinations of pre-simulated spectral components:

$$S^i(E) \approx \sum_{k=1}^{K} w_k^i B_k^i(E), \tag{3}$$

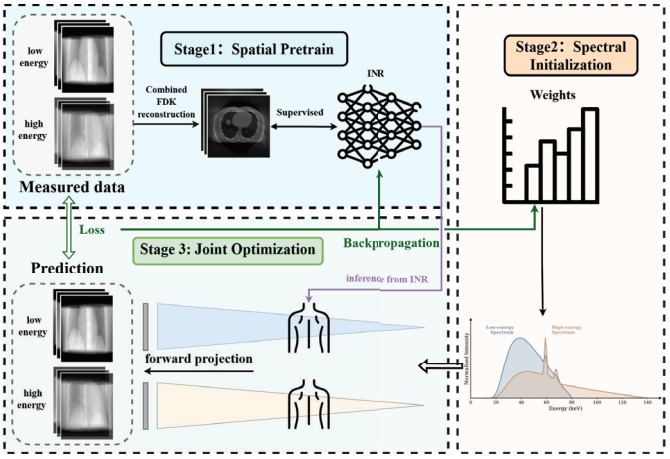


Fig. 1. Schematic overview of the proposed three-stage optimization pipeline.

where $\{B_k^i(E)\}_{k=1}^K$ are the spectral bases for energy spectrum i, and w_k^i are the weights subject to physical constraints which satisfy $w_k^i \geq 0, \sum_k w_k^i = 1$.

C. Approach

We model the spatial distribution of basis materials as a continuous function parameterized by an implicit neural representation (INR). Let $\Phi_\theta : \mathbb{R}^3 \to \mathbb{R}^D$ denote the network with learnable parameters θ, which maps a spatial coordinate $\mathbf{r}$ to the material-equivalent densities $\boldsymbol{\rho}(\mathbf{r}) = [\rho_1(\mathbf{r}), \ldots, \rho_D(\mathbf{r})]$.

To overcome the "spectral bias" of standard MLPs and capture high-frequency anatomical details, we employ Instant Multi-resolution Hash Encoding [5]. The architecture comprises a multi-scale hash table backed by a shallow MLP, which enhances the capability to represent fine-grained anatomical structures compared to pure frequency-encoding based networks. Additionally, to strictly enforce physical constraints (non-negativity and normalization), the spectral weights $\{w_k^i\}$ are parameterized via a Softmax operation on learnable latent variables.

The joint estimation of energy spectra and material images is a highly non-convex problem. To avoid local minima and ensure stability, we implement a three-stage optimization pipeline, as illustrated in Fig. 1.

1) Spatial Pre-training: To embed structural priors, we initialize the INR via supervised learning on analytic Feldkamp-Davis-Kress (FDK) [6] reconstructions. The target volumes are generated adaptively: for full-scan protocols, we can use separately reconstructed dual-energy volumes; for limited-angle protocols, we utilize a unified volume reconstructed from combined (and intensity-scaled) dual-energy projections to suppress streak artifacts. This step establishes a robust geometric backbone, providing a stable starting point for the non-convex optimization.
2) Spectral Initialization: With the spatial parameters θ fixed, we optimize solely the spectral weights to minimize the projection error. This aligns the global spectral model with the pre-trained volumetric geometry.
3) Joint Optimization: Finally, the network parameters θ and spectral weights are refined using an alternating minimization strategy. We iteratively update the spatial representation and spectral components to minimize the total objective function, recovering accurate material densities and energy spectra simultaneously.

The objective function is defined using the ℓ_1-norm to promote robustness against outliers:

$$\mathcal{L} = \sum_{i \in \{L,H\}} \sum_{j \in \mathcal{J}_{\text{batch}}} \| g_j^i(\Phi_\theta, S^i) - \hat{g}_j^i \|_1. \tag{4}$$

Here, $\hat{g}_j^i$ represents the measured projection data, $g_j^i(\cdot)$ is the forward model defined in Eq. (2), and $\mathcal{J}_{\text{batch}}$ denotes the subset of rays sampled in each batch.

III. Simulation Study

A. Experimental Setup

1) Data Simulation and Physical Modeling: The polychromatic X-ray spectra were modeled using the SPEKTR toolkit [7]. We simulated two standard clinical protocols: a low-energy setting at 80 kVp and a high-energy setting at 140 kVp. To capture a wide range of spectral variations and construct a comprehensive spectral base, we applied Aluminum filtration with thicknesses ranging from 0 mm to 18 mm in 2 mm increments. The energy spectra were discretized with a bin width of 1 keV. For the forward projection model, the object attenuation was approximated using a two-basis material decomposition, consisting of water and cortical bone. The corresponding energy-dependent mass attenuation coefficients were sourced from the NIST database via the Xraylib library [8].

2) Phantoms: To evaluate the decomposition accuracy, we constructed a 2D numerical phantom comprising a water-equivalent background and four circular inserts (12.8 mm diameter). The inserts were assigned normalized densities of 1.0 (bone), 0.3, and 0.1 (soft tissues) to cover a broad dynamic range. For clinically-oriented validation, we employed the 3D XCAT phantom [9]. Effective ground-truth basis maps were derived by averaging dual-energy decompositions obtained from virtual monochromatic images (VMIs) generated across 20–140 keV. This multi-energy consensus strategy ensures robust representation of the effective attenuation properties.

B. Implementation Details

1) Network Architecture: The spatial representation utilizes Instant Hash Encoding ($L = 16, F = 2, T = 2^{19}, N_{\min} = 16$), followed by a Layer Normalization (LN) module [10] to align feature distributions. The features are processed by an 8-layer MLP (256 hidden units with ReLU activation), terminating in a 2-channel output with Sigmoid activation to predict water and bone densities.

2) Training Strategy: The framework is implemented in PyTorch 1.11.0. Network parameters and spectral weights are updated via an alternating minimization strategy using the Adam optimizer ($lr = 10^{-4}$). The spectral weights are initialized uniformly. We adopted distinct training protocols tailored to the data dimensionality and acquisition geometry. For 2D experiments, we utilized full-batch updates (50,000

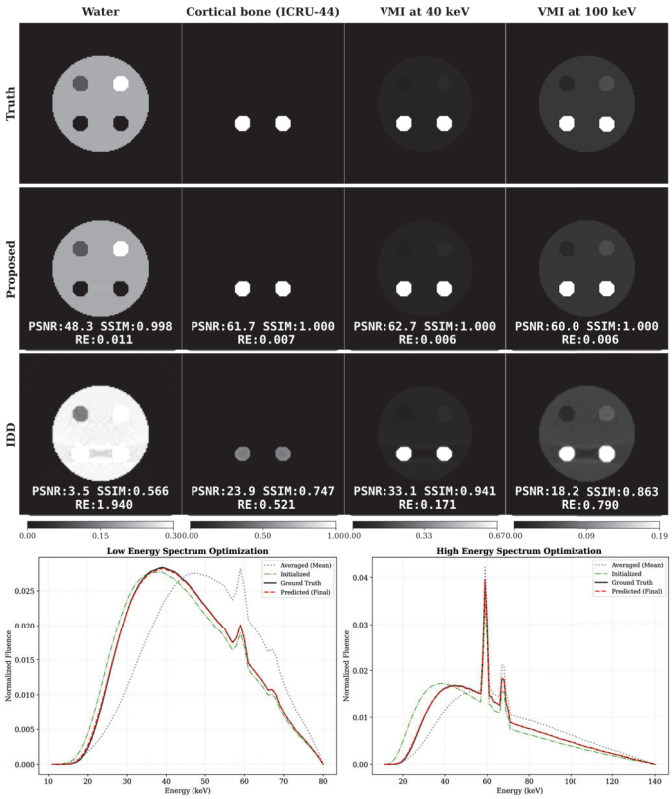


Fig. 2. Results for the 2D numerical phantom. Top: Comparison of basis material maps and VMIs (40/100 keV) between the proposed method and the IDD baseline. Bottom: Evolution of spectral estimation, showing the initial average, the result after Stage 2 initialization, the final estimate, and the ground truth.

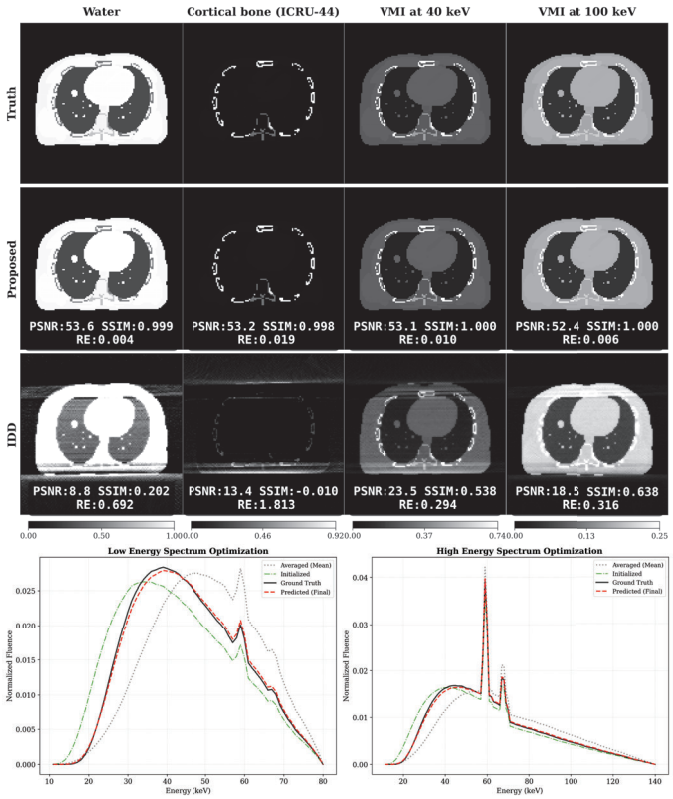


Fig. 3. Results for the 2D slice of XCAT. Top: Reconstructed basis materials and VMIs under the geometry-inconsistent setting. Bottom: Spectral estimation process.

epochs). For 3D experiments, given the memory constraints, we employed a view-based random sampling strategy (120 views per batch, 3,000 epochs). For full-scan data (360 views uniformly distributed over 360° for both low and high-energy scans), we applied standard alternating updates (i.e., 1 step for INR and 1 step for spectra). In contrast, for the limited-angle scenarios, we adopted an asymmetric update schedule: 50 inner steps for the INR parameters θ followed by 10 inner steps for the spectral weights per cycle. This asymmetric strategy stabilizes the geometric recovery under data incompleteness. All computations were performed on a single NVIDIA V100 GPU.

3) Baseline Method: For comparison, we employ a two-step Image-Domain Decomposition (IDD) method. We adopt effective mono-energetic assumptions, setting 50 keV and 90 keV as the effective energies for the low- and high-energy protocols, respectively.

C. Experimental Results

1) 2D Numerical Phantom (Full-scan Baseline Validation): We simulated a 2D fan-beam geometry ($d_{so} = 100.0$ cm, $d_{sd} = 150.0$ cm) with 512 detector bins. For this baseline study, noiseless projections were acquired using a standard full-scan configuration (360 views uniformly distributed over 360°) for both low- and high-energy scans. Reconstructions were performed on a 128×128 grid (0.1 cm voxel size).

As shown in Fig. 2, the proposed method achieves a high agreement with the ground truth. The boundaries of the inserts are well-preserved, and material densities are consistent with the reference, free from beam-hardening streaks. Quantitative metrics including PSNR, SSIM and relative error (RE) shown in the figure further confirm the decomposition accuracy. The estimated spectra exhibit high fidelity to the ground truth, achieving normalized root mean square error (NRMSE) of 0.85% and 0.83% for the low- and high-energy spectrum, respectively.

2) 2D XCAT Phantom (Geometry-Inconsistent 2D validation): Using the same fan-beam geometry, we simulated a one-kVp-switching protocol to introduce geometric inconsistency. The angular range was partitioned into two disjoint sectors: 180 low-energy views (80 kVp) were acquired over $[0°, 180°)$, and 180 high-energy views (140 kVp) were acquired over $[180°, 360°)$. Consequently, the two energy spectra possess non-overlapping angular coverage.

The results are illustrated in Fig. 3. Despite the disjoint angular sampling, our method effectively recovers anatomical structures. In contrast to the IDD method, which suffers from severe artifacts due to the limited-angle acquisition, the proposed framework yields artifact-free basis maps visually comparable to the ground truth. This robustness is supported by the quantitative improvements shown in Fig. 3. Furthermore, the spectral recovery remains highly accurate, yielding NRMSE values of 2.79% and 2.57% for the low- and high-energy spectra, respectively (see Fig. 3 bottom). This demonstrates the method's capability to capture true spectral characteristics

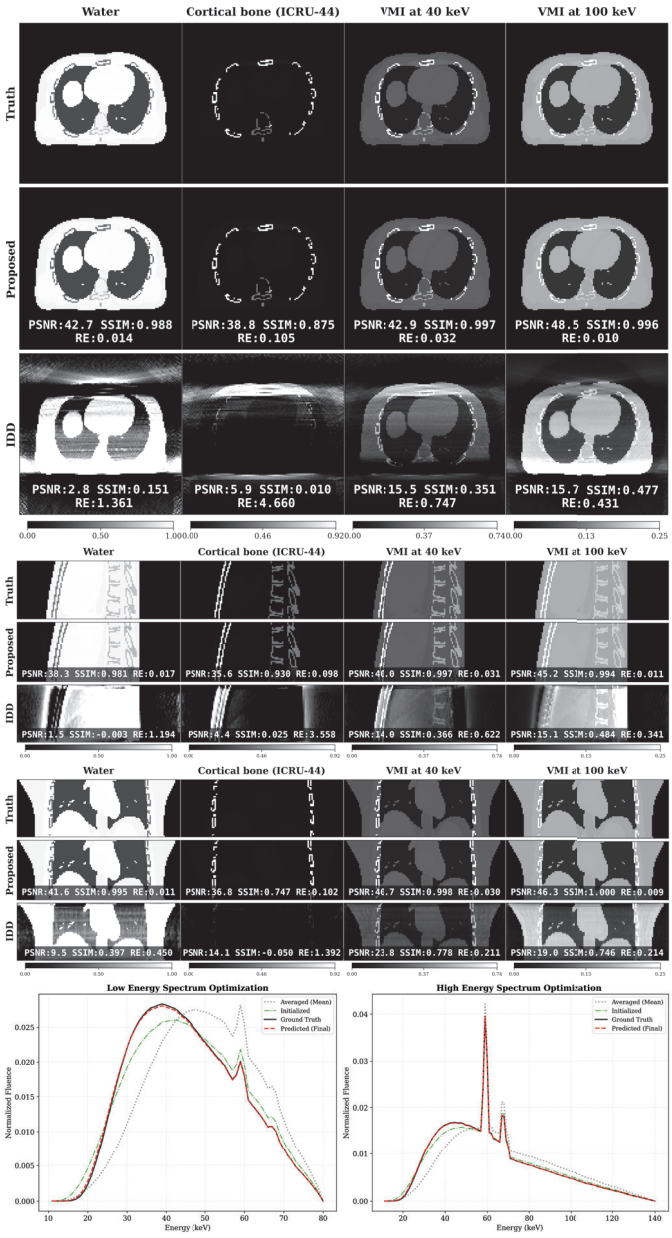


Fig. 4. Volumetric results for the 3D XCAT phantom. Rows 1-3: Representative orthogonal views (Axial, Coronal, and Sagittal) of reconstructed basis materials and VMIs. Bottom: Evolution of spectral estimation.

even with mismatched projections.

3) 3D XCAT Phantom (Geometry-Inconsistent Volumetric Validation): We extended the validation to a 3D CBCT geometry ($d_{so} = 100.0$ cm, $d_{sd} = 150.0$ cm) equipped with a flat-panel detector (384×256 pixels). A geometry-inconsistent protocol was applied: 180 low-energy views over $[0°, 180°)$ and 180 high-energy views over $[180°, 360°)$. This configuration presents a challenging limited-angle problem for each energy, violating the Tuy sufficiency condition. The volume was reconstructed on a $46 \times 128 \times 128$ grid with voxel size of 0.2 cm.

Fig. 4 presents representative orthogonal slices (axial, coronal, and sagittal) of the reconstructed volume. Despite the ill-posedness caused by incomplete angular sampling, the proposed framework successfully suppresses typical limited-angle streaks and geometric distortions. The cross-sectional views confirm that fine structural details are preserved with high fidelity throughout the volume. Quantitative assessments (PSNR/SSIM/RE) further validate the method's scalability and robustness. Beyond achieving high image fidelity, the algorithm maintains accurate spectral recovery with NRMSEs of 1.28% (low-energy) and 0.77% (high-energy) in this fully volumetric, geometry-inconsistent scenario.

IV. Conclusion

In this work, we presented a joint framework capable of reconstructing basis material images and recovering unknown energy spectra directly from limited-angle, geometry-inconsistent dual-energy projection data. By using implicit neural representation and low-rank, physics-driven spectrum representation, the proposed method effectively overcomes the ill-posedness inherent in mismatched projection data and unknown spectra. Experimental results on 2D and 3D phantoms validated that our approach achieves superior decomposition accuracy compared to the image-domain baseline method. This framework shows significant promise for enabling quantitative DECT applications on standard CBCT hardware without complex pre-calibration. Future work will focus on validating the method with real scan data, incorporating scatter correction mechanisms to address CBCT artifacts. Additionally, we plan to extend the framework to handle anatomical mismatches caused by patient motion (e.g., respiration) by jointly optimizing deformable motion fields alongside the spectral-spatial representation.

References

[1] J. H. Hubbell and S. M. Seltzer, "Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements Z= 1 to 92 and 48 additional substances of dosimetric interest," National Inst. of Standards and Technology-PL, Gaithersburg, MD (United States), Tech. Rep., 1995.

[2] W. Zhao, L. Xing, Q. Zhang, Q. Xie, and T. Niu, "Segmentation-free x-ray energy spectrum estimation for computed tomography using dual-energy material decomposition," *Journal of Medical Imaging*, vol. 4, no. 2, p. 023506, Apr 2017.

[3] S. Chang, M. Li, H. Yu, X. Chen, S. Deng, P. Zhang, and X. Mou, "Spectrum estimation-guided iterative reconstruction algorithm for dual energy ct," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 246–258, 2020.

[4] Q. Wu, H. Wei, J. Yu, S. K. Zhou, and Y. Zhang, "Jsover: Joint spectrum estimation and multi-material decomposition from single-energy ct projections," 2025.

[5] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding," *ACM Trans. Graph.*, vol. 41, no. 4, Jul. 2022.

[6] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am. A*, vol. 1, no. 6, pp. 612–619, Jun 1984.

[7] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. H. Siewerdsen, "Technical note: spektr 3.0—a computational tool for x-ray spectrum modeling and analysis," *Medical Physics*, vol. 43, no. 8Part1, pp. 4711–4717, 2016.

[8] T. Schoonjans, A. Brunetti, B. Golosio, M. Sanchez del Rio, V. A. Solé, C. Ferrero, and L. Vincze, "The xraylib library for x-ray–matter interactions. recent developments," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 66, no. 11, pp. 776–784, 2011.

[9] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, 2010.

[10] Z. Xu, H. Li, D. Sun, Z. Li, Y. Li, Q. Kong, Z. Cheng, N. Navab, and S. K. Zhou, "Nerf-based cbct reconstruction needs normalization and initialization," 2025.

Enhancing Pediatric Dual-Layer Spectral CT Quantification Using K-Edge Filtration: Experimental Study

Yinglin Ge[1,2], Olivia F. Sandvold[1,2], Roland Proksa[1], Amy E. Perkins[3], Thomas Koehler[4], Heiner Daerr[4], Yannan Jin[3], Kevin M. Brown[3], Leening P. Liu[1], Peter B. Noël[1]

***Abstract*—Accurate quantitative imaging at low radiation doses is essential for diagnostic pediatric spectral CT. In this experimental study, we demonstrate that combining a K-edge filter with a dual-layer spectral CT system enhances quantitative performance without increasing radiation exposure. A pediatric-sized phantom with tissue-mimicking inserts was scanned using an experimental spectral CT bench system with and without a K-edge filter. The K-edge filter thickness and tube current were optimized in a Python simulation to achieve a higher overall signal-to-noise ratio (SNR) at the same dose compared to the system with no filter. In the experiment, image noise was reduced by 10-24% on average across all inserts in iodine density maps using the Ho filter compared to the non-filtered scan. The bias in the Ho-filtered iodine images showed a 59-69% reduction across all inserts, and all biases were less than 1.9 mg/mL. Water density measurements remained accurate within a 5% error margin in both filtered and non-filtered cases. Overall, the experimental results are consistent with simulations. The combination of Ho K-edge filtration and dual-layer spectral CT significantly enhances quantitative imaging at low pediatric dose levels. This improvement in iodine accuracy may reduce reliance on extra non-contrast scans, streamlining workflow and minimizing radiation exposure in pediatric care.**

***Index Terms*—spectral computed tomography (CT), DECT, quantitative imaging, pediatric imaging**

I. Introduction

SPECTRAL computed tomography (CT) collects X-rays at two or more sensitivity profiles, enabling differentiation of materials [1], [2]. The spectral data also enables the generation of the virtual monoenergetic image (VMI) and material maps based on material decomposition [3]. The iodine density map highlights tissue perfusion and quantifies iodine concentration, which are crucial for accurate diagnosis and treatment. Compared to adults, children are more sensitive to radiation, and their life expectancy is longer, which increases the lifetime risk with additional exposure. Although CT is one of the first-line imaging modalities due to its short scan time and high resolution, dose level controls are particularly stringent for pediatric patients. Spectral CT offers important benefits for pediatric imaging, including improved soft-tissue contrast, reduced artifacts, and the ability to generate virtual non-contrast (VNC) images, which can eliminate the need for an additional scan [4]. In spectral CT, spectral separation is crucial for accurate and stable material decomposition and quantification of contrast agents [5]. However, low tube voltages are commonly used in pediatric patients, which induces noise and bias caused by the poor spectral separation. Additionally, most of the current spectral CT systems are designed for adults, with a limited number of studies focusing on instrumental optimization for children. We propose adding a K-edge material filter to improve spectral separation and adjusting system settings simultaneously to achieve better spectral quantification without increasing dose. Our previous work has demonstrated that adding K-edge material to spectral CT can improve the overall monoenergetic(monoE) SNR in spectra simulation [6-7]. This experimental study extends the previous work by testing the design in a dual-layer CT benchtop using a pediatric-sized phantom. Spectral quantification is evaluated

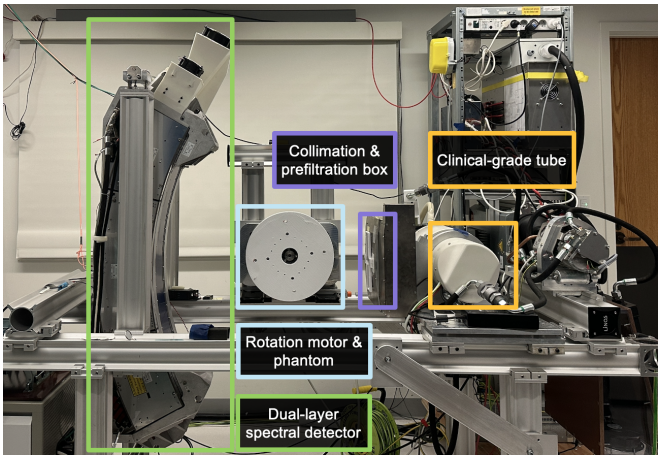


Fig. 1. Photograph of the experimental spectral CT benchtop comprised of clinical-grade tungsten anode tube, collimation and pre-filtration box, rotation motor and mounted phantom stage, and clinical dual-layer spectral CT detector.

This work was supported by the National Institutes of Health (R01EB030494, R01EB035908) and Philips Healthcare.

Y. Ge, O.F. Sandvold, are with [1]Department of Radiology and [2]Department of Bioengineering, University of Pennsylvania, Philadelphia, USA. (e-mail: gyl@seas.upenn.edu). R. Proksa, L.P. Liu, P.B. Noël are with [1]Department of Radiology, University of Pennsylvania, Philadelphia, USA. A. E. Perkins, K. M. Brown, Y. Jin are with [3]Philips Healthcare, Cleveland, USA. T. Koehler, H. Daerr are with [4]Philips Innovative Technologies, Hamburg, Germany.

in reconstructed 2D images. To the best of our knowledge, this study provides the first experimental validation of K-edge filtration for pediatric dual-layer spectral CT at a matched radiation dose.

II. Methods

A. CT benchtop

The spectral CT benchtop system is shown in **Fig 1**. The clinical-grade tungsten anode tube (Philips Healthcare, Cleveland, OH) can achieve voltages between 60 kVp and 150 kVp with a maximum tube power limit of 80 kW. The incident beam collimation was set to a width of 10 slices (less than 10 mm) and the detector scatter was reduced using a 2D anti-scatter grid. The system is equipped with a clinical dual-layer detector (Philips Healthcare, Cleveland, OH). The pediatric phantom is mounted to the rotation motor and spins at a rate of 1 Hz. The source-to-filter distance (SFD) was 150 mm, the source-to-isocenter distance (SID) was 523 mm, and the source-to-detector distance (SDD) was 1042 mm (**Fig. 2**).

B. System simulation and filter design

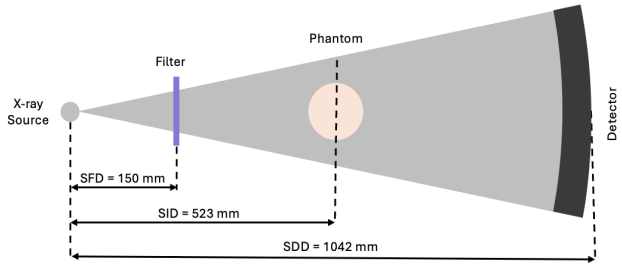


Fig. 2. Sketch of benchtop geometry.

The dual-layer CT benchtop system was modeled in Python using a fan-beam geometry to enable optimization of acquisition parameters and K-edge filter design [6-7]. The variance of the Cramér-Rao lower bound (CRLB) was used to estimate noise in the photoelectric and Compton scatter material decomposition domains [8]. We obtained the signal-to-noise ratio (SNR) of discrete mono-energetic levels and then summed over mono-energies 40-140 kVp in 1-kV increments to estimate the overall performance, represented by the area under the monoE curve (AUMC). The dose was calculated from the total photon energy reaching the phantom surface. The reference system without K-edge filtration was simulated at 100 kVp and 70 mA, following a clinical pediatric protocol. At the center ray, accessible K-edge materials of different thicknesses were tested, and the tube current was adjusted within the system power limitation to match the dose with the non-filter reference system. Titanium was also tested as a non-K-edge comparison. The optimal K-edge material was defined as the material yielding the highest AUMC. The updated current with optimal K-edge material and thickness is used for image collection.

C. Air kerma measurement

Air kerma is a suitable surrogate measure of radiation dose. A dose probe (Accu-Gold+, Radcal, CA, USA) was used to measure the air kerma and to verify consistency with the simulated dose estimates. The probe was placed at the phantom surface, and the air kerma at three tube voltages (100, 120, 140 kVp) with the currents of 70, 80, 100 mA were collected. Then the air kerma at the non-filtered system (no filter, reference current) was compared to that at the K-edge filter system (K-edge filter, updated current) to check the dose consistency.

D. Image collection and analysis

A polylactic acid (PLA) phantom with a diameter of 100 mm was used to mimic a pediatric patient. Four material inserts (Sun Nuclear, FL, USA): water [1 g/mL], 1, 2, 10 mg/mL iodine were placed inside the phantom (**Fig. 3**). The scan time was 1 second, and 2000 projections were collected. The data from the non-filtered system and the K-edge filter system were acquired. Subsequently, a 2D projection-based material decomposition was performed to generate iodine and water-based images without denoising or additional scatter correction.

For the reconstructed images, circular regions of interest (ROIs) were placed at four inserts and in two locations in the phantom background. The mean value (physical density) and standard deviation (noise) were calculated at each insert per slice. Bias was defined by the difference between the measured value and the expected iodine density and water density in mg/mL. The results of the two background ROIs were averaged. Contrast-to-noise ratio (CNR) was calculated by:

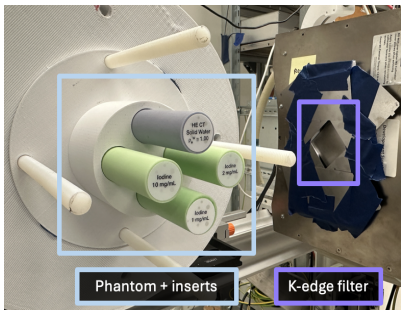


Fig. 3. Pediatric-sized PLA phantom with four inserts. A 0.3 mm Ho filter is attached in front of the collimation box. 0.2 mm tungsten (W) filters are attached around the center K-edge filter to reduce off-focal radiation.

$$CNR = \frac{signal_{insert} - signal_{background_{mean}}}{noise_{background_{mean}}}$$

Noise, bias, and CNR were computed for each insert across the central ten slices, and the mean and standard deviation across these slices were calculated.

III. Results

A. Filter design

Among all K-edge materials, the best two materials for the dual-layer setting are holmium (Ho) and erbium (Er). Ti represents a non-K-edge filter comparison (**Fig. 4**). For K-

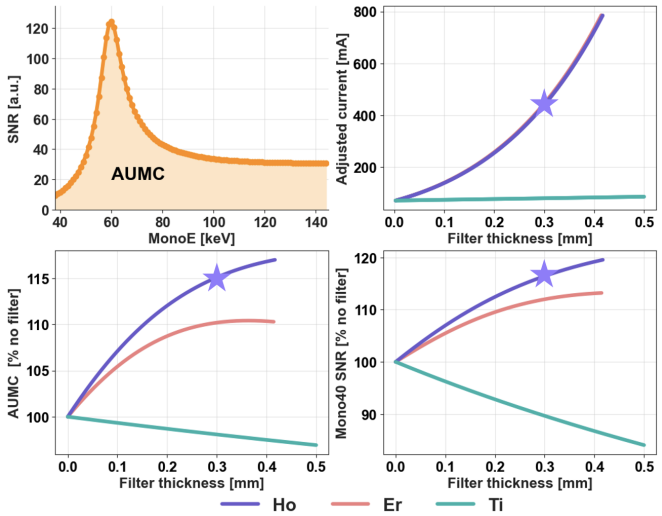


Fig. 4. K-edge filter optimization at center ray. *Upper left*: Plot depicting mono-energetic SNR and the AUMC integral area for the non-filtered system. *Upper right:* Tube current adjusted with filter thickness. *Lower left:* AUMC after adding a filter compared to no filter. *Lower right:* Mono 40 SNR compared to no-filter. The chosen filter and thickness were marked with a star.

edge filter optimization, the tube current was increased to compensate for the extra attenuation. The current maximum was 800 mA due to the tube power limit. The AUMC after adding K-edge materials also increases with thickness. The highest AUMC was achieved with a 0.42 mm Ho, resulting in an approximately 16% increase compared to no filter. Considering the accessibility and potential for damage to the tube, a 0.3 mm Ho was chosen, resulting in approximately a 15% AUMC improvement. After attaching the Ho filter, the tube current will increase to 416 mA to match the dose of the non-filtered system. The Mono 40 SNR curve for different filter thicknesses also follows a similar trend with AUMC, showing that 0.3 mm Ho has both good AUMC and Mono 40 performance. In comparison, the Ti filter has a negative contribution to the AUMC and Mono 40 SNR.

B. Air kerma measurement

The air kerma without the Ho filtration increased linearly with tube current for each tube voltage (**Fig. 5**). Then, the Ho filter was attached, and the tube current was increased to 416 mA, as determined by the simulation. The air kerma with and without the filter matched well, with a difference of less than 10%.

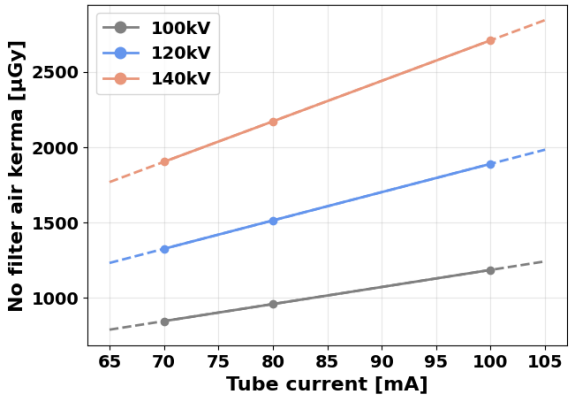


Fig. 5. No filter dose of varying tube voltage and current pairs.

C. Iodine/water image quantification

Noise. For all four inserts in the iodine image, noise decreased after adding the Ho filter and adjusting the current, see **Table I. A**. The highest noise decrease was 24.4% for the 2 mg/mL iodine insert, and the average decrease for iodine inserts was 18.4%. Generally, the absolute noise values for iodine inserts after K-edge filtration increase with higher iodine concentrations, achieving the highest value of 0.49 mg/mL at the 10 mg/mL iodine insert. The variation between slices was low (less than 0.02 mg/mL).

For the water image, the noise followed a similar trend to that in the iodine image (**Table I. B**). The absolute noise value for all ROIs of inserts and background was lower than 0.02 g/mL, and the standard deviation across slices was less than 0.008 g/mL. For the background ROI, the noise decreased by around 23% and 12% for iodine and water images, respectively.

Bias. In the iodine density images, both with and without the Ho filter scans, a negative bias was observed, indicating that the measured iodine concentrations were lower than the true values. The absolute values of this bias increased with higher iodine concentrations, reaching a maximum of 1.78 mg/mL without the Ho filter and 0.73 mg/mL with the filter for the 10 mg/mL insert. Compared to non-filtered acquisitions, bias analysis of filtered images showed a reduction across all inserts, with bias as low as 0.09 mg/mL. The water insert showed the highest decrease, with about 70% reduction. The average bias percentage decrease for iodine inserts compared to no filter was 63%.

The measured values were all higher than the true value in the water image, resulting in biases greater than zero. The bias increased with iodine concentration in the inserts, sharing the

TABLE I. A
AVERAGE NOISE [MG/ML] OF IODINE IMAGE

Insert		no filter	Ho filter	noise decrease [% no filter]
Water		0.48±0.01	0.37±0.01	22.4
Iodine	1 mg/mL	0.47±0.01	0.37±0.01	20.9
	2 mg/mL	0.53±0.02	0.40±0.01	24.4
	10 mg/mL	0.54±0.02	0.49±0.02	9.9
Background		0.47±0.01	0.36±0.00	22.9

TABLE I. B
AVERAGE NOISE [MG/ML] OF WATER IMAGE

Insert		no filter	Ho filter	noise decrease [% no filter]
Water		17.11±0.30	13.23±0.40	22.6
Iodine	1 mg/mL	16.83±0.48	13.61±0.32	19.2
	2 mg/mL	19.04±0.75	14.13±0.43	25.8
	10 mg/mL	18.30±0.58	15.53±0.64	15.2
Background		18.67± 0.40	16.47 ±0.32	11.7

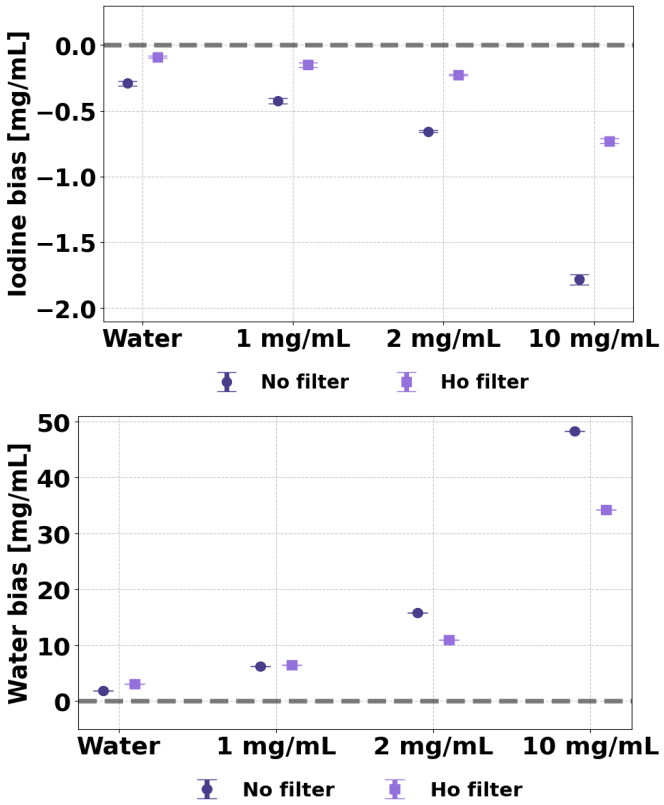


Fig. 6. Bias of iodine and water image.

trend observed in the iodine map. Water density measurements remained accurate within a 5% error in both filtered and non-filtered cases. After filtration, the absolute bias across all ROIs was lower than 0.04 g/mL. The 2 and 10 mg/mL inserts exhibited approximately 30% reductions in bias, whereas the remaining inserts had biases near zero, making further reductions difficult to discern.

CNR. With Ho filtration, CNR increased for all inserts, with 30% improvement for the water insert and 38%-45% improvement for the iodine inserts relative to the no-filter condition. The largest gain occurred for the 2 mg/mL insert, with a 45.3% improvement.

IV. Discussion

In this study, we experimentally demonstrate that filtration can serve as a practical and affordable strategy to enhance pediatric-specific spectral CT performance. Incorporating a K-edge filter and adjusting the tube current, we reduced spectral overlap and suppressed the correlated noise in material decomposition results without increasing dose. This was demonstrated by the decreased noise in water/iodine images. The decreased noise in decomposed results can lead to the improvement of material differentiation in spectral applications. The resulting gains in spectral accuracy can be traded for dose reduction while maintaining image quality.

The optimal filter material and thickness depend strongly on the K-edge energy and tube settings, as attenuation differences in the spectral-overlap region drive most of the achievable separation. For the dual-layer detector system evaluated here, holmium (K-edge: 55.6 keV) generally provides the most favorable performance, with its ideal thickness varying across tube voltages. The Ti filter degrades spectral performance because it primarily hardens the beam, narrowing the effective spectrum. This reduced spectral separation is unfavorable for detector-based spectral imaging, where broader spectral diversity between energy channels is essential for accurate material decomposition.

Dose measurements before and after adding the filter differ by roughly 10%. While dose does not directly affect quantitative bias, it influences noise and contrast (since the bias is noise induced) [9-10]. Image noise scales with the inverse square root of dose and CNR scales with its square root [11] if the bias is neglectable. In this study, the dose after adding the filter and adjusting the tube current was approximately 8% higher than that of the non-filtered reference, which may contribute to improved quantitative accuracy but remains within the expected experimental uncertainty of dose measurements. However, based on the known relationship, an equivalent reduction in tube current to match the reference dose would only partially offset these gains. Consequently, even after dose matching, the Ho-filtered acquisition is expected to retain improved noise performance, higher CNR, and reduced quantitative bias compared with the non-filtered case.

There are some limitations to this approach. Tube-power constraints limit its use to pediatric protocols, where low voltage and low current are typically employed. Moreover, although the optimization is based on simulated AUMC values, comparable experimental measurements are not always straightforward to obtain.

To better guide filter design, we are also developing a 2-D image simulation toolbox that reflects the benchtop geometry and acquisition settings. This framework enables pixel-wise AUMC evaluation and visualization, supporting more effective optimization. For the experiment, future work will evaluate the impact of K-edge filtration on conventional image quality to more fully characterize its clinical benefits. Moreover, this approach is not limited to dual-layer detector systems and can be extended to other spectral CT platforms, including photon-counting CT.

V. Conclusion

This study demonstrates the experimental feasibility of combining a K-edge filter with appropriate system parameter adjustments to achieve high spectral quantification accuracy in low-dose pediatric CT scans. These results provide experimental validation of our earlier simulation work[6-7]. Future work will focus on demonstrating and evaluating potential clinical benefits using lifelike phantoms.

REFERENCES

[1] K. Taguchi, I. Blevis, and K. Iniewski, *Spectral, Photon Counting Computed Tomography: Technology and Applications*, 1st ed. First edition. | Boca Raton : CRC Press, 2020. | Series: Devices, circuits, & systems: CRC Press, 2020. doi: 10.1201/9780429486111.

[2] A. Parakh *et al.*, "Dual-Energy CT Images: Pearls and Pitfalls," *RadioGraphics*, vol. 41, no. 1, pp. 98–119, Jan. 2021, doi: 10.1148/rg.2021200102.

[3] D. Cester, M. Eberhard, H. Alkadhi, and A. Euler, "Virtual monoenergetic images from dual-energy CT: systematic assessment of task-based image quality performance," *Quant. Imaging Med. Surg.*, vol. 12, no. 1, pp. 726–741, Jan. 2022, doi: 10.21037/qims-21-477.

[4] S. Gallo-Bernal, V. Peña-Trujillo, and M. S. Gee, "Dual-energy computed tomography: pediatric considerations," *Pediatr. Radiol.*, vol. 54, no. 13, pp. 2112–2126, Oct. 2024, doi: 10.1007/s00247-024-06074-5.

[5] B. Krauss, K. L. Grant, B. T. Schmidt, and T. G. Flohr, "The Importance of Spectral Separation: An Assessment of Dual-Energy Spectral Separation for Quantitative Ability and Dose Efficiency," *Invest. Radiol.*, vol. 50, no. 2, pp. 114–118, Feb. 2015, doi: 10.1097/RLI.0000000000000109.

[6] O. F. Sandvold, Y. Ge, R. Proksa, and P. B. Noël, "Double bowtie design for high sensitivity pediatric spectral CT," *Conf. Proc. Int. Conf. Image Form. X-Ray Comput. Tomogr.*, vol. 2024, pp. 268–271, Aug. 2024.

[7] Y. Ge, O. F. Sandvold, A. E. Perkins, R. Proksa, and P. B. Noël, "High-fidelity prefiltration using a double bowtie design for quantitative low-dose pediatric spectral CT imaging," in *Medical Imaging 2025: Physics of Medical Imaging*, J. M. Sabol, S. Abbaszadeh, and K. Li, Eds., San Diego, United States: SPIE, Apr. 2025, p. 4. doi: 10.1117/12.3046287.

[8] E. Roessl and C. Herrmann, "Cramér–Rao lower bound of basis image noise in multiple-energy x-ray imaging," *Phys. Med. Biol.*, vol. 54, no. 5, pp. 1307–1318, Mar. 2009, doi: 10.1088/0031-9155/54/5/014.

[9] L. W. Goldman, "Principles of CT: Radiation Dose and Image Quality," *J. Nucl. Med. Technol.*, vol. 35, no. 4, pp. 213–225, Dec. 2007, doi: 10.2967/jnmt.106.037846.

[10] M. J. Siegel, B. Schmidt, D. Bradley, C. Suess, and C. Hildebolt, "Radiation Dose and Image Quality in Pediatric CT: Effect of Technical Factors and Phantom Size and Shape," *Radiology*, vol. 233, no. 2, pp. 515–522, Nov. 2004, doi: 10.1148/radiol.2332032107.

[11] X. Tian and E. Samei, "Accurate assessment and prediction of noise in clinical CT images: Assessment and prediction of noise in clinical images," *Med. Phys.*, vol. 43, no. 1, pp. 475–482, Dec. 2015, doi: 10.1118/1.4938588.

Dual-Domain Posterior Mean-Guided Diffusion Bridge Model for Sparse-View CT Reconstruction

Xun Zhang, Hewei Gao, Li Zhang, Yuxiang Xing, Changyu Chen*, and Zhiqiang Chen*

***Abstract*—Sparse-view computed tomography (CT) is a typical ill-posed problem that poses significant challenges for high-quality reconstruction. While deep learning-based methods show promise, CNN-based approaches often suffer from over-smoothing and diffusion-based methods typically face a fidelity-perception trade-off. In this paper, we propose a novel dual-domain Posterior Mean-Guided Diffusion Bridge model (PMG-DB) for sparse-view CT reconstruction. PMG-DB efficiently combines the strengths of CNN- and diffusion-based methods in a dual-domain framework. A Posterior Mean Estimation Network (PME-Net) in the projection domain compensates for missing data and estimates the posterior mean as an informative initialization for the diffusion bridge model. An Image-Domain Diffusion Bridge Model (IDDB) further refines the reconstruction by modeling the clean image distribution. Experimental results on low-dose sparse-view CT scans demonstrate that PMG-DB achieves high fidelity and improved perceptual quality, outperforming existing methods in terms of reconstruction quality and robustness.**

***Index Terms*—Sparse-view CT, image reconstruction, deep learning, diffusion bridge model**

I. Introduction

Sparse-view computed tomography (CT) is a promising strategy for reducing radiation dose and accelerating the scanning process. However, the limited and incomplete data acquisition introduces significant challenges for accurate image reconstruction by conventional filtered back-projection (FBP) and iterative reconstruction methods.

Recent advance in deep learning (DL) methods have shown great promise in various ill-posed problems. Convolutional neural network (CNN)-based methods have been widely explored for sparse-view CT reconstruction including artifact suppression in the image domain [1], missing data compensation in the projection domain [2], and more comprehensive dual-domain optimization frameworks [3]. Despite their effectiveness, CNN-based approaches are prone to over-smoothing effects and may struggle to effectively capture fine structural details. Diffusion models [4] address the over-smoothing issue through a progressive denoising process. However, their reverse process is typically initialized from pure Gaussian noise, which lacks structural information about the target image distribution and results in high computational cost. Diffusion bridge models [5] show great promise for CT reconstruction by learning a stochastic transformation between two specific distributions, enabling the reverse process to be initialized directly from the corrupted image which accelerates convergence speed and also contains more structural priors. However, these diffusion-based models typically face a fidelity and perception trade-off [6], as they approximate the posterior distribution rather than strictly enforcing data consistency. While some methods [7] address distortion by incorporating projection data consistency constraints, they rely heavily on the accuracy of the projection data itself. When projection data are severely corrupted by noise, these constraints become unreliable and may introduce additional reconstruction errors.

In this work, we propose a novel dual-domain Posterior Mean-Guided Diffusion Bridge model (PMG-DB) for sparse-view CT reconstruction, which combines the fidelity preservation strengths of CNN-based methods with the powerful detail generation capabilities of diffusion bridge models. Specifically, PMG-DB integrates a posterior mean estimation network to construct a high-fidelity initialization and an image-domain diffusion bridge model to refine the reconstruction by generating fine structural details. Experimental results demonstrate that the proposed method reveals more reliable structures and achieves superior image quality, while exhibiting stronger generalization capability on a higher noise level.

II. Methods

A. Formulation of Sparse-View CT

Let $\mathbf{A}$ is the system matrix for full-view projection y^*, $\mathbf{M}$ the sparse-view subsampling mask. The sparse-view projection y of the imaged object x can be formulated as

$$y \triangleq \mathbf{M}(\mathbf{A}x + n) = \mathbf{M}y^* \tag{1}$$

where n is the noise term.

B. Diffusion Bridge Model

Let $X_t \in \mathbb{R}^d$ denote a d-dimensional, continuous-time stochastic process over $t \in [0, 1]$. Diffusion bridge models aim to sample from the distribution $X_0 \sim p(X_0)$, conditioned on a corrupted observation $X_1 \sim p(X_1|X_0)$. In the sparse-view CT context, X_0 denotes the clean image distribution with its realization x, while X_1 corresponds to the degraded image distribution, e.g. direct FBP reconstruction from sparse-view projection with its realization x_{FBP}.

Following the formulation in I^{2}SB [5], the intermediate state X_t can be defined as the product of Gaussian distributions

This work was supported in part by the National Natural Science Foundation of China under Grant 12327809, and in part by China Postdoctoral Science Foundation No. 2025M780861, in part by the Beijing Natural Science Foundation No. L252021.

Xun Zhang, Hewei Gao, Li Zhang, Yuxiang Xing, Changyu Chen, and Zhiqiang Chen are with the with Key Laboratory of Particle and Radiation Imaging (Tsinghua University), Ministry of Education, China and Department of Engineering Physics, Tsinghua University, Beijing 100084, China. (Corresponding authors: Changyu Chen, Email: ccy@mail.tsinghua.edu.cn; Zhiqiang Chen, Email: czq@mail.tsinghua.edu.cn)

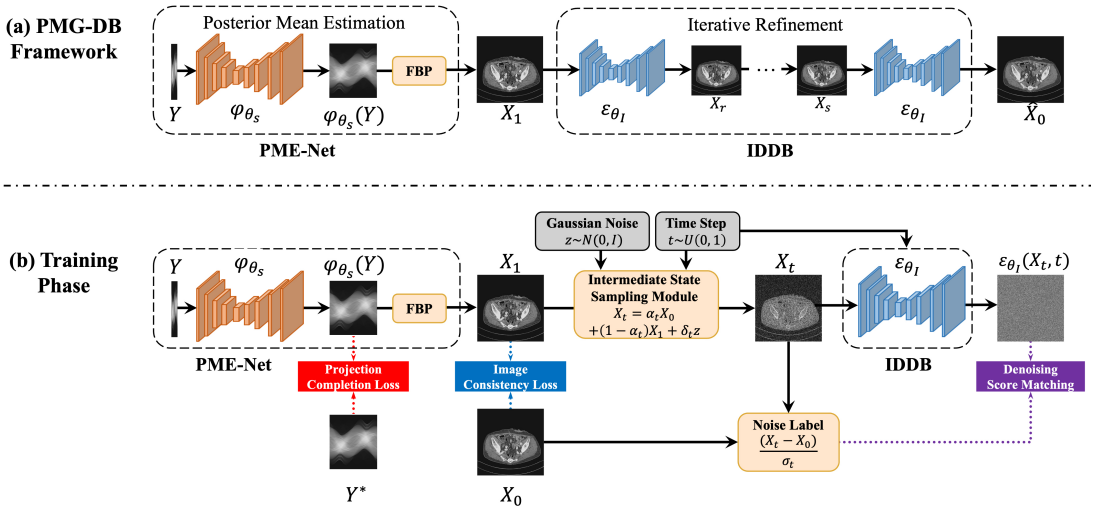


Fig. 1. Diagram for the proposed PMG-DB for sparse-view CT reconstruction. (a) The framework of PMG-DB. (b) The end-to-end training process.

$\mathcal{N}(X_t; X_0, \sigma_t^2)$ and $\mathcal{N}(X_t; X_1, \bar{\sigma}_t^2)$. The posterior distribution of X_t is given by

$$p(X_t|X_0, X_1) = \mathcal{N}(X_t; \alpha_t X_0 + (1-\alpha_t)X_1, \delta_t^2 I), \quad (2)$$

where $\alpha_t = \frac{\bar{\sigma}_t^2}{\sigma_t^2+\bar{\sigma}_t^2}$, $\delta_t^2 = \frac{\sigma_t^2\bar{\sigma}_t^2}{\sigma_t^2+\bar{\sigma}_t^2}$. By the reparameterization trick, X_t can be sampled as

$$X_t = \alpha_t X_0 + (1-\alpha_t)X_1 + \delta_t z, \quad z \sim \mathcal{N}(0, I). \quad (3)$$

The objective is to train a time-dependent neural network $\varepsilon_\theta(\cdot)$ that predicts the clean image X_0 from any intermediate state X_t. Analogous to denoising score matching (DSM) [8], the training objective can be formulated as

$$\min_\theta \mathbb{E}_{X_0, X_1, t}\left[\left\|\varepsilon_\theta(X_t, t) - \frac{X_t - X_0}{\sigma_t}\right\|_2^2\right]. \quad (4)$$

After training, X_0 can be reconstructed by iterative refinement from X_1 using DDPM-style ancestral sampling. However, directly applying the diffusion bridge model with FBP-based initialization faces several challenges. On the one hand, diffusion bridge models are inherently constrained by the fidelity–perception trade-off, which limits their ability to simultaneously achieve high perceptual quality and accurate structural recovery. This trade-off compromises the reconstruction fidelity necessary for reliable clinical diagnosis. On the other hand, severe artifacts in the FBP reconstruction can obscure critical structural details, leading to a degraded and biased initialization that may misguide the subsequent diffusion bridge generation process.

C. Dual-domain Posterior Mean-Guided Diffusion Bridge Model

To address these issues, it is essential to revisit the data-driven learning mechanism. Diffusion bridge models approximate the data distribution, enhancing perceptual quality but potentially introducing structural inconsistencies. In contrast, CNN-based methods produce deterministic outputs with high fidelity, but often converge to the posterior mean conditioned on the observed data, resulting in over-smoothing reconstructions that lack fine details. By combining these complementary objectives, a balance can be achieved between accurate mean estimation and faithful distribution modeling.

For sparse-view CT reconstruction, since projection data retain object-specific information, simultaneous learning in both the projection and image domains is more comprehensive. CNNs, with their ability to capture local spatial dependencies, are well-suited for completing missing projections and suppressing noise in the projection domain. Meanwhile, diffusion bridge models, with their strong generative capabilities, excel at recovering fine structural details and improving image quality in the image domain.

Building on these analyses, we propose a Posterior Mean-Guided Diffusion Bridge model (PMG-DB), which contains a posterior mean estimation network (PME-Net) $\varphi_{\theta_S}(\cdot)$ and an image-domain diffusion bridge model (IDDB) $\varepsilon_{\theta_I}(\cdot)$, as illustrated in Fig. 1. The PME-Net completes sparse-view projection Y and subsequently provide a reliable approximation of the posterior mean $\varphi_{\theta_S}(Y)$. The completed projections are reconstructed via FBP to generate an initial image, which serves as the starting point for diffusion bridge sampling. The process can be formulated as

$$X_1 = \mathbf{R}^{-1}(\varphi_{\theta_S}(Y)). \quad (5)$$

where $\mathbf{R}^{-1}$ is the FBP operator with RamLak filtration. The IDDB performs iterative sampling over the clean image distribution X_0, enabling further recovery of fine structural details and high-quality generation.

By integrating PME-Net and IDDB, the training of PMG-DB approximates to solving the following constrained optimization problem:

$$\min_{\theta_S, \theta_I} \mathbb{E}_{X_0, Y}\left[\left\|X_0 - \hat{X}_0\right\|_2^2\right], \text{s.t.} \quad p(\hat{X}_0) = p(X_0), \quad (6)$$

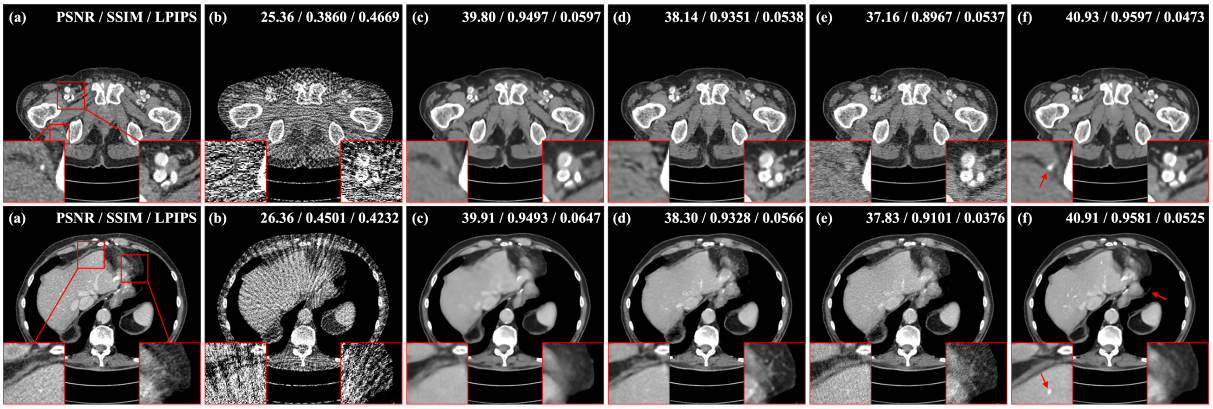


Fig. 2. Qualitative comparison among different methods: (a) Ground Truth, (b) FBP, (c) FBPConvNet, (d) I^{2}SB, (e) CDDB, (f) PMG-DB. The display window is [-160, 240] HU.

where $\hat{X}_0$ is the final reconstruction of PMG-DB via PME-Net estimation and IDDB sampling given the sparse-view projection Y. This formulation ensures high fidelity posterior mean estimation by minimizing the ℓ_2 reconstruction error, while simultaneously enabling high-quality sample generation through distribution alignment.

D. Network Architecture and Loss Function

PME-Net adopts a standard U-Net architecture [9] with bilinear upsampling, while IDDB employs a canonical denoising U-Net architecture widely used in denoising diffusion probabilistic model (DDPM) [4]. Instead of directly employing a pretrained sinogram-domain network, which may introduce distribution mismatch across different datasets, we train PME-Net and IDDB in an end-to-end manner. By iteratively sampling X_t, the end-to-end training objective is formulated as

$$\min_{\theta_I,\theta_S} \mathbb{E}_{X_0,Y,t}\Bigg[\underbrace{\left\|\varepsilon_{\theta_I}(X_t,t)-\frac{\mathrm{sg}(X_t)-X_0}{\sigma_t}\right\|_2^2}_{\text{IDDB denoising score matching}} + \underbrace{\lambda_1\cdot\|\varphi_{\theta_S}(Y)-Y^*\|_2^2}_{\text{PME-Net projection completion}} + \underbrace{\lambda_2\cdot\left\|\mathbf{R}^{-1}\big(\varphi_{\theta_S}(Y)\big)-X_0\right\|_2^2}_{\text{PME-Net image consistency}}\Bigg], \tag{7}$$

where $\mathrm{sg}(\cdot)$ is the stop-gradient operator, and λ_1 and λ_2 are weighting parameters that balance the training of the PME-Net and IDDB. In this way, the estimated posterior mean distribution is jointly optimized in dual domains, enabling adaptive alignment between the projection domain representation and the image domain generative prior.

III. Results

A. Experimental Set-up

We evaluate our method on "2016 NIH-AAPM-Mayo Clinic Low-Dose CT Grand Challenge" dataset, which contains 5936 1mm thickness slices from 10 patients (8 for training, 1 for validation and 1 for testing). Scanning is simulated under an equidistant fan-beam configuration. The distances from source to detector and from source to rotation center are set to 1240.00 mm and 720.00 mm, respectively. There are 960 detector elements with a size of 1.00 mm. All the slices have the same pixel size of 0.70 mm $\times$ 0.70 mm. The full-view projection is collected from 720 views over an angular range of 2π, and we uniformly extract 90 views for sparse-view. Poisson noise and electronic noise are added to obtain the simulated noisy projection. The incident photon number is set to $I_0 = 2\times10^5$, and the variance of electronic noise is fixed at $\sigma_e^2 = 10$. Peak signal to noise ratio (PSNR), structural similarity index (SSIM) and learned perceptual image patch similarity (LPIPS) are used for quantitative evaluation.

We employ four methods as baselines for evaluation: (i) the conventional FBP reconstruction, (ii) an image-domain CNN network, FBPConvNet [1], (iii) an image-domain diffusion bridge I^{2}SB [5], (iv) an image-domain diffusion bridge model with projection data consistency, CDDB [7]. All the methods are reproduced according to their original implementation. For diffusion bridge sampling, the number of function evaluations (NFE) is set to 10 to balance fidelity and perceptual quality.

B. Experimental Results

Two representative slices from the test patient are displayed in Fig. 2. The FBP reconstruction suffers from severe streaking artifacts. Although FBPConvNet alleviates artifacts, it leads to excessive smoothing and the loss of fine texture details. Conversely, I^{2}SB preserves more texture details, but some of them are inconsistent with the ground truth structures. Besides, CDDB achieves more accurate structural recovery, yet yields noticeable noise. Remarkably, PMG-DB is able to recover fine structures (see the red arrows in Fig. 2) that are absent in other methods, producing visually superior results replete with the least noise and richer texture details. The quantitative metrics among all test slices are listed in Table II, where PMG- achieves the highest PSNR and SSIM and ranks the second in LPIPS. The inferior LPIPS compared to CDDB is mainly due to the noise introduced by CDDB from noise projection, which may potentially leads to better perceptual metrics.

TABLE I
QUANTITATIVE METRICS AMONG ALL TEST SLICES

Metric	PSNR↑	SSIM↑	LPIPS↓
FBP	25.67	0.4204	0.4577
FBPConvNet	39.94	0.9448	0.0709
I^{2}SB	38.02	0.9161	0.0662
CDDB	37.13	0.8865	**0.0557**
PMG-DB (Ours)	**40.59**	**0.9498**	0.0576

We further evaluate the methods on the same test patient on a higher noise level (corresponding to $I_0 = 10^5$) to assess the generalization capability of the models. The same slice is shown in Fig. 3 and the quantitative metrics are listed in Table II. FBPConvNet sill introduces blurring across the entire image and fails to completely suppress artifacts (see the red arrows in Fig. 3). I^{2}SB generates more inconsistent structures, while CDDB suffers from amplified noise, which significantly degrade reconstruction fidelity. In contrast, PMG-DB consistently achieves superior reconstruction performance and the best metrics, demonstrating stronger robustness on a high noise level.

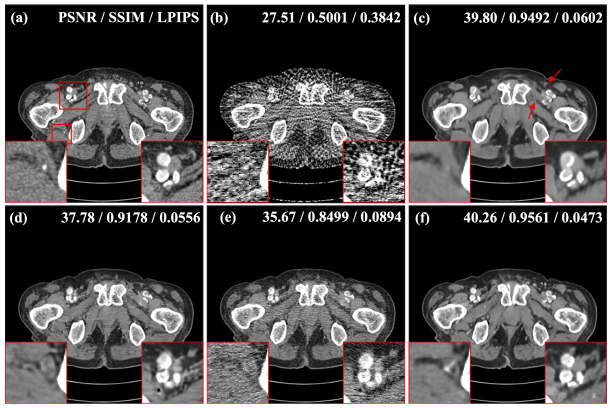


Fig. 3. Qualitative comparison among different methods on a higher noise level ($I_0 = 10^5$): (a) Ground Truth, (b) FBP, (c) FBPConvNet, (d) I^{2}SB, (e) CDDB, (f) PMG-DB. The display window is [-160, 240] HU.

TABLE II
QUANTITATIVE METRICS AMONG ALL TEST SLICES ON A HIGH NOISE LEVEL ($I_0 = 10^5$)

Metric	PSNR↑	SSIM↑	LPIPS↓
FBP	23.97	0.3269	0.5247
FBPConvNet	39.58	0.9420	0.0730
I^{2}SB	37.72	0.9082	0.0651
CDDB	35.71	0.8416	0.0879
PMG-DB (Ours)	**40.02**	**0.9459**	**0.0568**

IV. Discussion

Sparse-view CT reconstruction is an ill-posed problem induced by angular undersampling. Conventional FBP reconstruction inevitably suffers from severe streaking artifacts. Although CNN-based image-domain networks like FBPConvNet effectively suppress most artifacts, they produce blurry results and compromise texture details. Diffusion bridge models are capable of preserving texture details by modeling clean image distribution. However, the distribution-driven nature also makes them generate structures deviate from the ground truth. Posterior sampling is a principled way to improve reconstruction fidelity by applying data consistency constraints, but straightforward consistency constraints such as that adopted in CDDB are highly sensitive to measurement noise, which can amplify inconsistencies and destroy reconstruction stability on the contrary. Our proposed PMG-DB mitigates this issue by fully leveraging the projection data to construct a noise-robust posterior mean distribution. Furthermore, such a distribution provides a more informative starting point for diffusion bridge sampling, which helps to alleviate distortion-perception tradeoff and allows reconstruction of high quality. We will present more results under different sparse-view sampling rate and on various datasets in the conference.

V. Conclusion

We propose a novel pipline named PMG-DB for sparse-view CT reconstruction. PMG-DB leverages posterior mean from projection estimation as an initialization and performs iterative sampling to refine the reconstruction. Experimental results on low-dose sparse-view CT scans demonstrate that our method achieves high fidelity and improved perceptual quality. To further enhance the initial distribution through diffusion-priors during training phase, future work will explore more principled loss design within the end-to-end training framework.

References

[1] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4509–4522, 2017.

[2] H. Lee, J. Lee, and S. Cho, "View-interpolation of sparsely sampled sinogram using convolutional neural network," in *Medical Imaging 2017: Image Processing*, vol. 10133. SPIE, 2017, pp. 617–624.

[3] W. Wu, D. Hu, C. Niu, H. Yu, V. Vardhanabhuti, and G. Wang, "Drone: Dual-domain residual-based optimization network for sparse-view ct reconstruction," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3002–3014, 2021.

[4] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[5] G.-H. Liu, A. Vahdat, D.-A. Huang, E. A. Theodorou, W. Nie, and A. Anandkumar, "I 2̂ sb: Image-to-image schr\" odinger bridge," *arXiv preprint arXiv:2302.05872*, 2023.

[6] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 6228–6237.

[7] H. Chung, J. Kim, and J. C. Ye, "Direct diffusion bridge using data consistency for inverse problems," *Advances in Neural Information Processing Systems*, vol. 36, pp. 7158–7169, 2023.

[8] A. Hyvärinen and P. Dayan, "Estimation of non-normalized statistical models by score matching." *Journal of Machine Learning Research*, vol. 6, no. 4, 2005.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

Causal Statistical Foundations and Causal Invariance Learning for Material Basis Estimations from Single-kV CT Images

Zilin Jiang, Ran Zhang, Yijing Wu, Shujie Jin, Zhihua Qi, Ke Li, and Guang-Hong Chen

***Abstract*—In this work, we establish a causal–statistical foundation for material inference from single-energy CT and propose a causal invariance learning framework that enables estimation of iodine and water basis images from single-kV CT data. We show that conditioning on tissue or organ class induces population-level statistical dependence between low- and high-kV CT measurements via collider-induced correlation. This prediction is validated using clinical dual-energy abdominal CT data from 84 subjects, which reveal strong, organ-specific correlations between low- and high-kV CT numbers.**

Building on this foundation, we introduce an energy-agnostic deep learning architecture that combines a transformer backbone, U-Net refinement, and multi-environment causal invariance training. The model is trained on virtual monoenergetic images spanning 40–200 keV at 10 keV increments and evaluated on a dense 1 keV grid. The results demonstrate accurate, energy-invariant material decomposition across unseen energies, establishing a principled framework for single-energy material inference in CT.

***Index Terms*—Dual-energy CT, material decomposition, causal invariance, transformer, physics-consistency, foundation model.**

I. Introduction

FROM a measurement–theoretic viewpoint, a computed tomography (CT) scanner probes the x–ray attenuation of an unknown object. Under two distinct spectra, commonly referred to as low–kV and high–kV acquisitions, each voxel yields two CT numbers. If the object is allowed to span the full space of physically admissible attenuation behaviors, there is no reason *a priori* to expect statistical dependence between these two measurements.

This statistical independence underpins the technological foundation of dual–energy CT (DECT). Two independent spectral measurements are required to estimate two independent material basis images, such as effective electron density and effective atomic number, or any equivalent linear material representation. These material basis images provide physically meaningful image features that support material classification and quantitative tissue characterization.

The above argument implicitly assumes that the object may span the full space of physically admissible attenuation behaviors—an assumption that is overly broad in clinical CT. Medical imaging is constrained by strong prior knowledge, at minimum that the object is a human body composed of a limited set of constituent elements. More restrictively, voxels often belong to specific organs or tissue classes, such as liver, kidney, or muscle. Conditioning on such class information restricts the admissible combinations of attenuation coefficients and therefore alters the joint measurement statistics. As a result, although low-kV and high-kV acquisitions remain physically independent, they are not statistically independent once conditioned on tissue class—an effect formalized within causal statistical inference, as developed by Pearl [1], [2].

In a causal graphical representation, the low–kV and high–kV CT measurements act as independent parent variables whose joint influence determines underlying material properties. When no conditioning is applied, the two measurements are statistically independent. However, conditioning on the material class—or more generally on a variable derived from the material properties—induces statistical dependence between the two measurements. This phenomenon, known as *collider–induced correlation*, is a fundamental result in modern causal statistics [1], [2].

A practical implication is that, within a fixed organ class, low–kV and high–kV CT numbers exhibit predictable population–level correlations. While not implying a deterministic voxel–wise mapping, the conditional distribution is structured enough that one measurement can inform the expected value of the other in an organ–conditional sense. This provides a statistical pathway for single–energy material inference: patient–specific single–kV measurements can be combined with population–level organ statistics to infer dual–energy information conditionally, enabling material basis estimation even when only one spectrum is acquired.

II. Collider-Induced Correlations Between Low-kV and High-kV CT Numbers: Experimental Validation

The preceding causal analysis predicts that conditioning on organ or tissue class induces statistical dependence between low-kV and high-kV CT numbers. Because this prediction departs from the conventional physics-only understanding and has rarely been examined explicitly in CT, we retrospectively validated it using clinical dual-energy abdominal CT data.

We initially identified 100 contrast-enhanced dual-energy examinations and excluded cases with substantial metal artifacts, severe motion, or incomplete acquisitions, yielding 84

Z. Jiang, S. Jin, Y. Wu, R. Zhang, K. Li, and G.-H. Chen are with the Departments of Medical Physics and Radiology, University of Wisconsin–Madison, Madison, WI, USA.

Z. Qi is with the Department of Radiology, Henry Ford Health, Detroit, MI, USA.

K. Li is also with the Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

Correspondence: Guang-Hong Chen (e-mail: gchen7@wisc.edu).

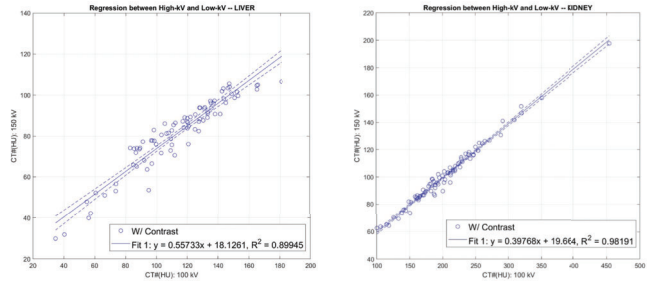


Fig. 1. Population-level relationship between low-kV and high-kV CT numbers measured from contrast-enhanced abdominal dual-energy CT scans. Paired CT numbers were extracted from organ-specific regions of interest in liver and kidney across 84 clinical cases acquired using a Siemens dual-source CT system. Each point represents a paired measurement from a single ROI. Within each organ class, the paired measurements are well described by an affine model, indicating a strong, organ-dependent statistical correlation between low- and high-kV CT numbers at the population level.

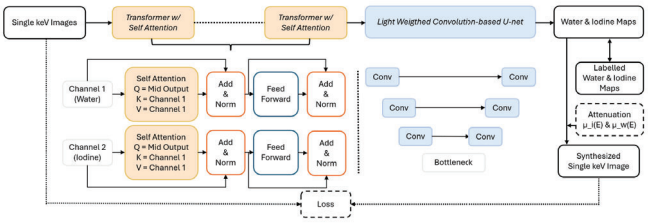


Fig. 2. Overview of the proposed deep learning framework for material decomposition from single-energy CT. Single-energy CT images are provided as input to a transformer-based backbone that learns organ-conditioned latent representations. A lightweight U-Net refinement module suppresses structured artifacts associated with attention-based features. During training, a physics-consistency pathway synthesizes the input single-energy CT images from the predicted material basis images using externally supplied, energy-dependent attenuation coefficients, thereby enforcing physically valid decomposition and promoting invariance across energy environments.

eligible studies. All scans were acquired on a Siemens dual-source system, with approximately 300–500 axial slices per examination.

For each case, we extracted paired low-kV and high-kV CT numbers from randomly selected ROIs from each organ. We focus on liver and kidney across all 84 subjects and aggregate paired measurements across the cohort. Fig. 1 shows the resulting scatter plots.

As shown in Fig. 1, within each organ class the paired CT numbers are well described by an affine model with organ-specific slope and intercept. Thus, at the population level, the high-kV CT number is highly predictable from the low-kV CT number, and vice versa, once tissue class is specified. Moreover, regression parameters differ systematically between liver and kidney, consistent with collider-induced dependence: conditioning on different tissue classes imposes different constraints on admissible attenuation properties and therefore induces distinct correlation structures. These findings provide direct experimental support for the central hypothesis and motivate single-energy inference methods that exploit organ-conditioned measurement structure.

III. Technical Implementation: Cascaded Transformer and U-Net Architecture

In this work, we adopt a substantially different architectural and learning strategy from the Deep-En-Chroma framework proposed by Li *et al.* [3]. In Deep-En-Chroma, a U-Net backbone is coupled with multiple physics-informed subnetworks, which encode physics constraints during training to enable the material decomposition of single-kV CT images. While effective, this approach relies on explicit forward-model constraints and task-specific physics modules.

Here, we propose a simpler, statistically grounded architecture: a transformer backbone followed by a lightweight U-Net refinement module. Rather than hard-coding physics into the network, we enforce *causal invariance learning* (Section IV) to promote energy-agnostic material estimation by leveraging organ-conditioned population structure.

The transformer models long-range dependencies and learns energy-agnostic latent representations that capture organ-specific attenuation behavior, effectively absorbing the population-level coupling between low-kV and high-kV measurements demonstrated above. The subsequent U-Net refiner suppresses structured artifacts and local inconsistencies that can arise from attention-based representations, combining global modeling with local recovery at modest complexity.

An overview of the proposed deep learning framework, including the transformer backbone, refinement module, and physics-consistency pathway, is shown in Fig. 2.

IV. Causal Invariance Learning

A central challenge in single-energy material estimation is preventing the model from encoding input-energy footprints into the predicted material basis images. In classical dual-energy CT (DECT), energy dependence is removed by jointly exploiting low-kV and high-kV measurements, yielding material basis images that are independent of the acquisition spectrum. With only a single-energy input, this energy-removal mechanism is no longer explicit, and naive learning strategies may produce energy-dependent surrogate outputs, leading to poor generalization across acquisition protocols and effective energies.

This limitation reflects the underlying causal structure of DECT: low-kV and high-kV measurements act as independent causes whose joint influence determines the material basis images. **Energy invariance of the basis images is therefore not merely a desirable property but a direct consequence of the DECT causal mechanism**. To emulate this property using single-energy inputs, the training procedure must explicitly enforce invariance of the predicted material maps to variations in input effective energy.

We achieve this through *multi-environment causal invariance learning*, a conceptually simple framework for invariant prediction [4], [5]. The core idea is to construct multiple environments by intervening on nuisance factors that affect the input but are not causal for the target, and to train the predictor to recover the relationship that remains stable across these environments.

In the CT setting, we treat the input effective energy as the environment variable. Using paired training data in

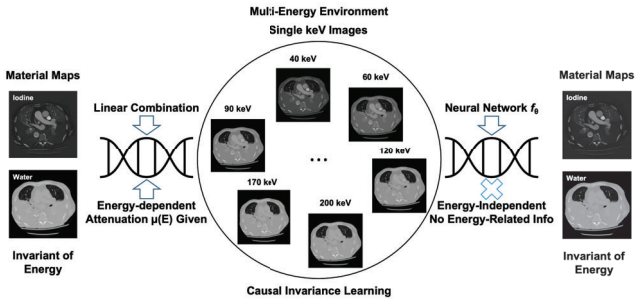


Fig. 3. Causal invariance learning for material decomposition from single-energy CT. Multiple energy environments are generated by synthesizing virtual monoenergetic images from fixed material basis maps. Joint training across these environments enforces invariance of the predicted material basis images to input energy, emulating the energy-removal mechanism of conventional dual-energy CT.

which material basis images are obtained from DECT, we generate additional virtual monoenergetic images (VMIs) by synthesizing images at different energies from fixed basis maps using known energy-dependent attenuation coefficients. These VMIs constitute controlled interventions on the input-energy variable while preserving the underlying material composition.

The resulting multi-environment dataset varies the input energy while keeping the iodine and water basis images fixed. As illustrated in Fig. 3, the model is trained jointly across these environments so that the predicted basis maps are invariant to changes in input energy. In effect, the network learns to remove the encoded energy dependence from single-energy CT inputs, enabling robust estimation of energy-independent material basis images from single-energy data.

V. Dataset and Experimental Design

We constructed paired training data consisting of a single-energy VMI and its corresponding iodine and water basis maps. In total, 272,427 VMI–material pairs were curated on a *coarse* energy grid spanning 40–200 keV in 10 keV increments (i.e., $\{40, 50, 60, \ldots, 200\}$ keV). Among these data, 205,411 pairs were used for training and 43,012 were reserved for validation, with the remaining 24,004 pairs held out for testing on the same coarse 10 keV grid.

To rigorously assess generalization across energy environments not explicitly seen during training, we further constructed a *dense* inference dataset by resampling the held-out cohort on a 1 keV energy grid spanning 40–200 keV (i.e., $\{40, 41, 42, \ldots, 200\}$ keV). The held-out cohort comprises 12 patients with a total of 1,412 axial slices. Generating VMIs at 161 integer energy levels for each slice yields 227,332 test instances for inference. This evaluation protocol deliberately decouples the energy resolution used during optimization (10 keV spacing) from that used during assessment (1 keV spacing), thereby ensuring that the model is evaluated at a large number of effective energies not explicitly represented during training.

This design directly reflects the principles of causal invariance learning. Specifically, the effective energy of the input VMI is treated as an environmental variable that is intervened upon during training and testing, while the underlying material basis images are regarded as the invariant causal targets. Training on a discrete set of energy environments and testing on a much denser and previously unseen set of environments provides a stringent test of whether the learned mapping captures energy-invariant material properties rather than overfitting to energy-specific correlations present in the training data.

Model performance is evaluated for both iodine and water basis maps using mean absolute error (MAE), root mean squared error (RMSE), and the structural similarity index measure (SSIM), with iodine concentration reported in units of mg/mL. To assess physics consistency, a synthesized VMI is generated from the predicted basis maps using known energy-dependent attenuation coefficients and compared with the corresponding input true VMI using MAE, RMSE (in HU), and SSIM. All metrics are computed independently at each energy level and subsequently averaged across the test cohort to produce energy-resolved performance curves over the full 40–200 keV range.

Together, this experimental design enforces and validates causal invariance by explicitly separating variations in the input energy environment from the material basis images to be estimated. Successful performance across the dense energy grid therefore provides direct evidence that the proposed framework learns energy-invariant material representations rather than energy-dependent surrogate mappings.

VI. Results

The results are presented to explicitly evaluate whether the proposed causal invariance learning framework produces material basis images that are independent of the effective energy of the input single-energy CT image.

A. Overall Quantitative Performance

Across the training-energy grid, the proposed model produces iodine and water basis maps with strong quantitative agreement to the reference labels. To assess robustness to spectral variation, we evaluate the model over a dense 1 keV grid spanning 40–200 keV, which includes many energy levels not explicitly represented during training. Table I summarizes overall performance on the dense inference set by reporting median $[Q_1, Q_3]$ MAE, RMSE, and SSIM aggregated across all test slices and all input energies from 40 to 200 keV.

B. Causal Invariance Across Energy Environments

In addition to the quantitative sweep, we report a representative qualitative showcase in Fig. 4 to illustrate causal invariance at the image level. Specifically, Fig. 4 shows water and iodine maps for the reference basis images and for predictions obtained from VMI inputs at 48 keV, 137 keV, and 198 keV, as well as from a clinical 140 kVp input image. The 140 kVp image is a true polychromatic clinical reconstruction rather than a VMI, and it is included to demonstrate model behavior under routine acquisition conditions.

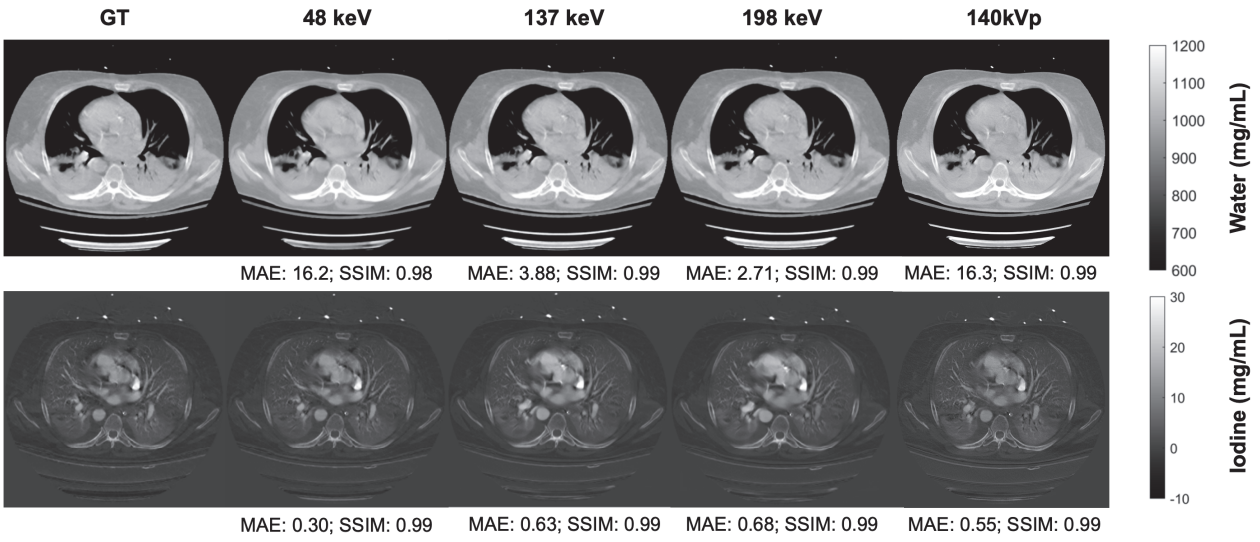


Fig. 4. Qualitative demonstration of causal invariance for a representative test slice. Top row: water maps; bottom row: iodine maps. The first column shows reference basis images, followed by predictions from single-energy inputs at 48, 137, and 198 keV, and from a clinical 140 kVp reconstruction (polychromatic, not a VMI). Display windows are fixed within each row for direct comparison across input energies.

TABLE I
OVERALL EVALUATION METRICS ACROSS INFERENCE SET AMONG INPUT ENERGIES FROM 40 TO 200 KEV. VALUES ARE REPORTED AS MEDIAN $[Q_1, Q_3]$ OVER ALL PATIENT SLICES. UNITS ARE MG/ML FOR IODINE AND WATER MAPS, HU FOR SYNTHESIZED VMI, AND UNITLESS FOR SSIM.

Metric	Water map	Iodine map	Synth. VMI
MAE	4.34 [3.23, 6.89]	0.38 [0.31, 0.48]	3.76 [2.83, 5.66]
RMSE	7.81 [5.51,12.75]	0.74 [0.54, 1.03]	6.84 [4.81,11.49]
SSIM	0.99 [0.99, 0.99]	0.98 [0.96, 0.99]	0.99 [0.99, 0.99]

Despite substantial differences in contrast and noise characteristics across these inputs, the predicted iodine and water maps remain visually consistent in spatial distribution and quantitative scale. For each input condition, we further report the corresponding per-image error summary by listing the MAE and SSIM beneath each panel in Fig. 4. Together, these results provide qualitative evidence that the learned material decomposition remains stable across energy environments and generalizes beyond the discrete energy settings observed during training, consistent with the causal invariance objective.

VII. CONCLUSION

We proposed an energy-agnostic material decomposition framework that enables estimation of iodine and water basis images from single-energy CT by explicitly enforcing causal invariance with respect to acquisition energy. Instead of relying on dual-energy measurements or energy-encoded inputs, the method leverages population-level, organ-conditioned statistical structure and multi-environment causal invariance learning to recover energy-independent material representations.

The framework combines a transformer-based backbone to capture long-range dependencies and disentangle spatial and spectral information with a lightweight U-Net refinement module to improve local consistency. A physics-consistency constraint, implemented by re-synthesizing the input image from the predicted material basis images using known energy-dependent attenuation coefficients, further restricts the solution space to physically interpretable decompositions and stabilizes training across energy environments.

Experiments demonstrate accurate and structurally consistent iodine and water maps across a wide range of input energies. Training on a sparse energy grid (10 keV spacing) and evaluation on a dense, unseen grid (1 keV spacing) confirm that the predicted material basis images remain stable with respect to input energy, providing direct empirical validation of the causal invariance objective.

More broadly, this work establishes a general statistical and computational paradigm for single-energy material inference in CT. By separating environmental variation (acquisition energy) from invariant causal targets (material basis images), the proposed framework provides a principled foundation for extending single-energy CT to broader spectral inference tasks and for efficient adaptation when limited dual-energy or material-specific data are available.

REFERENCES

[1] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. Cambridge, UK: Cambridge University Press, 2009.
[2] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*. New York, NY: Basic Books, 2018.
[3] Y. Li *et al.*, "A quality-checked and physics-constrained deep learning method to estimate material basis images from single-kV contrast-enhanced chest CT scans," *Medical Physics*, vol. 50, no. 6, pp. 3368–3388, 2023.
[4] J. Peters, P. Bühlmann, and N. Meinshausen, "Causal Inference by Using Invariant Prediction: Identification and Confidence Intervals," *J. R. Stat. Soc. B*, vol. 78, no. 5, pp. 947–1012, Nov. 2016, doi: 10.1111/rssb.12167.
[5] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, "Invariant Risk Minimization," arXiv:1907.02893, 2019.

Data-Driven Linearization of the Spectral CT Forward Model using Gaussian Mixture Models

Gauthier Multari, Guillaume Delorme, Frédéric Jolivet and Simon Rit

***Abstract*—Accurate material characterization in spectral CT requires modeling the polychromatic source, which inherently creates a non-linear inverse problem which is computationally expensive to solve. While linearizing the forward model can accelerate reconstruction, global approximations often introduce significant bias in material decomposition. In this work, we propose a framework to build a quantitatively accurate linear forward model by blending local linearizations. We approximate the joint distribution of material lengths and attenuation data using a Gaussian Mixture Model (GMM) fitted on a calibration basis. We derive a local linear model for each projection ray as a mixture of local linear approximations, weighted by the likelihood of the observed data. This formulation preserves the convexity of the linearized inverse problem while capturing the non-linear physics of the system. We evaluate the method on simulated dual-energy CT data of a multi material cylinder and a thorax phantom. Results demonstrate that the proposed GMM-based approach significantly reduces the bias compared to global linearization methods. It achieves material decomposition accuracy comparable to the computationally intensive complete physical model, while retaining the speed and stability of linear reconstruction frameworks.**

I. Introduction

Conventional Computed Tomography (CT) operates on the simplifying assumption of a monoenergetic X-ray source. This allows the attenuation process to be described by a linear forward model based on the Beer-Lambert law. While computationally efficient, this model overlooks the inherently polychromatic nature of X-ray sources, limiting the quantitative accuracy of the resulting images and leading to beam hardening artefacts in high attenuation scenarios.

In contrast, spectral CT leverages the full source spectrum to enable material-specific characterization [1] and extract quantitative information from the data. However, this physical accuracy comes at a cost: the forward model becomes nonlinear. The computational burden of solving this nonlinear problem is a significant challenge.

To mitigate this, previous works [2,3] proposed linearizing the forward model using a calibration basis. This yields a simpler objective function and faster model evaluation. However, because this linearization is global, it can introduce significant material decomposition bias [4]. While methods exist to correct these estimates [4,5], they rely on rectifying an initially biased solution.

In this work, we propose an alternative approach: rather than correcting a biased global linear approximation, we minimize linearization error directly at the calibration stage. To do so, we model the forward problem as a mixture of local linearizations, derived from a Gaussian Mixture Model (GMM) [6] fitted to a calibration basis. During reconstruction, we use the acquired data to obtain a local linear model. This approach improves quantitative accuracy while retaining the computational advantages of a linear model.

G. Multari is with THALES AVS, Moirans, France and with INSA-Lyon, Universite Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69373, Lyon, France (e-mail: gauthier.multari@thalesgroup.com).

G. Delorme and F. Jolivet are with THALES AVS, Moirans, France.

S. Rit is with INSA-Lyon, Universite Claude Bernard Lyon 1, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69373, Lyon, France and is a Royal Society Wolfson Visiting Fellow at the University College of London funded by the Wolfson Foundation and Royal Society. (e-mail: simon.rit@creatis.insa-lyon.fr).

This work is supported by the French National Association for Technical Research (ANRT). It was performed within the framework of the SIRIC LYriCAN+ (INCa-DGOS-INSERM-ITMO cancer 18003) and the LABEX PRIMES (ANR-11-LABX-0063) of Université de Lyon, within the program "Investissements d'Avenir" (ANR- 11-IDEX-0007) operated by the ANR.

Manuscript received January 28, 2026.

II. Materials & Methods

A. Forward Model

We consider the problem of dual-energy CT reconstruction, where the object is decomposed into $M = 2$ basis functions (e.g., water/bone or photoelectric/Compton [1]).

Let N_ν be the total number of voxels in the reconstruction volume and N_p the number of projection rays (defined as the product of detector pixels and acquisition angles). We define the unknown spatial map of the m-th material as $\mathbf{x}_m \in \mathbb{R}^{N_\nu}$ and let $\mathbf{x} = \{\mathbf{x}_m\} \in \mathbb{R}^{MN_\nu}$ denote the vector of stacked material maps.

The acquisition is performed across two spectral channels, denoted by the index $b \in \{\text{LE}, \text{HE}\}$ (Low-Energy and High-Energy). In this work, we assume that the acquisition geometry is identical for both spectral channels. Therefore, the geometric forward model relies on a common system matrix $\mathbf{A} \in \mathbb{R}^{N_p \times N_\nu}$, which acts as a discrete projection operator. The element $[\mathbf{A}\mathbf{x}_m]_i$ represents the line integral of material m along the i-th ray ($i \in \{1, \dots, N_p\}$). Consequently, the vector of projected material lengths is given by $\boldsymbol{\ell}_m = \mathbf{A}\mathbf{x}_m \in \mathbb{R}^{N_p}$.

The classical physics-based model discretizes the polyenergetic Beer-Lambert law. The expected photon count $\bar{I}_{b,i}$ for a spectral energy bin b and ray i is modeled as

$$\bar{I}_{b,i} = \sum_E S_{b,i}(E) \exp\left(- \sum_{m=1}^{M} \mu_m(E) \ell_{m,i} \right), \qquad (1)$$

where $S_{b,i}(E)$ is the effective spectral response and $\mu_m(E)$ is the energy-dependent attenuation coefficient for material m. The resulting log-transformed attenuation is

$$d_{b,i} = -\log\left(\frac{\bar{I}_{b,i}}{I^0_{b,i}} \right) \quad \text{with} \quad I^0_{b,i} = \sum_E S_{b,i}(E). \qquad (2)$$

This conventional physical forward model is non-linear with respect to the material projections $\ell_{m,i}$, rendering the resulting optimization problem non-strictly convex and computationally expensive. To facilitate efficient convex optimization, we consider the approach proposed by [4], which approximates the physics-based model with an empirical linear forward model learned on a calibration basis.

We formulate the full 2-channel linear forward model for ray i in matrix form:

$$m_i^{\text{LIN}}(\mathbf{x}) = \underbrace{\begin{pmatrix} c_{1,\text{LE},i} & c_{2,\text{LE},i} \\ c_{1,\text{HE},i} & c_{2,\text{HE},i} \end{pmatrix}}_{\mathbf{P}_i \in \mathbb{R}^{2\times 2}} \underbrace{\begin{pmatrix} [\mathbf{A}\mathbf{x}_1]_i \\ [\mathbf{A}\mathbf{x}_2]_i \end{pmatrix}}_{\boldsymbol{\ell}_i \in \mathbb{R}^2} + \underbrace{\begin{pmatrix} c_{0,\text{LE},i} \\ c_{0,\text{HE},i} \end{pmatrix}}_{\boldsymbol{c}_i \in \mathbb{R}^2}. \quad (3)$$

Note that the offset coefficients $\boldsymbol{c}_i$ are fixed to zero in the conventional linear framework [3–5].

The acquired data vector $\mathbf{d}_i = (d_{\text{LE},i}, d_{\text{HE},i})^T$ is thus expressed as:

$$\mathbf{d}_i = m_i^{\text{LIN}}(\mathbf{x}) + \boldsymbol{\varepsilon}_i^{\text{LIN}} = \mathbf{P}_i\, \boldsymbol{\ell}_i + \boldsymbol{c}_i + \boldsymbol{\varepsilon}_i^{\text{LIN}}, \quad (4)$$

where $\boldsymbol{\varepsilon}_i^{\text{LIN}} \in \mathbb{R}^2$ is the error vector accounting for the measurement noise and the linearization error.

However, this conventional linear model leads to significant significant material decomposition bias [4]. To capture the non-linearity of the forward model while retaining the computational benefits of a linear model, we propose a mixture of local linear approximations.

To automatically find the right mixture for each ray measurement, we model the joint distribution of material lengths $\boldsymbol{\ell}_i$ and data $\mathbf{d}_i$ using a Gaussian Mixture Model (GMM).

B. Calibration Step

To simplify notation, the ray index i is omitted in Sec. II-B and Sec. II-C.

In order to fit the GMM, we use a calibration basis $\boldsymbol{\xi}=\{\boldsymbol{\xi}_n\} \in \mathbb{R}^{N_c\times 4}$ of N_c pairs of known material lengths $(\ell_{1,n}^{\text{calib}}, \ell_{2,n}^{\text{calib}})$ and their corresponding measured attenuation $(d_{\text{LE},n}^{\text{calib}}, d_{\text{HE},n}^{\text{calib}})$. We write the mixture with K components for a given $\boldsymbol{\xi}_n \in \mathbb{R}^4$ as:

$$p(\boldsymbol{\xi}_n|\boldsymbol{\pi},\boldsymbol{\nu},\boldsymbol{\Sigma}) = \sum_{k=1}^{K} p(k)\, p(\boldsymbol{\xi}_n|k) = \sum_{k=1}^{K} \pi_k\, \mathcal{N}(\boldsymbol{\xi}_n|\boldsymbol{\nu}_k, \boldsymbol{\Sigma}_k), \quad (5)$$

where each Gaussian has its own mean

$$\boldsymbol{\nu}_k = (\nu_{\ell_1,k}, \nu_{\ell_2,k}, \nu_{d_{\text{LE}},k}, \nu_{d_{\text{HE}},k})^T = \begin{pmatrix} \boldsymbol{\nu}_{\boldsymbol{\ell},k} \\ \boldsymbol{\nu}_{\mathbf{d},k} \end{pmatrix} \in \mathbb{R}^4$$

and full covariance

$$\boldsymbol{\Sigma}_k = \begin{array}{c} \\ \ell_1 \\ \ell_2 \\ d_{\text{LE}} \\ d_{\text{HE}} \end{array} \begin{array}{c} \begin{array}{cc} \ell_1 \quad & \ell_2 \end{array} \quad \begin{array}{cc} d_{\text{LE}} \quad & d_{\text{HE}} \end{array} \\ \begin{pmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\ell},k} & \boldsymbol{\Sigma}_{\boldsymbol{\ell}\mathbf{d},k} \\ \boldsymbol{\Sigma}_{\mathbf{d}\boldsymbol{\ell},k} & \boldsymbol{\Sigma}_{\mathbf{d},k} \end{pmatrix} \end{array} \in \mathbb{R}^{4\times 4}.$$

The parameters of the mixture ($\boldsymbol{\pi}=\{\pi_k\}, \boldsymbol{\nu}=\{\boldsymbol{\nu}_k\}, \boldsymbol{\Sigma}=\{\boldsymbol{\Sigma}_k\}$) are learned to approximate the joint density $p(\boldsymbol{\ell}, \mathbf{d})$ by maximizing the log-likelihood of the calibration basis, given by

$$\ln p(\boldsymbol{\xi}|\boldsymbol{\pi},\boldsymbol{\nu},\boldsymbol{\Sigma}) = \sum_{n=1}^{N_c} \ln\left(\sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{\xi}_n|\boldsymbol{\nu}_k, \boldsymbol{\Sigma}_k)\right), \quad (6)$$

via expectation-maximization (EM).

C. Derivation of the models

We now consider the conditional distributions $p(\boldsymbol{\ell}|\mathbf{d})$ and $p(\mathbf{d}|\boldsymbol{\ell})$ to define estimators for $\boldsymbol{\ell}$ and $\mathbf{d}$. Both conditional distributions are mixtures of Gaussians themselves. It follows that their expectation can be computed in closed form as the weighted sum of the components' expectations [6]. Furthermore, a key property of multivariate Gaussians is that their conditional expectation is linear with respect to the conditioning variable. We therefore select the expectation as our estimator. We refer the reader to Sec. 2.3.2 of [6] for the derivation of the conditional distribution of a multivariate Gaussian.

a) Inverse model: Formally, we denote the conditional expectation for the k-th component as $\boldsymbol{\nu}_{\boldsymbol{\ell}|\mathbf{d},k} = \mathbb{E}[\boldsymbol{\ell}|\mathbf{d}, k]$. The inverse model is derived as the conditional expectation, which smoothly blends the local linear models:

$$\begin{aligned} \mathbb{E}[\boldsymbol{\ell}|\mathbf{d}] &= \sum_{k=1}^{K} p(k|\mathbf{d})\, \boldsymbol{\nu}_{\boldsymbol{\ell}|\mathbf{d},k} \\ &= \sum_{k=1}^{K} p(k|\mathbf{d})\, (\underbrace{\boldsymbol{\Sigma}_{\boldsymbol{\ell}\mathbf{d},k}\boldsymbol{\Sigma}_{\mathbf{d},k}^{-1}}_{\mathbf{Q}_k \in \mathbb{R}^{2\times 2}} (\mathbf{d} - \boldsymbol{\nu}_{\mathbf{d},k}) + \boldsymbol{\nu}_{\boldsymbol{\ell},k}) \\ &= \sum_{k=1}^{K} p(k|\mathbf{d}) (\mathbf{Q}_k\mathbf{d} \underbrace{-\mathbf{Q}_k\boldsymbol{\nu}_{\mathbf{d},k} + \boldsymbol{\nu}_{\boldsymbol{\ell},k}}_{\boldsymbol{c}_k' \in \mathbb{R}^2}), \end{aligned} \quad (7)$$

with

$$p(k|\mathbf{d}) = \frac{\pi_k \mathcal{N}(\mathbf{d}|\boldsymbol{\nu}_{\mathbf{d},k}, \boldsymbol{\Sigma}_{\mathbf{d},k})}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{d}|\boldsymbol{\nu}_{\mathbf{d},j}, \boldsymbol{\Sigma}_{\mathbf{d},j})}, \quad (8)$$

the likelihood, or responsibility, of component k given the data $\mathbf{d}$. Given that $\sum_{k=1}^{K} p(k|\mathbf{d}) = 1$, we can formulate the inverse estimator as

$$\hat{\boldsymbol{\ell}} = \widetilde{\mathbf{Q}}\mathbf{d} + \widetilde{\boldsymbol{c}}' + \boldsymbol{\varepsilon}', \quad (9)$$

with $\widetilde{\mathbf{Q}} = \sum_{k=1}^{K} p(k|\mathbf{d})\mathbf{Q}_k$ and $\widetilde{\boldsymbol{c}}' = \sum_{k=1}^{K} p(k|\mathbf{d})\boldsymbol{c}_k'$. This inverse model can be used for projection-based material decomposition.

b) Forward model: We are now looking for a linear forward model of the form of Eq. 4. The GMM provides the conditional expectation

$$\mathbb{E}[\mathbf{d}|\boldsymbol{\ell}] = \sum_{k=1}^{K} p(k|\boldsymbol{\ell})(\mathbf{P}_k\boldsymbol{\ell} + \boldsymbol{c}_k) \quad (10)$$

where $\mathbf{P}_k = \boldsymbol{\Sigma}_{\mathbf{d}\boldsymbol{\ell},k}\boldsymbol{\Sigma}_{\boldsymbol{\ell},k}^{-1}$ and $\boldsymbol{c}_k = \boldsymbol{\nu}_{\mathbf{d},k} - \mathbf{P}_k\boldsymbol{\nu}_{\boldsymbol{\ell},k}$.

However, the responsibilities $p(k|\boldsymbol{\ell})$ depend on the material lengths $\boldsymbol{\ell}$, which are unknown prior to reconstruction. To overcome this, we substitute $p(k|\boldsymbol{\ell})$ with the data responsibilities $p(k|\mathbf{d})$. While generally $p(k|\mathbf{d}) \neq p(k|\boldsymbol{\ell})$, this assumes that the measured data $\mathbf{d}$ is the best available information to compute the correct local linear region, allowing for a direct estimation without requiring a prior estimate of $\boldsymbol{\ell}$.

By substituting $p(k|\mathbf{d})$ for $p(k|\boldsymbol{\ell})$, we define our local linear forward model as

$$m^{\text{LIN}}(\mathbf{x}) = \sum_{k=1}^{K} p(k|\mathbf{d})(\mathbf{P}_k\boldsymbol{\ell} + \boldsymbol{c}_k) \tag{11}$$

As for the inverse model, this can be written in the simple linear form $m^{\text{LIN}}(\mathbf{x}) = \widetilde{\mathbf{P}}\boldsymbol{\ell} + \widetilde{\boldsymbol{c}}$, where the model parameters are now data-dependent. This approach provides a local, more accurate linearization of the system compared to a single global model.

Note that the relevance of the local linear approximation will decrease as the noise in the data increases. We model the acquisition noise as an additive, zero-mean Gaussian noise $\epsilon \sim \mathcal{N}(0, \boldsymbol{\Sigma}_0)$, where $\boldsymbol{\Sigma}_0$ is the (diagonal) noise covariance matrix. The marginal distribution of $\mathbf{d}$ for component k, as learned by the GMM, is $p(\mathbf{d}|k) = \mathcal{N}(\mathbf{d}|\boldsymbol{\nu}_{\mathbf{d},k}, \boldsymbol{\Sigma}_{\mathbf{d},k})$. The distribution of the observed (noisy) data is the convolution of the GMM marginal and the noise model. This results in a new Gaussian where the covariances are summed:

$$p(\mathbf{d}_{\text{noisy}}|k) = \mathcal{N}(\mathbf{d}|\boldsymbol{\nu}_{\mathbf{d},k}, \boldsymbol{\Sigma}_{\mathbf{d},k} + \boldsymbol{\Sigma}_0)$$

Therefore, a more robust responsibility computation which accounts for the acquisition noise, is

$$p(k|\mathbf{d}_{\text{noisy}}) = \frac{\pi_k \mathcal{N}(\mathbf{d}_{\text{noisy}}|\boldsymbol{\nu}_{\mathbf{d},k}, \boldsymbol{\Sigma}_{\mathbf{d},k} + \boldsymbol{\Sigma}_0)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{d}_{\text{noisy}}|\boldsymbol{\nu}_{\mathbf{d},j}, \boldsymbol{\Sigma}_{\mathbf{d},j} + \boldsymbol{\Sigma}_0)}. \tag{12}$$

D. Optimization problems

We consider the projection based material decomposition problem as well as one-step reconstruction. In both cases we assume the Gaussian Mixture Model parameters were learnt for each pixel of the acquisition system prior to the reconstruction.

a) Material decomposition: Let $\tilde{\mathbf{d}} \in \mathbb{R}^{N_p}$, such as $\tilde{\mathbf{d}}_i = \mathbf{d}_i - \widetilde{\boldsymbol{c}}_i$, be the vector of acquisition data with offset of the model. Assuming a Gaussian noise model and a linear forward model, the material decomposition problem in projection space reads

$$\hat{\boldsymbol{\ell}}_i \in \arg\min_{\boldsymbol{\ell}_i \in \mathbb{R}^2} \left\| m_i^{\text{LIN}}(\boldsymbol{\ell}_i) - \mathbf{d}_i \right\|_2^2 \tag{13}$$

$$= \arg\min_{\boldsymbol{\ell}_i} \left\| \widetilde{\mathbf{P}}_i \boldsymbol{\ell}_i - \tilde{\mathbf{d}}_i \right\|_2^2, \tag{14}$$

and admits a trivial closed form solution.

b) One-step reconstruction: The goal of one-step reconstruction is to directly estimate material-specific volumes $\mathbf{x} \in \mathbb{R}^{MN_\nu}$. This inverse problem can be formulated as the following variational optimization problem:

$$\hat{\mathbf{x}} \in \arg\min_{\mathbf{x} \in \mathbb{R}^{MN_\nu}} \mathcal{F}(\mathbf{x}) + \lambda \mathcal{R}(\mathbf{x}), \tag{15}$$

with $\mathcal{F}$ the data fidelity term, $\mathcal{R}$ a regularization term and λ the regularization parameter balancing the two terms.

When considering a Gaussian noise model and a linear forward model, the data fidelity term is

$$\mathcal{F}^{\text{LIN}}(\mathbf{x}) = \sum_{i=1}^{N_p} \left\| m_i^{\text{LIN}}(\mathbf{x}) - \mathbf{d}_i \right\|_2^2 \tag{16}$$

$$= \sum_{i=1}^{N_p} \left\| \widetilde{\mathbf{P}}_i \mathbf{A}_i \mathbf{x} - \tilde{\mathbf{d}}_i \right\|_2^2. \tag{17}$$

E. Numerical experiments

We simulated a calibration basis of PMMA (0–20 cm) and aluminum (0–5 cm) as well as a 120 kVp source filtered by 2 mm of aluminum and 0.4 mm of copper. The source spectra was simulated with SpekPy [7]. We fixed the number of Gaussian components to $K = 20$ for all pixels as it empirically showed the best results.

a) Material Decomposition: We compared our method against three baselines: (1) A global linear regression on the full calibration basis; (2) the corrected linear model proposed by Alvarez [5] and (3) the complete physical model (discretized at 1 keV). While all the methods based on a linear forward model yield a closed form solution to the material decomposition problem, the complete physical model requires an iterative resolution. We used a limited memory quasi-Newton algorithm (L-BFGS [8]) to solve it. For this experiment, data was simulated without noise to isolate the inherent error introduced by each approach. We evaluated 15 points along three lines in the space of pairs of material lengths: $\boldsymbol{\ell}_{\text{PMMA}} = 0$, $\boldsymbol{\ell}_{\text{Al}} = 0$ and $\boldsymbol{\ell}_{\text{PMMA}} = \boldsymbol{\ell}_{\text{Al}}$. The four final points along each trajectory lie outside the calibration basis to evaluate the extrapolation capabilities of each empirical model.

b) Image Reconstruction: We evaluated the different forward models in a one-step reconstruction setting. We simulated 620 evenly spaced projections around a 205° circular axis with a tube loading of 1.25 mAs on a 300x400 mm detector with pixels of 0.616x0.616 mm. Both eletronic and photonic noise were simulated.

We considered an 8 cm water cylinder phantom with 2 cm inserts of air, PMMA, aluminum and bone as well as an XCAT [9] thorax phantom. In both cases, the voxels were 1 mm^3.

We used the L-BFGS [8] algorithm with Huber regularization to reconstruct the volumes. Regularization was manually tuned to yield similar noise levels in all reconstructions. We did not evaluate results for the corrected linear model proposed by Alvarez [5] as it is not applicable to a one-step reconstruction scenario. The error was computed as the difference between the reconstructed maps and a theorical ground truth. The theorical ground truth was determined by solving a least-squares minimization between the attenuation coefficients of the known phantom materials and the basis' functions in the energy range of $[20 - 120]$ keV.

III. Results

a) Material Decomposition: Fig. 1 displays material decomposition results along the bold lines in material space. While all methods but the global linear model performed well inside the convex hull of the calibration basis, the corrected linear model performed poorly when extrapolating. The linearized GMM, however, gave significantly better results.

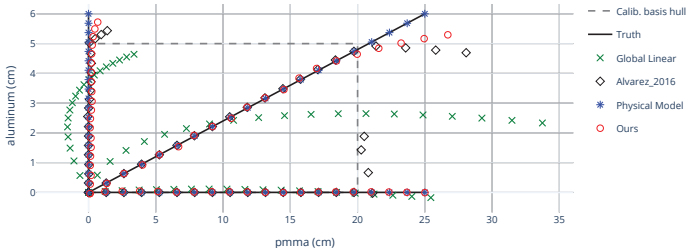


Fig. 1. Material decomposition results for 15 points along three material combination lines (in bold) of the basis functions without acquisition noise.

b) Image reconstruction: Fig. 2 and 3 show the one-step reconstruction results with the different forward models. The global linear model exhibited significant errors, most notably beam hardening artefacts in high attenuation scenarios. Our GMM-based approach efficiently reduced these errors, yielding reconstructed volumes and error maps close to the much more computationally expensive complete physical model.

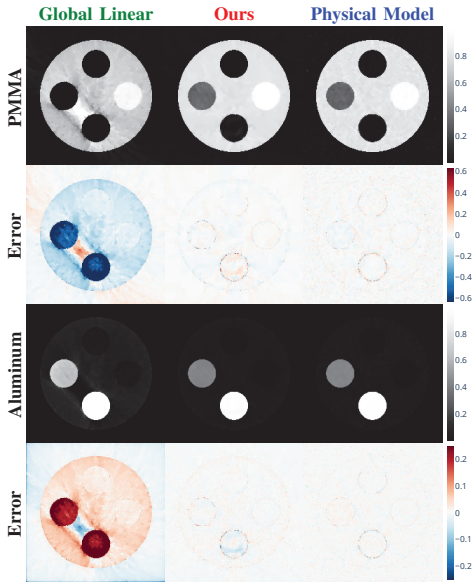


Fig. 2. One-step material reconstructions, in fraction of material, of a 8 cm water cylinder with 2 cm inserts of air (top), PMMA (right), aluminum (bottom) and bone (left) and corresponding error maps.

IV. Discussion

This GMM-based approach exhibits computation times comparable to a conventional linear approach. While the calibration step is more computationally intensive than a simple linear regression, it is trivially paralellizable and can be performed prior to the time sensitive reconstruction step. The only additional cost at inference time is a pre-computation step where the responsibilities $p(k|\mathbf{d}_i)$, $\widetilde{\mathbf{P}}_i$ and $\tilde{\mathbf{d}}_i$, are calculated for each pixel using measured data before reconstruction. The memory cost is therefore no higher than a conventional linear model per pixel and the additional computational cost is negligible in an iterative reconstruction scenario, making this approach feasible for medical and other time sensitive contexts.

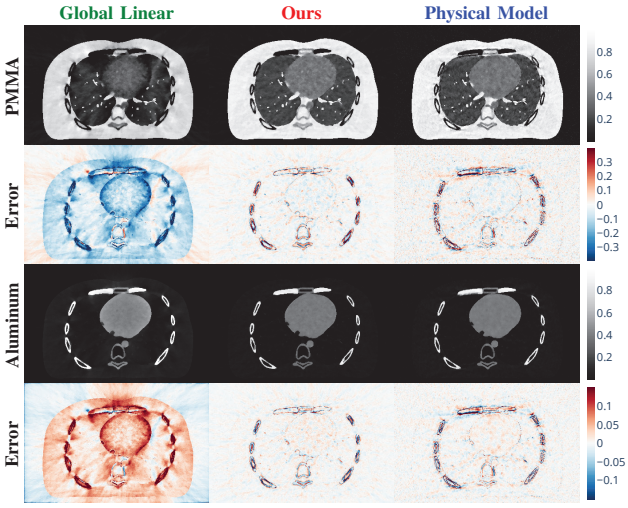


Fig. 3. One-step material reconstructions, in fraction of material, of a XCAT [9] thorax phantom and corresponding error maps.

V. Conclusion

We have presented a framework for constructing a quantitatively accurate linear forward model in spectral CT. By encoding the non-linear relationship between material lengths and attenuation data into a mixture of local linear models, our method strongly reduces the bias inherent in global linear approximations. The proposed GMM-based approach yields material decomposition results that closely match the accuracy of the much more computationally intensive physical model. Future works will extend this evaluation to varied dose levels, more spectral channels and experimental data to fully establish clinical applicability.

References

[1] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Physics in Medicine & Biology*, vol. 21, no. 5, p. 733, sep 1976. [Online]. Available: https://doi.org/10.1088/0031-9155/21/5/002

[2] K. C. Zimmerman and T. G. Schmidt, "Experimental comparison of empirical material decomposition methods for spectral CT," *Phys. Med. Biol.*, vol. 60, no. 8, pp. 3175–3191, 2015.

[3] F. Jolivet, J. Lesaint, C. Fournier, M. Garcin, and A. Brambilla, "An efficient one-step method for spectral ct based on an approximate linear model," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 4, pp. 528–536, 2021.

[4] R. E. Alvarez, "Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis," *Medical Physics*, vol. 38, no. 5, pp. 2324–2334, 2011. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3570658

[5] ——, "Efficient, non-iterative estimator for imaging contrast agents with spectral x-ray detectors," *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 1138–1146, 2016.

[6] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[7] G. Poludniowski, A. Omar, R. Bujila, and P. Andreo, "Technical note: Spekpy v2.0—a software toolkit for modeling x-ray tube spectra," *Medical Physics*, vol. 48, no. 7, pp. 3630–3637, 2021. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.14945

[8] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Mathematical Programming*, vol. 45, no. 1, pp. 503–528, 1989. [Online]. Available: https://doi.org/10.1007/BF01589116

[9] L. Dahal, M. Ghojoghnejad, L. Vancoillie, D. Ghosh, Y. Bhandari, D. Kim, F. C. Ho, F. I. Tushar, S. Luo, K. J. Lafata, E. Abadi, E. Samei, J. Y. Lo, and W. P. Segars, "Xcat 3.0: A comprehensive library of personalized digital twins derived from ct scans," *Medical Image Analysis*, vol. 103, p. 103636, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841525001835

The Sufficiency of Material Bases in Polychromatic CT: A Quantitative Effective Energy Perspective

Xiaopeng Yu, Ran Zhang, and Guang-Hong Chen

***Abstract*—Material decomposition in CT is fundamentally limited by the polychromatic nature of x-ray spectra, which gives rise to beam hardening and nonlinear attenuation behavior. Although the concept of effective energy is commonly used as a qualitative descriptor of polychromatic measurements, its rigorous definition, dependence on underlying physical factors, and role in quantitative CT modeling remain unclear. In this work, we establish a rigorous definition of effective energy for polychromatic measurements based on the mean value theorem for integrals. We analyze its dependence on physical factors and demonstrate how it can be used to quantitatively assess the sufficiency of material bases for polychromatic CT. Experimental studies were performed for one-, two-, and three-material decomposition using polyethylene (PE), polytetrafluoroethylene (PTFE), and aluminum (Al) at tube potentials of 50, 95, and 125 kV. In single-material decomposition, effective energy exhibits strong thickness dependence. In two-material decomposition, this dependence is substantially reduced and typically requires only first-order corrections. Crucially, in three-material decomposition, effective energy becomes largely insensitive to material thickness. These results suggest that three-material decomposition provides a sufficient representation of attenuation coefficients in polychromatic CT, as dominant spectral hardening effects can be explicitly captured by the three material bases.**

***Index Terms*—Computed tomography, beam hardening, material decomposition, effective energy, spectral imaging, photon counting.**

I. Introduction

In x-ray computed tomography (CT), a long-standing and fundamental challenge is the beam hardening effect. When x-rays interact with matter, photons are attenuated according to the Beer–Lambert law. For a monochromatic x-ray beam, the logarithm of the ratio between the incident and transmitted photon numbers corresponds exactly to a line integral of the linear attenuation coefficient at the given photon energy. This linearity forms the theoretical foundation for analytic image reconstruction. However, practical x-ray sources emit photons over a broad energy spectrum. The measured detector signal represents an energy-weighted integral of exponentially attenuated photons. Because the logarithm is a nonlinear operation, the log-transform of this spectrally averaged signal does not equal the spectral average of the corresponding monochromatic log-transformed data. This nonlinearity results in beam hardening artifacts, such as cupping and streaking, if not properly corrected [1].

Xiaopeng Yu and Guang-Hong Chen are with the Department of Medical Physics, School of Medicine and Public Health, University of Wisconsin–Madison, Madison, WI, USA. Ran Zhang is with the Department of Radiology, School of Medicine and Public Health, University of Wisconsin–Madison, Madison, WI, USA.

In single spectral CT, beam hardening correction schemes exploit the approximate scaling behavior of mass attenuation coefficients for materials with low effective atomic number (low-Z) [2]. Over the diagnostic energy range, many biologically relevant materials exhibit similar energy dependence of their mass attenuation coefficients. The essential assumption underlying this approximation is an approximate decoupling between spatial dependence and energy dependence of the attenuation coefficient for unknown materials. Dual-energy CT (DECT) relaxes this approximation by representing the attenuation coefficient as a linear combination of two basis materials, which enables improved beam hardening correction and provides additional material-specific information beyond conventional single-energy CT [3].

This progression naturally suggests broader questions: How many basis materials are sufficient for accurate material decomposition in medical CT imaging, where low-Z materials dominate typical object compositions? In what precise mathematical or physical sense can such sufficiency be defined? While multi-material decomposition is explored in photon-counting CT [4], there remains a lack of systematic theoretical analysis regarding the "limit" of sufficiency.

The purpose of this work is to address these two questions. We invoke the mean value theorem for integrals to associate each polychromatic measurement with a corresponding *effective energy*, formally linearizing the attenuation line integrals. We demonstrate that the dependence of this effective energy on the thickness provides a new perspective for quantification of the sufficiency of material decomposition in x-ray CT.

II. Theory

A. Forward Model and the Mean Value Theorem

We define the energy-dependent line integral of attenuation along the ray ℓ as $P_\ell(E) \equiv \int_\ell \mu(\vec{x}, E)\,\mathrm{d}\ell$. Let $\Omega(E)$ denote the normalized polychromatic x-ray spectrum such that $\int \Omega(E)\,\mathrm{d}E = 1$. The normalized transmission is given by:

$$\frac{N_\ell}{N_{0\ell}} = \int \Omega(E)\,\exp\big(-P_\ell(E)\big)\,\mathrm{d}E = \int \Omega(E)\,f(E)\,\mathrm{d}E, \tag{1}$$

where $N_{0\ell}$ is the total number of incident photons along ray ℓ, N_ℓ is the total number of transmitted photons of ray ℓ, and $f(E) \equiv \exp\big(-P_\ell(E)\big)$.

Since $f(E)$ is continuous and $\Omega(E)$ is a non-negative weighting function, the weighted mean value theorem for integrals guarantees the existence of at least one energy $E_\ell^* \in [E_{\min}, E_{\max}]$ such that:

$$\int \Omega(E)\,f(E)\,\mathrm{d}E = f(E_\ell^*). \tag{2}$$

Consequently, taking the negative logarithm yields:

$$\ln\left(\frac{N_{0\ell}}{N_\ell}\right) = P_\ell(E_\ell^*) = \int_\ell \mu(\vec{x}, E_\ell^*)\,\mathrm{d}\ell. \tag{3}$$

Equation (3) demonstrates that a polychromatic measurement is mathematically equivalent to a monochromatic measurement evaluated at a ray-dependent effective energy E_ℓ^*.

B. Nature and Determinants of the Effective Energy

It is worth noting that the mean value theorem is an existence theorem. It guarantees that, for each ray ℓ, there exists at least one energy $E_\ell^* \in [E_{\min}, E_{\max}]$ for which the polychromatic measurement is equivalent to a monochromatic one. However, it does not provide a closed-form expression for E_ℓ^*, nor does it ensure uniqueness or prescribe how this energy depends on physical parameters. In other words, E_ℓ^* is an emergent quantity resulting from the combined effects of the incident x-ray spectrum, object attenuation, and detector response along ray ℓ. For a multi-material object, the energy-dependent line integral can be written as

$$P_\ell(E) = \sum_m L_{\ell,m}\,\mu_m(E), \tag{4}$$

where $L_{\ell,m}$ denotes the path length of ray ℓ through material m. Accordingly, the effective energy depends on both the material composition and the thicknesses encountered along the ray, as well as on the energy dependence of the attenuation coefficients $\mu_m(E)$. Because $P_\ell(E)$ is a line integral, different rays traverse different material compositions and thicknesses. This intrinsic ray dependence of the effective energy constitutes the fundamental origin of beam hardening. Nevertheless, it should not be interpreted as the energy of photons directly measured by the detector, nor as a unique characteristic energy of the spectrum. Rather, E_ℓ^* is an *implicit equivalence energy* defined by Eq. (2), such that the polychromatic transmission equals that of a monochromatic beam at E_ℓ^*. In general, this equivalence energy may be ray dependent and need not be unique.

C. Material-Thickness Dependence

Suppose that the incident spectrum $\Omega(E)$ and the detector system, including filtration, energy response, and detection efficiency, are fixed, and that the attenuation functions $\mu_i(E)$ in Eq. (4) are known material properties. Under these conditions, the only quantities that vary across rays are the material path lengths $\{L_i\}_{i=1}^n$, where n denotes the number of basis materials used for material decomposition. Equation (4) then implicitly defines the effective energy as a deterministic function of the material thicknesses alone,

$$E_\ell^* = E^*(L_1, L_2, \ldots, L_n), \tag{5}$$

with the dependence on $\mu_i(E)$ entering parametrically.

With this definition, the log-transformed measurement along ray ℓ can be written in the formally linear-looking form

$$\ln\left(\frac{N_{0\ell}}{N_\ell}\right) = P_\ell(E_\ell^*) = \sum_{i=1}^n \mu_i[E^*(L_1, L_2, \ldots, L_n)]\; L_i. \tag{6}$$

Equation (6) resembles a monochromatic line-integral model evaluated at a single energy. However, unlike the true monochromatic case, the effective energy itself depends on the material thicknesses. As a result, material thickness enters the measurement both explicitly through the factors L_i and implicitly through the dependence of E^* on $(L_1, L_2, \ldots, L_n)$. Consequently, the mapping

$$(L_1, L_2, \ldots, L_n) \;\mapsto\; \ln\left(\frac{N_{0\ell}}{N_\ell}\right) \tag{7}$$

is intrinsically nonlinear and nonseparable, despite its superficially linear appearance shown in Eq. (3).

This hidden nonlinearity explains why global linearization using a fixed effective energy is, in general, not possible and motivates the use of calibration-based or explicitly nonlinear material decomposition strategies. In particular, differentiating the log-measurement with respect to an individual path length L_i yields spectrum-weighted effective attenuation coefficients, rather than the monochromatic coefficients $\mu_i(E_\ell^*)$. This distinction plays a central role in both calibration procedures and multi-material decomposition experiments.

III. Methods

A. Effective-Energy Parameterization

We model the ray-dependent effective energy as a low-order polynomial function of the material thickness vector $\mathbf{L} = (L_1, L_2, \ldots, L_n)$,

$$E^*(\mathbf{L}) = E^*(\mathbf{L}; \boldsymbol{c}) = \boldsymbol{\phi}(\mathbf{L})^\mathsf{T}\boldsymbol{c}, \tag{8}$$

where $\boldsymbol{\phi}(\mathbf{L})$ denotes a prescribed polynomial basis and $\boldsymbol{c}$ is the corresponding vector of unknown coefficients. In practice, the basis expansion is sometimes truncated at first or second order to balance model flexibility against stability and to mitigate overfitting.

Under this effective-energy parameterization, the log-transformed polychromatic measurement is approximated by the surrogate forward model

$$\hat{y}(\mathbf{L}; \boldsymbol{c}) = \sum_{i=1}^n \mu_i[E^*(\mathbf{L}; \boldsymbol{c})]\; L_i. \tag{9}$$

This representation enforces a single effective energy for all materials along a given ray, while allowing that energy to vary smoothly with the material thickness vector $\mathbf{L}$. We investigate three levels of model complexity:

1) Zeroth-order model: Assumes a constant effective energy, ignoring beam hardening variations:

$$E^*(\mathbf{L}) = c_0. \tag{10}$$

2) First-order model: Introduces linear dependence on thickness, capturing the dominant hardening trend:

$$E^*(\mathbf{L}) = c_0 + \sum_{i=1}^n c_i L_i. \tag{11}$$

3) Second-order model: Includes quadratic and interaction terms to capture higher-order spectral effects:

$$E^*(\mathbf{L}) = c_0 + \sum_{i=1}^n c_i L_i + \sum_{i=1}^n c_{ii} L_i^2 + \sum_{i<j} c_{ij} L_i L_j. \tag{12}$$

B. Calibration and Optimization

To determine the sufficiency of the basis size n, we perform calibration experiments. Given a set of measurements $y^{(j)}$ with known thicknesses $\mathbf{L}^{(j)}$, we estimate the parameters $\boldsymbol{c}$ by minimizing the sum of squared errors:

$$\hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{c}} \sum_{j=1}^{N} \left[y^{(j)} - \hat{y}(\mathbf{L}^{(j)}; \boldsymbol{c}) \right]^2. \tag{13}$$

Because the attenuation functions $\mu_i(E)$ are obtained from tabulated databases (e.g., NIST XCOM) and are not analytic, we solve Eq. (13) using Powell's derivative-free optimization method. This method iteratively performs one-dimensional minimizations along adaptive search directions, avoiding the need for gradients of the attenuation coefficients.

Model performance is assessed using the Root Mean Square Error (RMSE) between the predicted $\hat{y}^{(j)}$ and measured log-transmission values $y^{(j)}$,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{j=1}^{N} (y^{(j)} - \hat{y}^{(j)})^2}. \tag{14}$$

C. Sufficiency Quantification

Sufficiency is quantified by the stability of the inferred effective energy parameter (E^*) as model complexity increases. Specifically, a set of material basis is defined as sufficient when the optimal E^* remains effectively constant across the object's thickness range, even when the model permits higher-order polynomial variations. Operationally, this criterion is satisfied when introducing first- or second-order thickness-dependent terms yields negligible further reduction in RMSE—indicating that the added degrees of freedom in $E^*(\mathbf{L})$ provide no additional explanatory power. In this regime, the basis materials themselves capture the dominant spectral nonlinearities, rendering the effective energy a stable, system-level parameter rather than a variable dependent on object composition.

IV. Experimental Studies

A. Data Acquisition

Experimental measurements were acquired using homogeneous slabs of three materials: polyethylene (PE), polytetrafluoroethylene (PTFE), and aluminum (Al). Data were acquired using a benchtop PCD-CT system equipped with a CdTe-based PCD (XC-Hydra FX50, Varex Imaging, USA). A tungsten anode diagnostic x-ray tube (G1952 with B-180H housing, Varian Medical Systems, Inc., USA) with an additional 2 mm copper filter was used. Data acquisitions were performed at three tube potentials (50, 95, and 125 kV). We analyzed three distinct decomposition scenarios: One-Material ($n = 1$), Two-Material ($n = 2$), and Three-Material ($n = 3$).

B. Results: One-Material Decomposition

We first examined the case where the object is modeled using a single material basis. Table I summarizes the RMSE, effective energy E_ℓ, and range for PE, PTFE, and Al.

TABLE I
One-Material Calibration Accuracy

Material	kV	Order	RMSE	Mean E^* (keV)	Range (keV)
PE	125	0	0.0108	51.57	0.00
		1	0.0089	52.06	0.89
		2	0.0093	52.07	1.09
		3	0.0093	51.97	1.37
		4	0.0074	51.61	2.58
		5	0.0076	51.53	2.60
	95	0	0.0123	45.94	0.00
		1	0.0087	46.22	0.93
		2	0.0089	45.65	1.53
		3	0.0094	46.27	1.38
		4	0.0081	45.53	1.90
		5	0.0081	45.59	1.81
PTFE	125	0	0.0452	69.01	0.00
		1	0.0385	68.91	1.57
		2	0.0396	69.20	0.59
		3	0.0323	68.74	2.74
		4	0.0305	68.80	2.36
		5	0.0224	67.63	5.15
	95	0	0.0481	60.60	0.00
		1	0.0400	61.14	1.20
		2	0.0401	61.30	0.38
		3	0.0321	60.72	2.65
		4	0.0299	60.55	2.69
		5	0.0275	59.47	5.72

For single-material decomposition, accurate representation of the measured polychromatic data requires an effective energy model with at least fifth-order dependence on material thickness. This strong nonlinearity reflects the fundamental difficulty of capturing beam hardening using a single material basis in polychromatic CT. Moreover, effective energy varies substantially with tube potential and material choice, making it inherently difficult for a single material, such as water, to fully account for beam hardening—a limitation recognized for decades.

C. Results: Two-Material Decomposition

We extended the analysis to pairs of materials (e.g., PTFE + Al). Table II reports the RMSE, effective energy E^*, and range.

The dependence of effective energy on thickness is noticeably reduced compared to the one-material case.

- **Trend:** Across all tube voltages, introducing a first-order dependence (linear correction) substantially reduces RMSE compared to the zeroth-order model. For PTFE-Al at 125 kV, RMSE drops from 0.0752 to 0.0398.
- **Diminishing Returns:** Second-order models do not yield consistent further improvements. In some cases (e.g., PTFE-Al at 125 kV), the RMSE actually increases slightly due to potential overfitting or instability in the higher-order surface fit.

These results indicate that for two-material decomposition, the dominant nonlinearity is captured by the basis materials themselves, leaving a residual dependency that is largely linear.

D. Results: Three-Material Decomposition

Finally, we analyzed the three-material basis (PE + PTFE + Al). This setup allows the model to span a much larger attenuation space.

TABLE II
TWO-MATERIAL CALIBRATION ACCURACY

Pair	kV	Order	RMSE	Mean E^* (keV)	Range (keV)
PTFE + Al	125	0	0.0752	66.71	0.00
		1	0.0398	66.36	5.07
		2	0.0445	66.30	5.64
		3	0.0378	65.67	6.50
		4	0.0401	65.37	6.09
	95	0	0.0773	59.81	0.00
		1	0.0469	59.91	2.97
		2	0.0673	60.41	0.38
		3	0.0489	59.99	3.95
		4	0.0432	59.21	4.44
	50	0	0.1245	41.04	0.00
		1	0.1107	41.45	0.13
		2	0.0701	40.51	1.38
		3	0.0740	40.61	1.29
		4	0.0789	39.81	2.16
PE + PTFE	125	0	0.0653	67.74	0.00
		1	0.0584	67.90	2.92
		2	0.0624	68.44	0.51
		3	0.0539	66.84	3.12
		4	0.0563	65.52	6.64
	95	0	0.0640	59.60	0.00
		1	0.0596	60.27	0.89
		2	0.0637	60.06	0.71
		3	0.0530	59.94	2.49
		4	0.0544	59.56	1.72
	50	0	0.1309	39.79	0.00
		1	0.1090	39.46	0.15
		2	0.1184	40.21	0.82
		3	0.0949	39.59	2.46
		4	0.1085	38.51	4.56
PE + Al	125	0	0.0409	60.83	0.00
		1	0.0303	60.14	2.27
		2	0.0315	61.20	1.23
		3	0.0388	61.33	0.63
		4	0.0333	60.52	2.02
	95	0	0.0570	55.56	0.00
		1	0.0493	55.92	0.48
		2	0.0431	56.29	1.00
		3	0.0518	55.55	0.52
		4	0.0515	55.65	1.64
	50	0	0.1390	38.58	0.00
		1	0.1390	39.07	0.48
		2	0.1125	38.97	1.15
		3	0.1125	39.21	0.72
		4	0.1166	39.32	1.14

TABLE III
THREE-MATERIAL CALIBRATION ACCURACY

kV	Order	RMSE	Mean E^* (keV)	Range (keV)
125	0	0.069	65.81	0.00
	1	0.063	66.44	0.58
	2	0.062	66.46	0.60
	3	0.057	66.91	4.25
	4	0.055	65.08	4.64
95	0	0.072	59.70	0.00
	1	0.062	59.49	0.02
	2	0.067	60.46	0.17
	3	0.055	59.52	2.96
	4	0.051	59.05	4.38
50	0	0.156	40.62	0.00
	1	0.115	40.48	0.10
	2	0.115	40.40	0.82
	3	0.115	40.58	2.37
	4	0.109	40.01	2.99

The results in Table III demonstrate a profound stability in the effective energy:

- **Insensitivity to Thickness:** The improvement from zeroth-order to first-order is marginal. At 125 kV, the RMSE reduction is from 0.069 to 0.063, a smaller relative gain than in the 1-material case.
- **Stability at Low Voltage:** At 95 kV, using a first-order model yields a fitted effective energy range of only 0.02 keV. This implies the effective energy is essentially constant.
- **Model Consistency:** The system behaves nearly monochromatically. The physics of beam hardening are fully resolved by the superposition of the three basis materials.

V. DISCUSSION AND CONCLUSION

The experimental hierarchy shows strong thickness dependence for $n = 1$, linear dependence for $n = 2$, and near-independence for $n = 3$, which provides a quantitative definition of "sufficiency."

In the one material decomposition setting, the effective energy exhibits the strongest dependence on material thickness, necessitating the inclusion of higher-order thickness terms in order to adequately capture the underlying polychromatic physics. This behavior reflects the fact that, in the absence of additional basis materials, all spectral hardening effects must be absorbed into a single effective-energy parameterization.

When extending to two-material decompositions, the thickness dependence of the effective energy is substantially reduced. The introduction of a second basis material allows part of the spectral variation to be represented explicitly through material-specific attenuation terms, thereby weakening the residual nonlinearity that must be modeled through the effective energy itself. In this case, first-order or, in some instances, second-order thickness dependence is sufficient to achieve near-optimal agreement with the measured data.

In contrast, for three-material decompositions, the effective energy becomes largely insensitive to material thickness, with higher-order models providing minimal additional benefit. This indicates that the dominant spectral hardening effects are effectively accounted for by the material decomposition itself, leaving only weak residual dependence to be absorbed by the effective energy parameter. From this perspective, the order required in the effective-energy model decreases systematically as the number of basis materials increases, with the one-material case representing the most demanding scenario in terms of required model complexity.

REFERENCES

[1] J. Hsieh, "Computed tomography: principles, design, artifacts, and recent advances," 2003.
[2] Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelrieß, "Empirical beam hardening correction (ebhc) for ct," *Medical physics*, vol. 37, no. 10, pp. 5179–5187, 2010.
[3] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting ct: technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293–312, 2018.
[4] Y. Long and J. A. Fessler, "Multi-material decomposition using statistical image reconstruction for spectral ct," *IEEE transactions on medical imaging*, vol. 33, no. 8, pp. 1614–1626, 2014.

Physics-Informed Adaptive Filtering Based on VVBP τ-Coordinate Properties for Low-Dose CT Image Recovery

Junhao Wu, Yuanxiang Gao, Haomin Hou, Deyang Kong, Yankun Wang, Yule Li, Jingyi Liao, Hao Wang, Zhaoying Bian, Jianhua Ma and Dong Zeng

***Abstract*—In low-dose computed tomography (CT) imaging, noise is highly coupled with anatomical structures, making it challenging for traditional image-domain denoising methods to effectively separate noise from true signals. This paper proposes an adaptive filtering method based on the physical properties of the τ-coordinate in View-by-View Back Projection (VVBP) for image recovery. The framework constructs a differentiable VVBP tensor that maps noisy projection data into a third-order tensor preserving complete angular information. By sorting the projection values along the view dimension, the original angular coordinate $\boldsymbol{\theta}$ is transformed into the structural coordinate $\boldsymbol{\tau}$. The $\boldsymbol{\tau}$-coordinate has a clear physical meaning: $\boldsymbol{\tau = 0.5}$ corresponds to the structural normal direction, while $\boldsymbol{\tau = 0/1}$ corresponds to the tangential direction. Leveraging the physical partitioning characteristics of the τ-coordinate, we design an adaptive filter that integrates median-slice priors and explicit anatomy embedding to restore monotonicity along the τ-dimension and enhance structural consistency. Experimental results demonstrate that on abdominal 1/8-dose data, our method outperforms traditional FBP and traditional denoising methods in terms of quantitative metrics.**

***Index Terms*—Adaptive Filtering, Low-Dose CT Image Recovery, VVBP sorting, τ-Coordinate**

I. Introduction

LOW-DOSE computed tomography (CT) technology reduces patient radiation exposure but introduces severe quantum noise and electronic noise, leading to degraded image quality in reconstructed images. Traditional filtered back projection (FBP) methods produce significant noise under low-dose conditions, affecting clinical diagnosis. In recent years, deep learning-based CT recovery methods have achieved remarkable progress. However, most methods rely on paired training data and often perform denoising in the image domain, failing to fully utilize the physical priors embedded in the projection data. Low-dose CT recovering methods can be broadly categorized into iterative recovery (IR) and deep learning-based approaches. IR methods, such as statistical iterative recovery (SIR), model noise statistics but are computationally intensive [1]. Deep learning methods have shown superior performance in recent years, with networks like RED-CNN [2] and LEARN [3] achieving state-of-the-art results. However, these methods typically require paired high-dose and low-dose images for supervised training, which may not always be available in clinical practice. Unsupervised methods, such as Tensor-Net [4] and CycleGAN [5] have been applied to CT denoising without clean targets. These methods rely on statistical independence or cycle consistency but often lack explicit physical constraints, which can lead to structural distortions or residual noise. VVBP theory provides a mathematical framework for understanding the relationship between projection data and local structures [6]. The sorting operation transforms the original angular coordinate into a structural coordinate τ, which has been shown to be effective for low-rank approximation and sparse-view reconstruction [7-8].

This work was supported in part by National Key R&D Program of China under Grants 2024YFC2417800 and 2024YFA1012000, National Natural Science Foundation of China under Grant 62571228, Beijing Natural Science Foundation under Grant Z250002, and Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515140146. *(Corresponding authors: Hao Wang, Dong Zeng.)*

Junhao Wu, Yuanxiang Gao, Haomin Hou, Deyang Kong, Yank un Wang, Yule Li, Jingyi Liao, Hao Wang, Zhaoying Bian and Do ng Zeng is with the School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China. (e-mail: wujunhao@smu.edu.cn, gyx2001@smu.edu.cn, houhm@smu.edu.cn, kongdy@smu.edu.cn, ykwang@smu.edu.cn, ylli@smu.edu.cn, ljy@smu.edu.cn, wang_hao@smu.edu.cn, zybian@smu.edu.cn and zd1989@smu.edu.cn).

Jianhua Ma is with the School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China. (e-mail: jhma@smu.edu.cn).

View-by-View Back Projection (VVBP) is a CT reconstruction theory based on sorting operations [9]. By sorting the multi-view back-projected values for each pixel, the original angular coordinate$\theta$is mapped to a structural coordinate τ, making local structures (edges, etc) of the object explicit in the sorted tensor slices. The τ-coordinate has a clear physical meaning: its value reflects the ray density distribution, which in turn is related to the noise distribution.

Building on VVBP theory, this paper proposes a τ-coordinate-guided adaptive filter. We construct an adaptive filter that can recover the low-dose CT images through filtering along τ -coordinate [9-10]. Leveraging the physical partitioning characteristics of the τ-coordinate to design adaptive filtering. Then, compare the recovery results with traditional Gaussian filtering and FBP methods .

II. Method

A. Distribution Characteristics of Sorted VVBP Tensor

The core of our approach lies in analyzing the distribution characteristics of the View-by-View Back-Projection (VVBP) tensor after sorting along the view dimension. When noisy projection data from low-dose CT scans undergo VVBP tensor construction followed by sorting, distinct distribution patterns emerge that directly inform our filtering strategy.

The sorting operation transforms the originally chaotic angular fluctuations into an organized profile along the τ-dimension, where τ represents the normalized structural importance coordinate. In ideal noise-free conditions, this profile exhibits strict monotonic increase from τ=0 to τ=1, reflecting the physical hierarchy from tangential to normal ray directions relative to structural boundaries. However, under low-dose conditions with amplified quantum noise, this monotonicity is significantly compromised. The high-pass amplification inherent in the ramp filter during FBP procedure substantially increases noise variance in the back-projected domain, creating what we term “sorting inconsistency”.

Our analysis reveals that noise distribution in the sorted τ-profile is highly non-uniform. The central region around τ = 0.5, which corresponds to the normal direction with maximum ray density, contains the strongest structural signals but also exhibits distinct noise characteristics. Due to the signal plateau in this region, small noise fluctuations can disrupt the expected monotonic trend, leading to local ranking errors. Conversely, the extreme regions near τ = 0 and τ = 1, corresponding to tangential directions with minimal ray density, contain predominantly noise with minimal structural information as shown in Fig 1. These regions demonstrate the highest relative noise levels and the least reliable sorting outcomes.

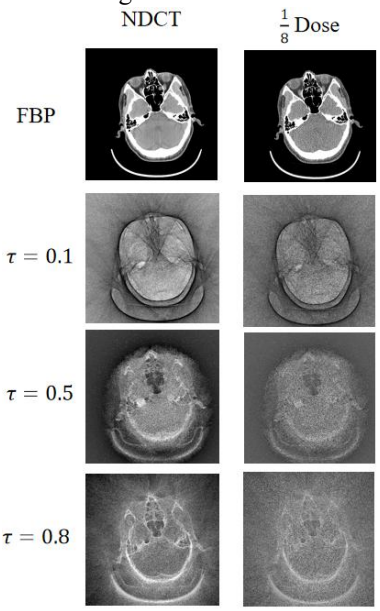


Fig. 1. VVBP-Tensor Slices of Noise Level of τ-corrdinated VVBP-Tensor of Normal Dose and 1/8 Dose Images

The transitional regions present the most complex distribution, containing rich structural details from the transition between grazing and penetrating rays while simultaneously being vulnerable to noise corruption. This analysis of noise distribution along the τ-dimension provides the foundation for our adaptive filtering approach, where different τ intervals require distinct filtering

strategies based on their specific noise and structural characteristics.

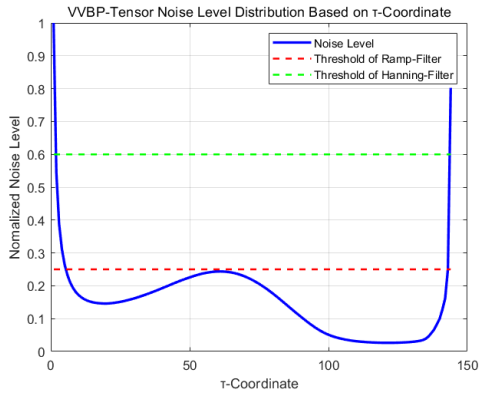


Fig. 2. Noise Level of unnormalized τ-corrdinated VVBP-Tensor

B. Adaptive Filtering with 3D Tensor Median Prior Recovery

Guided by our analysis of sorted VVBP tensor distributions, we propose an adaptive filter with 3D tensor median prior recovery for low-dose CT recovery. The complete processing pipeline, illustrated in Fig. 3.

After tensor construction, we first perform an initial analysis of the unsorted tensor to establish baseline noise characteristics. Applying the sorting operation along the view dimension for each spatial position, generating the sorted tensor $\mathcal{T}_{\text{sorted}}(\mathbf{r}, \tau)$. We divide the τ-dimension into three distinct intervals based on our distribution analysis: the central normal region ($\tau \in [0.4,0.6]$), (2) the transitional regions ($\tau \in [0.2,0.4] \cup [0.6,0.8]$) and (3) the tangential regions ($\tau \in [0,0.2] \cup [0.8,1.0]$).

A crucial component of our approach is the extraction and utilization of the median slice at $\tau = 0.5$ as a robust structural prior. This slice, denoted as $M = \mathcal{T}_{\text{sorted}}(:,:,V/2)$, contains the most reliable structural information due to maximum ray

density and exhibits statistical robustness against noise fluctuations. We analyze this median slice to extract two types of prior information: structural boundaries through edge detection and regional noise characteristics through local variance analysis.

The extracted prior information serves two purposes. First, it guides the adaptive filtering strength across different τ intervals. For regions corresponding to strong edges in the median slice, we apply weaker filtering to preserve structural details. For homogeneous regions in the median slice, we apply stronger filtering to suppress noise. Second, the median slice provides reference intensity values that help correct local intensity deviations in adjacent τ slices caused by sorting inconsistency.

III. EXPERIMENTS

A. Experimental Setup

Based on our distribution analysis and guided by the median slice prior, we implement interval-specific filtering strategies:

(1) Central Normal Region: This region receives moderate Gaussian filtering ($\sigma = 1.2$) combined with

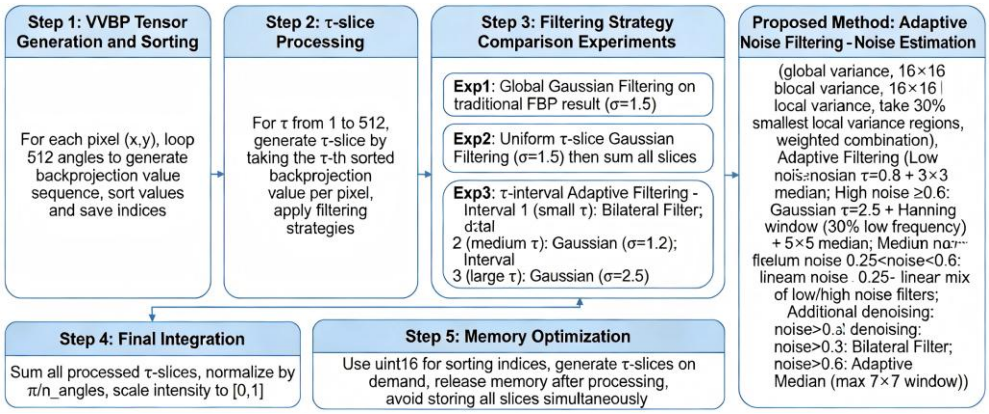


Fig. 3. Pipeline of the Denoising Procedure

edge-preserving bilateral filtering. The filtering strength is dynamically adjusted based on edge information from the median slice, with reduced filtering near structural boundaries and increased filtering in homogeneous areas.

(2) Transitional Regions: These regions employ structure-preserving filters that balance noise reduction with detail retention. We apply anisotropic diffusion filtering followed by light Gaussian smoothing ($\sigma = 0.8$). The diffusion parameters are modulated by the gradient magnitude calculated from the median slice to prevent oversmoothing of fine textures.

(3) Tangential Regions: These noise-dominated regions receive strong Gaussian filtering ($\sigma = 2.5$) followed by median filtering to effectively suppress noise while minimizing the introduction of artifacts.

After interval-specific filtering, we enforce consistency along the τ-dimension through a weighted averaging scheme. Each filtered τ-slice is adjusted by considering its immediate neighbors in the τ-domain, with weights determined by the similarity of local structural patterns. This step ensures smooth transitions between adjacent τ-slices and mitigates residual sorting inconsistencies.

The final recovery is obtained by summing the filtered and consistency-enforced tensor along the τ-dimension, followed by appropriate scaling.

In this paper, we used the simulated 1/8-dose projection data added Poisson and Gaussian noise which original data were obtained from clinical abdominal CT data. The dataset contained 100 cases and the filters used in different τ interval are shown in Table I. Comparative methods included FBP (Ram-Lak Filter), Global Gaussian (Gaussian Filter), NDCT (Normal Dose CT Image). In the experiments, the number of VVBP-Tensor Slices is 512 and the experiments were performed on an NVIDIA RTX 4090 GPU.

TABLE I
Filters used in τ-Interval

Filter type	Number of Slices Used	Percentage (%)	Application Condition (Noise Level)
Ramp Filter	497	97.07	Noise≤0.25
Hybrid Filter	12	2.34	0.25<Noise< 0.6
Hanning Filter	3	0.59	Noise≥ 0.6
Bilateral Filter	11	2.15	Noise> 0.3
Adaptive Median Filter	3	0.59	Noise> 0.6
Wavelet Denoising	4	0.78	Noise> 0.5
τ- Consistency	512	100.00	All Slices

B. *Evaluation Metrics*

To objectively evaluate denoising performance, this paper uses normal dose CT image as ground truth. Root mean square error (RMSE), peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) are employed as quantitative assessment metrics. By comparing the differences between the denoised images and the ground truth, the performance of this method in eliminating noise, suppressing artifacts, and reserving structural detail's quantitative reliability of the system is validated.

IV.RESULTS

Table 1 summarizes the dynamic filtering strategy was statistically analyzed across all 512 τ-slices. The distribution, as summarized in the provided table I, reveals a clear pattern aligned with the theoretical noise characteristics along the τ-dimension. A dominant 97.07% of slices required only mild Ramp-like filtering, corresponding to low estimated noise levels (≤ 0.25). This validates a key advantage of the VVBP sorting operation: it transforms random projection noise into high-frequency jitter along τ, concentrating the corrupted signal in a

minority of slices. Consequently, only approximately 3% of slices necessitated moderate or strong filtering to address severe noise corruption. This distribution demonstrates that our method is a targeted, physics-informed correction mechanism that applies significant processing only where physiologically indicated.

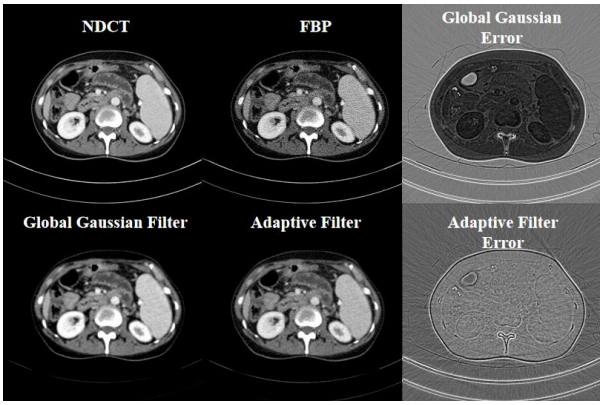


Fig. 4. Normal Dose CT Image and denoised Images.

Fig. 2 shows the direct experimental evidence for the theoretical partitioning of the τ-coordinate. The noise profile exhibits that noise is minimal at the center ($\tau \approx 0.5$) and increases symmetrically towards the boundaries. The boundaries correspond to tangential directions with minimal structural information, resulting in noise-dominated slices. This empirical noise map serves as the foundational justification for our adaptive filtering strategy, which applies progressively stronger filtering from the central normal region to the peripheral tangential regions. The quantitative performance of the proposed method is compared against traditional FBP in the table below:

TABLE II
Quantitative Indicator Comparison

	FBP	Global Gaussian	Adaptive Filter
PSNR	21.680±1.778dB	22.16±1.731dB	22.31±2.149dB
SSIM	0.3682±0.0598	0.4804±0.0356	0.6241±0.0349
RMSE	0.0841±0.0163	0.0764±0.0143	0.0820±0.0191

The τ-consistency enforcement and adaptive filtering successfully recover tissue boundaries and textural details that were obscured by noise. Fig. 4 shows that the pervasive grainy noise present in the FBP reconstruction, especially in homogeneous soft-tissue regions, is significantly reduced, yielding a smoother appearance. Critical anatomical features such as bone edges and organ boundaries remain sharp and well-defined, with no noticeable blurring or loss of resolution due to over-smoothing. The reconstruction does not introduce new streaking or blotchy artifacts, indicating a stable and controlled filtering process governed by the physical prior.

V. CONCLUSION

This paper has presented an adaptive filtering framework for low-dose CT recovery that fundamentally reinterprets the denoising problem through the lens of sorted VVBP tensors. By shifting the focus from image-domain filtering to the physics-informed, high-dimensional tensor domain, we address the core challenge of noise-structure coupling inherent in low-dose acquisitions.

In conclusion, this work establishes a principled, physics-integrated pathway for low-dose CT image recovery. Future work will focus on utilizing this method in neural network further bridging the gap between deep learning efficacy and imaging physics interpretability.

REFERENCES

[1] Fessler, J. A., Sonka, M., & Fitzpatrick, J. M. (2000). Statistical image reconstruction methods for transmission tomography. *Handbook of medical imaging*, *2*, 1-70.

[2] Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., ... & Wang, G. (2017). Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, *36*(12), 2524-2535.

[3] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "Learn: Learned experts' assessmentbased reconstruction network for sparse-data ct," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1333–1347, 2018..

[4] X. Tao, Y. Wang, L. Lin, Z. Hong, and J. Ma, "Learning to reconstruct ct images from the vvbp-tensor," *IEEE Transactions on Medical Imaging*, vol. 40, no. 11, pp. 3030–3041, 2021.

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. ICCV*, 2017, pp. 2223–2232.

[6] Tao, X., Zhang, H., Wang, Y., Yan, G., Zeng, D., Chen, W., & Ma, J. (2019). VVBP-tensor in the FBP algorithm: its properties and application in low-dose CT reconstruction. *IEEE transactions on medical imaging*, *39*(3), 764-776.

[7] W. Wang and H. Yu, "View-by-view back projection for structure-aware CT reconstruction," *IEEE Trans. Med. Imaging*, vol. 40, no. 5, pp. 1234–1245, May 2021.

[8] Zhao, X., Du, Y., & Peng, Y. (2025). VVBPNet: Deep learning model in view-by-view backprojection (VVBP) domain for sparse-view CBCT reconstruction. *Biomedical Signal Processing and Control*, *102*, 107195.

[9] Wang, W., Zhou, C., Chen, X., Tang, G., Ma, J., Gao, Y., & Li, S. (2025). A VVBP data-based pancreatic lesion detection model with noncontrast CT. *Computerized Medical Imaging and Graphics*, 102601.

[10] Zhu, M., Wang, Z., Wang, C., Zeng, C., Zeng, D., Ma, J., & Wang, Y. (2024). VBVT-Net: VOI-Based VVBP-Tensor Network for High-Attenuation Artifact Suppression in Digital Breast Tomosynthesis Imaging. *IEEE Transactions on Medical Imaging*.
[11] Zhu, M., Chen, G., Zhu, Q., Song, Y., Wang, Y., & Ma, J. (2023, April). VVBP-tensor-based deep neural network for metal artifact reduction in computed tomography. In *Medical Imaging 2023: Physics of Medical Imaging* (Vol. 12463, pp. 1035-1040). SPIE.

WICD: Wavelet-Informed Conditional Diffusion Model for Low-Dose CT Denoising

Changsheng Fang, Bahareh Morovati, Shuo Han, Yu Shi, Li Zhou, Shuyi Fan, Dayang Wang, Hengyong Yu, *Fellow, IEEE*

***Abstract*—Low-dose CT reduces radiation exposure, but introduces strong quantum noise that degrades structural fidelity. Existing diffusion-based denoisers operate in the image domain and require costly multi-step sampling, while failing to exploit the multi-scale nature of CT images. We propose WICD, a Wavelet-Informed Conditional Diffusion framework that performs generative modeling exclusively on the deepest wavelet low-frequency subband, enabling efficient reconstruction of global anatomy with drastically reduced dimensionality. A complementary high-frequency enhancement module further refines the directional subbands to restore fine textures suppressed by low-dose noise. After inverse wavelet synthesis, WICD produces clean and structurally accurate images. Experiments on the AAPM 2016 LDCT dataset show that WICD outperforms baseline methods by a significant margin. These results demonstrate the effectiveness of combining wavelet priors and low-dimensional diffusion for high-quality LDCT denoising.**

***Index Terms*—Computed Tomography, Diffusion Model, Low-dose CT Denoising, Wavelet Transform.**

I. Introduction

COMPUTED tomography (CT) plays a critical role in routine clinical diagnosis due to its high resolution and rapid acquisition capabilities. Nevertheless, radiation exposure remains a major concern that motivates the use of low-dose CT (LDCT) protocols. Reducing tube current or voltage effectively lowers radiation dose, but it also introduces substantial quantum noise, leading to degraded contrast, mottled textures, and loss of anatomically relevant details in reconstructed images. These limitations require robust LDCT denoising techniques to restore diagnostic fidelity.

Conventional filtered back-projection (FBP) reconstruction amplifies noise under low-dose conditions, while model-based iterative reconstruction (MBIR) approaches [1]–[4] suppress noise more effectively but require intensive computation and tend to oversmooth fine structures. In parallel, a wide range of deep learning–based LDCT denoising methods [5]–[7] have shown improved noise suppression but often struggle to balance texture preservation and generalization across dose levels. Recent advances in diffusion models [8]–[11] have demonstrated great potential for medical image restoration. However, most existing diffusion-based methods operate directly in the image domain and rely on multi-step stochastic sampling, resulting in high inference cost. Moreover, image-domain diffusion models treat all spatial information uniformly and thus fail to exploit the multi-scale nature of CT images, making it difficult to distinguish noise from meaningful structures.

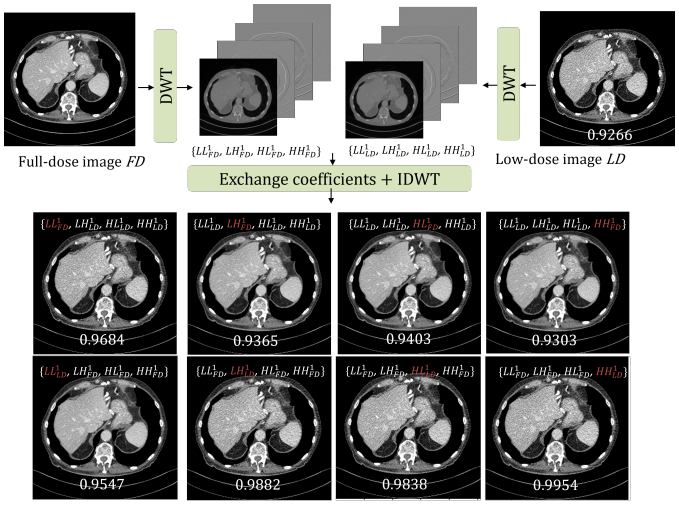


Fig. 1. Illustration of exchanging wavelet coefficients between low-dose and full-dose images. Replacing low-dose frequency components with full-dose counterparts consistently improves SSIM, while introducing low-dose components into the full-dose image leads to SSIM degradation. Among all subbands, low-frequency coefficients have the most significant impact, dominating global structure and contrast, whereas high-frequency components mainly affect local details and noise textures. This demonstrates the effectiveness of wavelet-based representations in characterizing structural fidelity in low-dose CT.

To better analyze the noise characteristics in LDCT, we perform the wavelet decomposition shown in Fig. 1. The deepest low-frequency subband preserves global anatomical structure with moderate corruption, whereas high-frequency subbands contain strong directional noises, but still retain important edge information. This indicates that low-frequency component is well suited for generative restoration, while high-frequency details are better handled through targeted refinement. Motivated by this observation, we propose a wavelet-informed conditional diffusion (WICD) framework for LDCT denoising as shown in Fig. 2.(a). The low-frequency conditional diffusion (LFCD) module performs generative modeling only on the deepest low-frequency subband, significantly reducing dimensionality and enabling efficient learning of structural priors. The high-frequency subbands are refined with a lightweight high-frequency enhancement module (HFEM) to recover textures and edge details. After inverse wavelet synthesis, the combined reconstruction produces clean and structurally consistent images without requiring iterative data-consistency procedures.

This work involved human subjects in its research. The authors confirm that the research is exempt from the Institutional Review Board at the University of Massachusetts Lowell (IRB# 23-043). This work was supported in part by NIH/NIBIB under grants R01EB032807 and R01EB034737.

The authors are with the Department of Electrical and Computer Engineering, University of Massachusetts Lowell, Lowell, MA 01854. H.Y. Yu serves as the corresponding author (email: hengyong-yu@ieee.org).

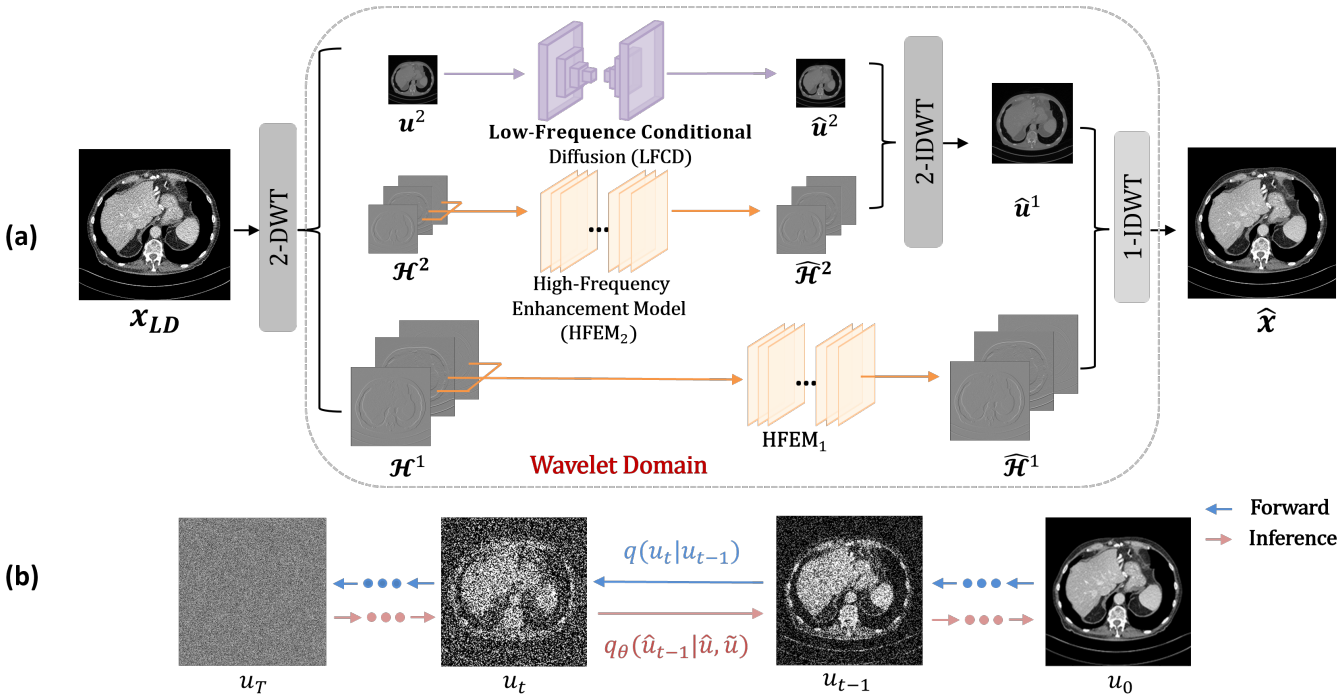


Fig. 2. Illustration of the proposed denoising network for $L = 2$. (a). Overview of the proposed WICD framework. (b). Forward and reverse processes of the low-frequency conditional diffusion (LFCD) module.

By relocating generative modeling from the image domain to the wavelet domain and complementing it with selective refinement of directional textures, the proposed WICD framework achieves an effective balance between denoising accuracy, computational efficiency, and structural fidelity. This dual-domain strategy enables a compact yet highly capable solution for LDCT restoration. The key contributions of this work are:

- Wavelet-domain diffusion: A low-frequency–only diffusion scheme to minimize complexity while preserving global structure.
- High-frequency enhancement: A lightweight module to selectively restore directional textures.
- Dual-domain design: A compact framework to combine generative structure recovery and efficient texture refinement.

II. Methodology

A. Discrete Wavelet Transformation

Given a low-dose CT image $x \in \mathbb{R}^{H_x \times W_x \times 1}$, we apply a 2D discrete wavelet transform (DWT) $\mathcal{W}$ to decompose it into one low-frequency and three high-frequency sub-bands:

$$\{LL^1, LH^1, HL^1, HH^1\} = \mathcal{W}(x), \tag{1}$$

where each sub-band lies in $\mathbb{R}^{\frac{H_x}{2} \times \frac{W_x}{2} \times 1}$. The LL component contains the global anatomical structure and is less affected by noise, while $\{LH, HL, HH\}$ capture directional details and are more sensitive to low-dose corruption.

We denote $u = LL$ and $\mathcal{H} = \{LH, HL, HH\}$, and recursively apply DWT to the low-frequency component to obtain a multi-scale representation:

$$\{u^l, \mathcal{H}^l\} = \mathcal{W}(u^{l-1}), \quad l \in [1, L], \tag{2}$$

where $\{u^l, \mathcal{H}^l\} \in \mathbb{R}^{\frac{H_x}{2^l} \times \frac{W_x}{2^l} \times 1}$. This hierarchical decomposition yields progressively compact structural features.

In our framework, diffusion operates only on the deepest low-frequency sub-band u^L, which provides a stable and informative target for generative restoration in LDCT imaging. The high-frequency components $\mathcal{H}^L$ are refined through a lightweight enhancement module to restore fine textures while suppressing noise. Restricting diffusion to u^L reduces spatial dimensionality by a factor of 4^L, significantly lowering computational and memory costs while preserving structural fidelity.

B. Low-frequency Conditional Diffusion (LFCD)

The proposed LFCD module restores global structural information in LDCT images by performing conditional diffusion in the wavelet domain. Instead of operating in full-resolution image space, LFCD applies diffusion only to the low-frequency component u extracted via wavelet decomposition. This compact representation contains the dominant anatomical structure and substantially reduces the computational cost of generative modeling. Forward diffusion and reverse denoising are integrated into a unified training framework, enabling the network to learn noise perturbation and conditional restoration directly from noisy LDCT inputs.

As shown in Fig. 2.(b), during training, the clean low-frequency coefficient u_0^L is diffused according to a variance

Algorithm 1 LFCD Training Process

input: Low-frequency coefficients of full/quarter dose image pairs u_0^L and $\tilde{u}$; time step T; number of implicit sampling steps S.
while not convergence **do**
  # Forward diffusion process
  $t \sim \text{Uniform}\{1, \ldots, T\}, \quad \epsilon_t \sim \mathcal{N}(0, I)$
  $u_t = \sqrt{\bar{\alpha}_t}\, u_0^L + \sqrt{1-\bar{\alpha}_t}\, \epsilon_t$
  Performing a gradient descent step for

$$\nabla_\theta \|\epsilon_t - \epsilon_\theta(u_t, \tilde{u}, t)\|^2$$

  # Reverse denoising process
  Initializing $\hat{u}_T \sim \mathcal{N}(0, I)$
  for $i = S : 1$ **do**
    $t = (i-1) \cdot T/S + 1$
    $t_{\text{next}} = (i-2) \cdot T/S + 1$ if $i > 1$, else 0
    $\hat{u}_{t_{\text{next}}} =$
    $\sqrt{\bar{\alpha}_{t_{\text{next}}}} \left(\frac{\hat{u}_t - \sqrt{1-\bar{\alpha}_t}\, \epsilon_\theta(\hat{u}_t, \tilde{u}, t)}{\sqrt{\bar{\alpha}_t}} \right) + \sqrt{1-\bar{\alpha}_{t_{\text{next}}}}\, \epsilon_\theta(\hat{u}_t, \tilde{u}, t)$
  end for
  Performing a gradient descent step for $\nabla_\theta \|\hat{u}^L - u_0^L\|^2$
end while
output: θ, $\hat{u}^L$

schedule $\{\beta_t\}_{t=1}^T$, producing noisy samples u_t. The reverse process begins with Gaussian noise $u_T \sim \mathcal{N}(0, I)$ and progressively restores the signal under the guidance of the noisy low-frequency input $\tilde{u}$. At each timestep, the network predicts the noise $\epsilon_\theta(u_t, \tilde{u}, t)$, and a reconstruction-consistency term is applied between the predicted clean coefficient $\hat{u}_0^L$ and ground truth u_0^L. The training loss is:

$$\mathcal{L}_{\text{LFCD}} = \mathbb{E}_{u_0,\, t,\, \epsilon_t} \left[\|\epsilon_t - \epsilon_\theta(u_t, \tilde{u}, t)\|^2 + \lambda \|\hat{u}_0^L - u_0^L\|^2 \right]. \tag{3}$$

During inference, LFCD starts with random noise u_T and performs conditional reverse diffusion guided by $\tilde{u}$ to obtain a clean low-frequency representation $\hat{u}^L$. This process provides stable denoising and significantly lower computational cost compared with image-domain diffusion methods. Algorithms 1 summarize the training procedures. Algorithm 1 details the conditional diffusion and reconstruction consistency learning on the low-frequency coefficients.

C. High-frequency Enhancement Module (HFEM)

Although the LFCD restores the global low-frequency structure of LDCT images, the high-frequency subbands obtained from wavelet decomposition remain strongly corrupted by quantum noise, leading to texture blurring and directional degradation. To refine these components, we introduce the High-frequency Refinement Module to enhance directional subbands through cross-feature interaction and lightweight hierarchical refinement.

As shown in Fig. 3, at the L^{th} wavelet level, each high-frequency subband $\mathcal{H}_k^L$ ($k \in \{LH, HL, HH\}$) is first processed with a depth-wise separable convolution to extract orientation-specific representations. The resulting features are

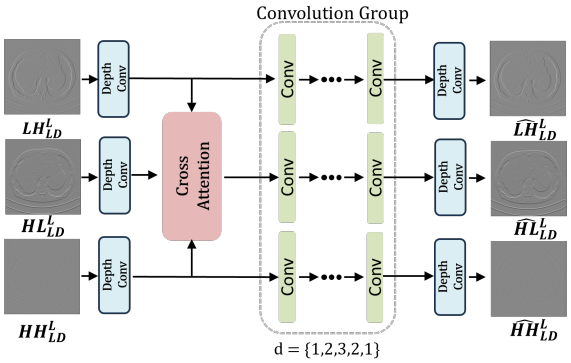


Fig. 3. Architecture of the proposed HFEM. Depth Conv denotes the depth-wise separable convolution.

then integrated by a cross-direction fusion module to adaptively aggregate complementary information across subbands. A compact enhancement block combining dilated and depth-wise separable convolutions further strengthens fine textures while expanding the receptive field.

Finally, the refined high-frequency coefficients $\mathcal{H}$ are fused with the predicted low-frequency component via the inverse discrete wavelet transform (IDWT) to reconstruct the output at the previous scale:

$$\hat{u}^{L-1} = \mathcal{W}^{-1}\big(\hat{u}^L, \hat{\mathcal{H}}^L\big). \tag{4}$$

III. Experiment and Result

A. Experimental Settings

1) Implementation Details: All experiments are implemented in PyTorch and trained on a single NVIDIA RTX A6000 (48GB) GPU. WICD is optimized end-to-end for 30 epochs using the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an initial learning rate of 2×10^{-4}, decayed by 0.8 every 2×10^3 iterations. The batch size is 16 with 256×256 patches.

Wavelet decomposition uses $L = 2$ levels. The diffusion model employs a U-Net backbone with residual and attention modules. We use $T = 300$ diffusion steps and $S = 10$ implicit sampling steps for both training and inference. Training includes reconstruction and data-consistency losses with light TV regularization for stability.

2) Baseline and Evaluation Metrics: To evaluate the proposed WICD, we compare it with several representative LDCT denoising baselines. FBP provides the traditional reconstruction reference under low-dose conditions. DeepGuess [12] is a CNN-based regression model that maps noisy inputs to clean CT images. DeepMamba [13] and MambaIR [14] employ state-space modeling to strengthen long-range feature interactions. All methods are trained and evaluated under identical data splits and preprocessing settings to ensure a fair comparison.

Performance is evaluated using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) for fidelity and Learned Perceptual Image Patch Similarity (LPIPS) for perceptual quality assessment.

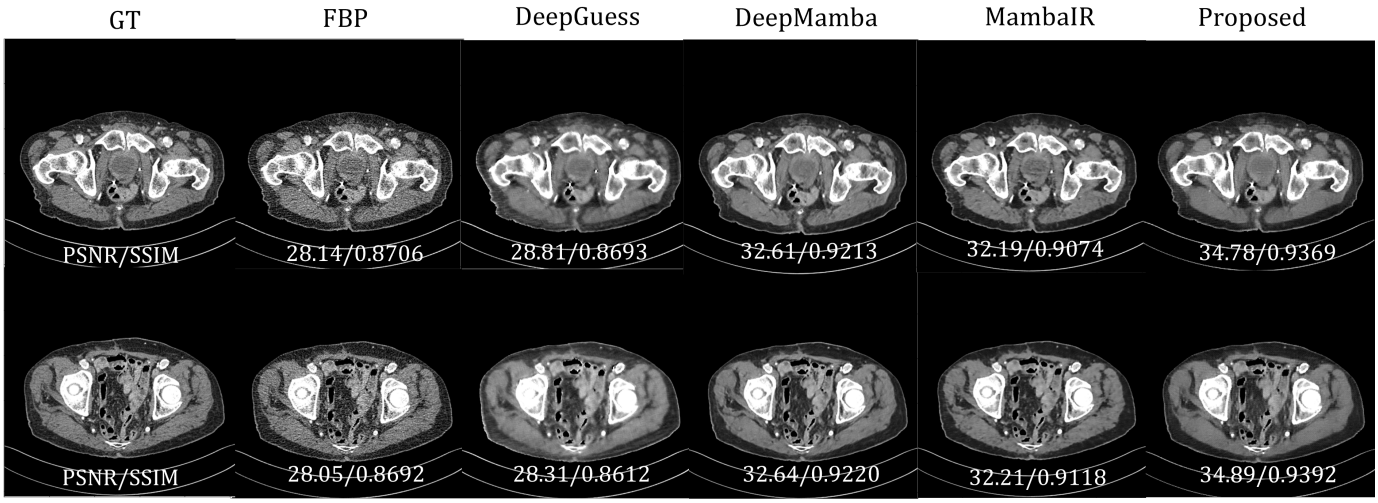


Fig. 4. Qualitative comparison of our method and competitive methods on the 2016 AAPM Low-Dose CT Grand Challenge dataset. The display window is [-160,240] HU.

TABLE I
QUANTITATIVE COMPARISON OF DIFFERENT METHODS ON TESTED DATASETS. THE BEST RESULTS ARE HIGHLIGHTED IN **BOLD**.

	Reference	PSNR	SSIM	LPIPS
FBP	-	28.3893±0.4499	0.8739±0.0071	0.0991±0.0082
MambaIR	ECCV'2024	32.4807±0.4519	0.9100±0.0055	0.0895±0.0102
DenoMamba	JBHI'2025	32.9556±0.4613	0.9144±0.0054	0.0894±0.0235
DeepGuess	CIMG'2025	28.6592±1.3072	0.8711±0.0027	0.1079±0.0060
Proposed	-	**35.5762±0.6802**	**0.9449±0.0042**	**0.0559±0.0121**

B. Data Preparation

Experiments were conducted on the abdominal CT dataset from the 2016 AAPM Low-Dose CT Grand Challenge. The scans were acquired on Siemens Somatom Definition AS+/Flash scanners (single-source mode) using a flying focal spot technique, automated exposure control (CareDose 4D), and automated tube potential selection (CarekV), with 120 kV and 200 effective mAs as the reference setting. Low-dose data were simulated by adding Poisson noise to the preprocessed projection data to approximate 25% of the original radiation dose. Ten patient cases were provided; we used patient L506 (526 slices) exclusively for testing, while the remaining nine patients (18,772 slices) were used for training. The dataset covers a range of patient sizes and includes normal anatomy as well as local liver lesions ($\leq$ 3cm), consistent with the challenge criteria.

C. Results and Analysis

Fig. 4 shows that the differences among the competing methods remain substantial. The quarter-dose FBP reconstructions are dominated by strong quantum noise, resulting in mottled textures, blurred boundaries, and poor anatomical visibility. DeepGuess reduces noise, but introduces clear oversmoothing, causing fine structures (*e.g.*, bone–soft tissue interfaces and muscle textures) to lose definition. DeepMamba and MambaIR produce cleaner and more stable reconstruction results; however, noticeable residual speckle noise persists in homogeneous regions, and a slight deformation appears around organ edges. Their visual limitations align with the moderate PSNR/SSIM values observed.

Table I summarizes the quantitative results of different methods in the testing datasets in terms of PSNR, SSIM, and LPIPS, where the proposed WICD consistently shows superior performance. Its reconstructions present uniform soft-tissue regions, sharp cortical bone outlines, and clearly preserved fine textures without introducing artificial patterns or directional artifacts. Both assessed slices show WICD exceeding those of all baseline methods. This substantial improvement reflects the effectiveness of modeling structural information through low-frequency wavelet-domain diffusion while refining high-frequency details with a lightweight enhancement module. The resulting images exhibit strong noise suppression and high structural fidelity, avoiding both oversmoothing and residual noise.

IV. DISCUSSION AND CONCLUSION

The results show that separating LDCT denoising into low- and high-frequency restoration pathways is highly effective. The wavelet-domain diffusion module reconstructs stable global structures while avoiding the computational burden and instability of image-domain diffusion. The proposed high-frequency refinement module further restores edges and textures that are otherwise blurred or lost under heavy noise. Together, these components enable WICD to deliver substantially higher fidelity than recent CNN-, and Mamba-based baselines, as reflected by both visual quality and large performance gains in PSNR/SSIM.

Remaining limitations include the reliance on multiple diffusion steps and the fixed Haar wavelet basis. Future work will explore faster deterministic solvers, learnable wavelet transforms, and broader anatomical validation. Overall, WICD offers a compact, efficient, and high-performing solution for LDCT denoising, providing strong noise suppression and structural accuracy suitable for clinical imaging settings.

REFERENCES

[1] S. V. Venkatakrishnan, K. A. Mohan, A. K. Ziabari, and C. A. Bouman, "Algorithm-driven advances for scientific ct instruments: From model-based to deep learning-based approaches," *IEEE Signal Processing Magazine*, vol. 39, no. 1, pp. 32–43, 2021.

[2] D. O. Baguer, J. Leuschner, and M. Schmidt, "Computed tomography reconstruction using deep image prior and learned reconstruction methods," *Inverse Problems*, vol. 36, no. 9, p. 094004, 2020.

[3] Y. Huang, O. Taubmann, X. Huang, V. Haase, G. Lauritsch, and A. Maier, "Scale-space anisotropic total variation for limited angle tomography," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 2, no. 4, pp. 307–314, 2018.

[4] C. Fang, X. Huang, and K. Xia, "Optimal design of fir filter based on improved artificial bee colony algorithm," in *2022 21st international symposium on distributed computing and applications for business engineering and science (DCABES)*. IEEE, 2022, pp. 155–158.

[5] L. Zhou, B. Morovati, D. Wang, Y. Xu, S. Han, S. Fan, C. Fang, Y. Shi, and H. Yu, "Co-retention featurepyramid network for low-dose ctdenoising via spatialand and frequency domain learning," *Available at SSRN 5718103*, 2025.

[6] L. Li, Y. Yang, C. Fang, and K. Xia, "A hybrid equilibrium optimizer algorithm and its optimization for lstm," in *2023 16th International Conference on Advanced Computer Theory and Engineering (ICACTE)*. IEEE, 2023, pp. 213–217.

[7] J. Fan, S. Fan, D. Jha, C. Fang, T. Zeng, H. Yu, and D. Wang, "Large language model evaluated stand-alone attention-assisted graph neural network with spatial and structural information interaction for precise endoscopic image segmentation," *arXiv preprint arXiv:2508.07028*, 2025.

[8] C. Fang, Y. Liu, B. Morovati, S. Han, Y. Shi, L. Zhou, S. Fan, and H. Yu, "Respf: Residual poisson flow generative model for efficient and physically consistent sparse-view ct reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2025.

[9] Y. Shi, S. Fan, C. Fang, S. Han, H. Li, L. Zhou, B. Morovati, D. Wang, and H. Yu, "Clinical metadata guided limited-angle ct image reconstruction," *arXiv preprint arXiv:2509.01752*, 2025.

[10] H. Li, S. Han, H. Mao, Y. Shi, C. Fang, J. Zhang, W. Wu, and H. Yu, "Cross-distribution diffusion priors-driven iterative reconstruction for sparse-view ct," *arXiv preprint arXiv:2509.13576*, 2025.

[11] C. Fang, Y. Liu, B. Morovati, S. Han, L. Zhou, and H. Yu, "Pocgm: Poisson-conditioned generative model for sparse-view ct reconstruction," *arXiv preprint arXiv:2511.13967*, 2025.

[12] E. Loli Piccolomini, D. Evangelista, and E. Morotti, "Deep guess acceleration for explainable image reconstruction in sparse-view ct," *Computerized Medical Imaging and Graphics*, vol. 123, p. 102530, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0895611125000394

[13] Şaban Öztürk, O. C. Duran, and T. Çukur, "Denomamba: A fused state-space model for low-dose ct denoising," *arXiv:2409.13094*, 2024.

[14] H. Guo, J. Li, T. Dai, Z. Ouyang, X. Ren, and S.-T. Xia, "Mambair: A simple baseline for image restoration with state-space model," in *European conference on computer vision*. Springer, 2024, pp. 222–241.

Noise and Anatomy Adaptive Convolution for Efficient Low-Dose CT Denoising

Minwoo Yu, Jongduk Baek, and Adam S. Wang

***Abstract*—To reduce radiation dose on patients, low-dose CT imaging techniques have significantly advanced. Specifically, while utilizing deep learning-based denoising networks for post-processing has become standard, the primary focus remains largely limited to learning the direct mapping between input noisy images and their clean target counterparts. This hinders the development of denoising architectures capable of utilizing CT imaging properties, in which noise intensity and its spatial correlations are highly dependent on the scan geometry, acquisition settings, and specific anatomy of the patient. Accordingly, we propose a modified convolution operation that adaptively modulates weights based on both noise and anatomical structures. It achieves high performance-to-computational efficiency, by simply replacing standard convolutional layers in existing networks. Furthermore, our proposed operation is implemented as a custom-engineered framework, optimizing memory overhead and accelerating inference speed in practice. Our code is fully available in this https://github.com/minwoo-yu/Noise-and-Anatomy-Adaptive-Convolution.git link.**

***Index Terms*—Anatomical Variance, Convolution with Adaptive Weight, CT image denoising, Noise Spatial Correlation.**

I. Introduction

Advancements in deep learning (DL)-based denoising methods for computed tomography (CT) imaging continue to reduce the burden of radiation dose on patients. Often, DL-based denoising model utilize low-dose CT (LDCT) images for input, and normal-dose CT (NDCT) images for ground-truth (GT) for training. Recently, a Noise2Noise (N2N)-based methodology - which adds independent noise into NDCT to address correlations with LDCT noise - has enabled higher quality noise reduction in CT imaging, equivalent to Noise2Clean [1], [2]. However, even when the X-ray tube current is consistent during scanning, the intensity and spatial correlations of noise vary significantly due to the scanned anatomical variations and acquisition settings, leading to non-stationary (i.e., spatially varying) noise properties. This is a consequence of the back-projection reconstruction process, where projection data from all view angles affect multiple pixels in the resulting CT image. These context-dependent noise characteristics violate the common assumption of independent and identically distributed (i.i.d.) noise, which is often made in denoising tasks. Consequently, the current lack of sufficient consideration for noise properties and anatomical variances suggests that most proposed models are not yet fully optimized for CT image data.

Recently, Kristof et al. [3] proposed the Noise-Augmented Deep Denoising (NADD) framework, which boosts denoising performance using augmented noise. NADD estimates augmented noise maps from LDCT projection data and utilizes them as a concatenated network input with the LDCT image. This demonstrated that the model can better account for the spatial correlation of noise solely through simple concatenation, without requiring any modifications to the model architecture. Nevertheless, applying NADD to the U-Net architecture [4], which is considered the standard model for denoising tasks, was not effective, and since the underlying computational operation remains unchanged, it cannot be considered a fundamental solution.

To address these limitations, we propose adjusting convolution weights using noise correlation and anatomical information. Our approach is inspired by the malleable convolution (MalleConv) operation, which applies spatial-varying convolution filters [5]. However, the MalleConv layer has several limitations: it is difficult to train, which restricts its use throughout the entire network; it lacks consideration for noise correlation; and it is not fully reproducible. On the contrary, our approach overcomes this limitation by enabling stable training and incorporating noise correlations in image space. Furthermore, our proposed architecture achieves high efficiency in terms of computational resource consumption. As illustrated in Fig. 1, adopting our method delivers a higher peak signal-to-noise ratio (PSNR) relative to multiply-accumulate operations (MACs) compared to traditional convolution-based models.

II. Methods

A. Preprocessing to obtain a Noise Correlation map

In general, the measurement noise, P_{noise} in the detector can be expressed as the summation of Poisson noise Pois and Gaussian noise N, as follows:

$$P_{noise} \sim \text{Pois}(P_{clean}) + N(0, \sigma_e^2), \tag{1}$$

where P_{clean} denotes the number of photons, and σ_e^2 denotes the variance value for electronic noise. Next, setting

Minwoo Yu and Jongduk Baek are with the Department of Artificial Intelligence, Yonsei University, South Korea, e-mail: (ymw9754@yonsei.ac.kr, jongdukbaek@yonsei.ac.kr)

Adam S. Wang is with the Department of Radiology and the Department of Electrical Engineering, Stanford University, Stanford, CA, 94305 USA, e-mail: (adamwang@stanford.edu)

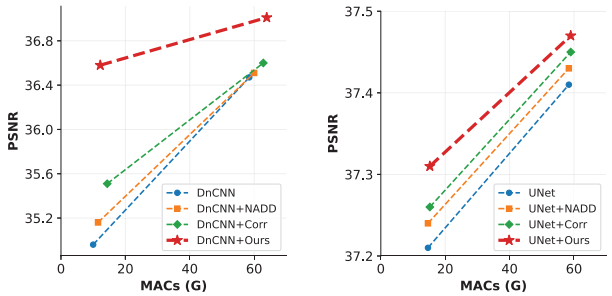


Fig. 1: MACs vs PSNR on the 10% dose denoising task. Our proposed approach demonstrates superior efficiency, proving particularly beneficial in lightweight DnCNN (left) and U-Net (right) configurations.

P_{noise} as the NDCT data P_{NDCT}, we can estimate the LDCT data by adding noise to P_{NDCT} as follows:

$$\begin{aligned} P_{LDCT} &\sim \text{Pois}(P_{clean} \cdot d) + N(0, \sigma_e^2) \\ &\approx d(P_{NDCT} + a(N(0, P_{NDCT}) + bN(0, \sigma_e^2)) \\ &\text{where } a = \sqrt{\frac{1}{d} - 1},\ b = \sqrt{\frac{1}{d} + 1} \end{aligned} \tag{2}$$

where $0 < d < 1$ denotes the relative radiation dose level to the NDCT data.

In general, CT denoising tasks only utilize LDCT images $I_{LDCT} \in \mathbb{R}^{H \times W}$ for input, where H and W denote the height and width of the image, respectively. However, the NADD framework extracts augmented noise maps $X_N \in \mathbb{R}^{N \times H \times W}$ from the LDCT data to generate noise-only images; these are subsequently utilized alongside the LDCT image as input as follows:

$$\begin{aligned} X_n &\sim FBP(-\log(P_{LDCT}/P_{clean})) \\ &\approx FBP(-\log((P_{LDCT} + N(0, P_{LDCT}) \\ &+ N(0, \sigma_e^2))/P_{LDCT})) \\ &\text{where } n = 0, 1, ...N \end{aligned} \tag{3}$$

However, NADD simply utilizes augmented noise maps as model inputs without any additional processing, which limits the ability of the model to exploit the region-dependent spatial correlation and intensity of noise. To handle this issue, as shown in Fig. 2, we estimate correlation map $\mathbf{C} \in \mathbb{R}^{p \times p \times H \times W}$, where p denotes the patch size, which is obtained by using noise maps X_N as follows:

$$\begin{aligned} \mathbf{C}(i, j, x, y) &= \frac{1}{Nk^2} \sum_{n=1}^{N} \sum_{(u,v) \in \Omega} \mathbf{X}_n(x + u, y + v) \\ &\cdot \mathbf{X}_n(x + u + i - \lfloor p/2 \rfloor, y + v + j - \lfloor p/2 \rfloor), \\ &\text{where } \Omega = \{-\lfloor k/2 \rfloor, \ldots, \lfloor k/2 \rfloor\}^2 \end{aligned} \tag{4}$$

where k denotes the spatial kernel size. Using Eq. 4, intensity and spatial correlation of CT noise are extracted within a localized boundary defined by the patch size p.

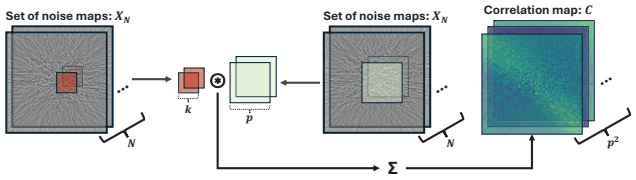


Fig. 2: Illustration of estimating the Noise Correlation Map.

B. *Convolution with Adaptable Weight*

Even if the correlation map is obtained from the noise map, simply providing this information to the model alongside the LDCT image, as in NADD, does not fully exploit the potential of the noise statistics. We note that even while CT images are acquired at a static tube current, noise intensity and correlation significantly vary across different regions; nevertheless, weight values of the convolution layers remain fixed. Therefore, inspired by Malleable Convolution [5], which adapts convolution weights based on features, we develop a structure that dynamically adjusts convolution weights according to both the feature map and correlation map. Our structure fuses extracted features from anatomy and correlation maps to generate adaptive kernel weight $W \in \mathbb{R}^{k^2 \times C \times H/8 \times W/8}$ and bias $B \in \mathbb{R}^{C \times H/8 \times W/8}$, where k and C denote the kernel size and the number of channels, respectively. Next, the output feature is calculated as a weighted sum of the input feature and adaptive kernel, employing an approach similar to depthwise convolution. Note that to enhance computational efficiency and performance of kernel generation, the weight and bias values are initially generated at an $8\times$ downscaled resolution and subsequently upsampled via bilinear interpolation.

However, generating output features after fully completing the interpolation leads to a severe memory footprint, which is not practical in clinical systems. To address this limitation, we develop a customized module that partitions the operation into blocks and performs on-the-fly interpolation only for the regions required for the weighted sum of each block; this minimizes both memory consumption and data loading time from the GPU memory. This overall pipeline is illustrated in Fig. 3.

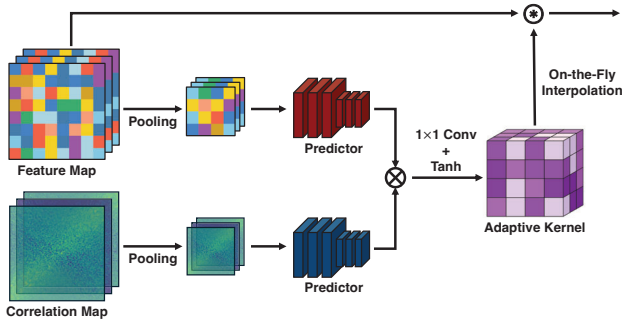


Fig. 3: Illustration of Weight-Adaptive Convolution.

	-		+NADD		+Corr		+Ours	
dose d	10%	25%	10%	25%	10%	25%	10%	25%
NDCT	23.11 / 0.5687	27.32 / 0.7376	-	-	-	-	-	-
DnCNN-lite	34.96 / 0.9472	37.17 / 0.9652	35.16 / 0.9499	37.33 / 0.9663	35.51 / 0.9534	37.64 / 0.9684	**36.58 / 0.9627**	**38.26 / 0.9717**
DnCNN	36.47 / 0.9621	38.16 / 0.9710	36.51 / 0.9622	38.23 / 0.9715	36.60 / 0.9627	38.30 / 0.9717	**37.01 / 0.9653**	**38.55 / 0.9728**
U-Net-lite	37.21 / 0.9666	38.63 / 0.9732	37.24 / 0.9668	38.66 / 0.9733	37.26 / 0.9669	38.68 / 0.9734	**37.31 / 0.9670**	**38.71 / 0.9735**
U-Net	37.41 / 0.9675	38.79 / 0.9737	37.43 / 0.9676	38.81 / **0.9738**	37.45 / 0.9676	38.82 / 0.9737	**37.47 / 0.9677**	**38.82** / 0.9737

TABLE I: Quantitative evaluation for each configuration in terms of PSNR / SSIM. The best results are indicated in **bold**, and the second-best results are underlined.

C. Configuration of Denoising Network

Similar to depthwise convolution, our proposed structure lacks interaction between channels. While it can be complemented by an additional 1×1 convolution to account for inter-channel dependencies, relying solely on 1×1 convolution may result in insufficient consideration of inter-channel relationships. To overcome this, we employ a sequential stacking of the proposed weight-adaptive convolution and standard convolution layers.

In addition, many denoising network architectures are inspired by the U-Net [4] structure, which resizes feature maps for high-quality and robustness. This hierarchical design results in varying feature size across levels; therefore, we apply average pooling to resize the noise information accordingly.

III. Experiments

In this work, we used the Mayo Clinic Low-Dose CT Challenge 2016 [6] dataset, which contains 10 patient scans reconstructed at a 512×512 resolution. We used normal-dose images as the clean GT dataset, from which clean projection data were acquired under a typical fan-beam CT geometry using the LivermorE AI Projector (LEAP) [7] framework. For noise insertion, the number of incident photons for the normal-dose setting and the electron noise standard deviation σ_e were set to 5×10^5 and 2, respectively. To evaluate the model across multiple scenarios, the dose levels d were set to 10% and 25%. For the training, we follow the N2N methodology by generating target images containing noise independent of the LDCT, enabling training without clean GT data.

For network architectures, we evaluated DnCNN [8] and U-Net [4] by employing four different configurations: (1) baseline models without modification, (2) models with increased number of input channels for augmented noise maps (+NADD), (3) models with increased number of input channels for the correlation map (+Corr), (4) selected convolutional layers are replaced with our weight-adaptive convolution (+Ours). Specifically, the number of augmented noise maps N was set to 10, with the kernel size k and patch size p for noise correlation calculation configured to 3 and 5, respectively. To maximize computational efficiency, the weight-adaptive convolution operation was implemented using Triton [9] to leverage GPU parallelization.

IV. Result

TABLE I shows the PSNR and structural similarity index measure (SSIM) values for each model configuration. In the case of NADD, the relatively slight gains in image quality indicate that simply apply noise augmentation fails to capture the intricate noise correlation required for denoising. In contrast, as proposed in this study, the process of converting noise maps into a correlation map facilitates a better consideration of the spatial correlation of noise, leading to improved image quality in most cases; this improvement is particularly pronounced in at relatively low radiation dose levels and in lightweight network configurations. Next, adopting our proposed convolution operation consistently improves both PSNR and SSIM values across all experimental settings. These results are achieved without increasing the number of MACs as shown in Fig. 1.

Next, qualitative evaluation presented in Fig. 4 confirms the superiority of our approach in terms of visual fidelity. In the case of DnCNN-lite, due to the insufficient number of convolutional layers, most configurations exhibit residual artifacts and degraded image texture quality. On the other hand, our proposed structure maintains superior quality thanks to its adaptive weights, demonstrating exceptional efficiency in terms of image quality per computational unit. U-Net-lite generally outperforms DnCNN in terms of overall texture quality; however, it fails to preserve complex structures, leading to low visual fidelity in several regions. Our approach overcomes structural inaccuracies by effectively utilizing both anatomical features and noise spatial correlation.

Batch size	Backbone	Memory (GB)	Runtime (ms)
8	Naive	16.51	324.50 ± 0.46
	Ours	3.19	193.70 ± 0.83
16	Naive	33.02	645.93 ± 0.93
	Ours	6.38	380.82 ± 0.63
32	Naive	Out-of-Memory	
	Ours	12.76	757.66 ± 1.01

TABLE II: GPU memory consumption and runtime measurements for various batch sizes and U-Net-based backbone configurations. All evaluations are performed on 512×512 images using an NVIDIA RTX A6000 GPU.

Although our approach shows superior efficiency in terms of PSNR vs MACs, it does not guarantee faster execution in real-world application. In practice, a naive PyTorch implementation suffers from significant time and memory overhead. This is because storing all interpolated weights results in massive memory consumption, and subsequent retrieval of stored values is severely bottlenecked due to limited memory transmission bandwidth. To address this limitation, we implemented an on-the-fly interpolation

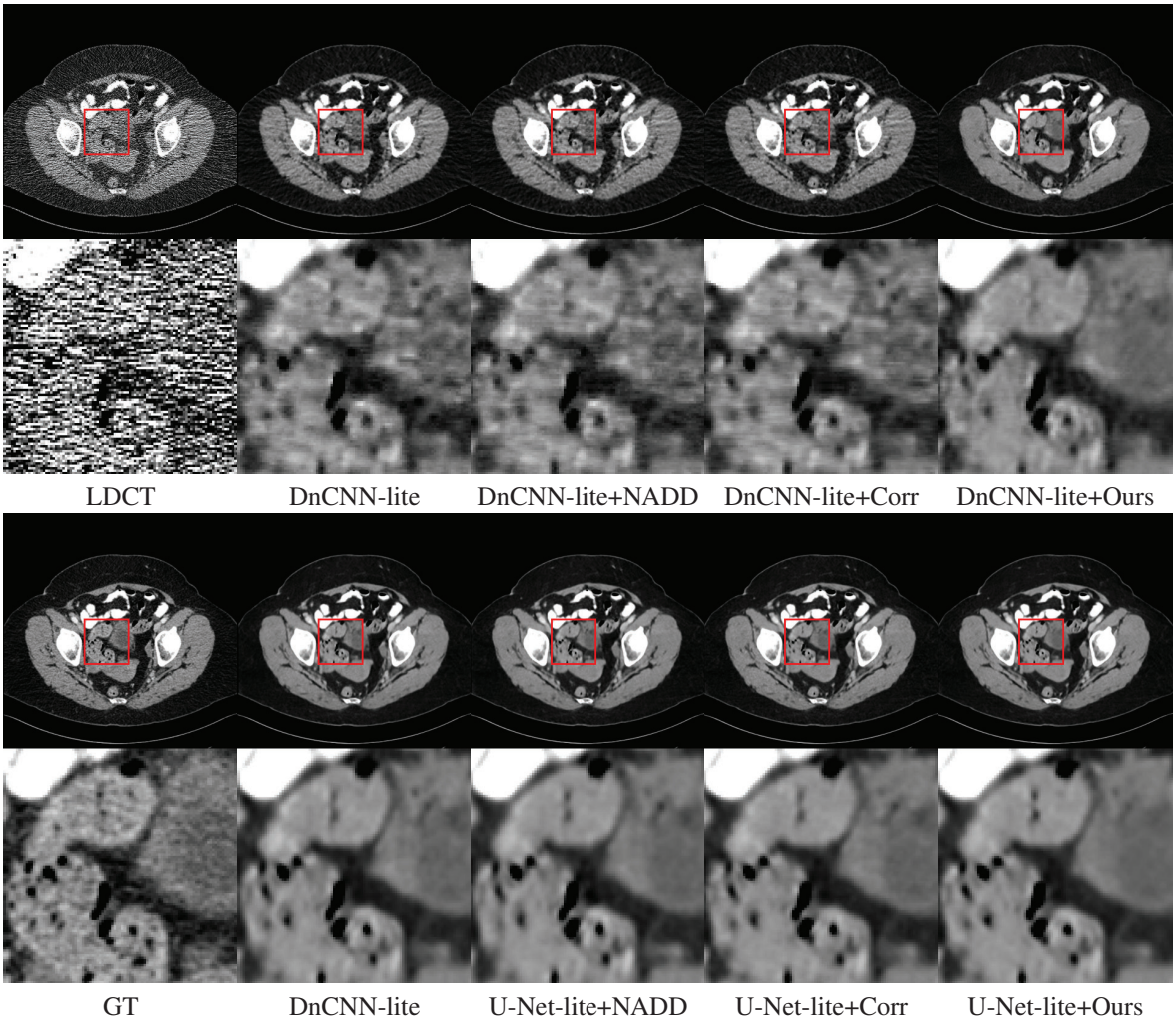


Fig. 4: Qualitative evaluation of 10% denoising task with zoomed ROI images. Red boxes indicate the ROI locations. The display window is [-135, 215] HU.

algorithm for modified convolution operation, and TABLE II demonstrates its efficiency. Our implementation is over five times more memory-efficient and approximately twice as fast in terms of runtime compared to the naive implementation.

V. Conclusion

In this study, we propose an approach that improves the efficiency and efficacy of convolution-based denoising by enabling them to better incorporate spatial noise correlation and anatomical information. Our method exhibits a significantly higher performance-per-computation efficiency than standard convolutions, providing a more optimized backbone for CT imaging tasks. Future work will focus on developing a framework that enables efficient parallelization for rapid noise map generation and further improving memory access patterns for our proposed operations.

Acknowledgment

This work was partially supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (RS-2023-00240135, and RS-2025-00553670), and Institute of Information & Communications Technology Planning & Evaluation (IITP) Grant funded by MSIT (No. RS-2020-II201361, Artificial Intelligence Graduate School Program (Yonsei University))

References

[1] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.

[2] S. Wang and A. S. Wang, "Simulating arbitrary dose levels and independent noise image pairs from a single ct scan," in *7th International Conference on Image Formation in X-Ray Computed Tomography*, vol. 12304. SPIE, 2022, pp. 460–466.

[3] G. Kristof, E. Eulig, and M. Kachelrieß, "Noise-augmented deep denoising: A method to boost ct image denoising networks," *Medical Physics*, vol. 52, no. 10, p. e18121, 2025.

[4] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[5] Y. Jiang, B. Wronski, B. Mildenhall, J. T. Barron, Z. Wang, and T. Xue, "Fast and high quality image denoising via malleable convolution," in *European Conference on Computer Vision*. Springer, 2022, pp. 429–446.

[6] T. R. Moen, B. Chen, D. R. Holmes, X. Duan, Z. Yu, L. Yu, S. Leng, J. G. Fletcher, and C. H. McCollough, "Low-dose ct image and projection dataset," *Medical Physics*, vol. 48, no. 2, pp. 902–911, Feb. 2021, epub 2020 Dec 16.

[7] H. Kim and K. Champley, "Differentiable forward projector for x-ray computed tomography," 2023. [Online]. Available: https://arxiv.org/abs/2307.05801

[8] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[9] P. Tillet, H. T. Kung, and D. Cox, "Triton: an intermediate language and compiler for tiled neural network computations," in *Proceedings of the 3rd ACM SIGPLAN International Workshop on Machine Learning and Programming Languages*, ser. MAPL 2019, 2019, p. 10–19.

Lag-Induced Spectral Contamination in Spectral CBCT: A Comparative Study of Fast kV-Switching and Dual-Layer Flat-Panel Detector Technologies

Yuting Chen, Hao Zhou, Wenxin Mo, Hewei Gao*

***Abstract*—Lag-induced spectral contamination compromises material decomposition (MD) accuracy in spectral cone-beam CT (CBCT). This study compares Fast kV-Switching (FKS) and Dual-Layer (DL) flat-panel detector (FPD) technologies using a simulation framework that models realistic lag conditions. We evaluated both matched and mismatched lag cases on two virtual phantoms. Results reveal distinct error mechanisms for each technology. DL is highly sensitive to lag mismatch between layers: in the elliptical phantom, mismatch shifted the mean deviation from 1.3 to -6.4 mg/mL, whereas FKS showed only a 0.4 mg/mL shift. However, FKS suffers from spectral cross-contamination due to alternating kVp exposures. Although expectation-maximization (EM)-based spectral calibration improves accuracy for uniform objects, it cannot compensate for history-dependent errors in non-uniform geometries. Under lag mismatch, changing from cylindrical to elliptical geometry increased the absolute mean deviation from 1.5 to 6.0 mg/mL for EM-calibrated FKS, compared to 4.5 to 6.4 mg/mL for DL.**

***Index Terms*—Lag, spectral contamination, spectral CBCT, fast kV-switching, dual-layer flat-panel detector.**

I. Introduction

SPECTRAL CBCT can be achieved through different acquisition strategies. Among these, two prominent implementation approaches are widely recognized. Fast kV-Switching (FKS) rapidly alternates tube voltage between consecutive frames to acquire low- and high-energy projections [1]. The dual-layer (DL) configuration allows for the preferential detection of lower-energy photons in the top layer and higher-energy photons in the bottom layer, enabling simultaneous spectral differentiation [2].

However, detector lag poses a fundamental challenge to both technologies. In amorphous silicon (a-Si) flat-panel detectors, lag arises from multiple sources: scintillator afterglow due to delayed luminescence in the phosphor crystals, charge trapping and release in the photodiode layer, and readout electronics delays [3]. These mechanisms cause residual signals from previous frames to persist into subsequent acquisitions.

In FKS, lag-retained signals from one energy spectrum directly contaminate measurements at the opposing energy, causing spectral cross-contamination. In DL, differences in scintillator thickness and energy deposition depth lead to mismatched lag characteristics between the top and bottom layers, resulting in inconsistent spectral contamination between energy channels.

This study presents a comparative simulation of lag-induced spectral contamination in FKS and DL technologies, aiming to provide practical guidance for spectral CBCT system design.

This project was supported in part by Beijing Natural Science Foundation (No. L252021), and in part by the National Key R&D Program of China under grant No. 2024YFC2417500. *(Corresponding author: Hewei Gao)*

The authors are with the Department of Engineering Physics, Tsinghua University, Beijing 100084, China, and are also with the Key Laboratory of Particle and Radiation Imaging, Ministry of Education, Tsinghua University, Beijing 100084, China (e-mail: hwgao@tsinghua.edu.cn).

II. Method

A. Lag Modeling and Its Impact on Projection Data

In dual-energy material decomposition, the projection data at low-energy (P_L) and high-energy (P_H) are typically modeled as:

$$P_{\mathrm{L}} = -\ln\left(\frac{\int S_{\mathrm{L}}(E)e^{-\mu_1(E)\int \rho_1 \mathrm{dl}-\mu_2(E)\int \rho_2 \mathrm{dl}}\mathrm{dE}}{\int S_{\mathrm{L}}(E)\mathrm{dE}}\right), \quad (1)$$

$$P_{\mathrm{H}} = -\ln\left(\frac{\int S_{\mathrm{H}}(E)e^{-\mu_1(E)\int \rho_1 \mathrm{dl}-\mu_2(E)\int \rho_2 \mathrm{dl}}\mathrm{dE}}{\int S_{\mathrm{H}}(E)\mathrm{dE}}\right), \quad (2)$$

where $S_L(E)$ and $S_H(E)$ are the incident X-ray spectra at low and high energies, $\mu_1(E)$ and $\mu_2(E)$ are the mass attenuation coefficients of the two basis materials, and ρ_1 and ρ_2 are the material densities along the ray path.

This formulation assumes an ideal, memory-free detector response. In practice, however, amorphous-silicon (a-Si) FPDs exhibit significant detector lag in X-ray imaging. The discrete-time impulse response function (IRF) can be described as a multi-exponential model:

$$h(k) = b_0\delta(k) + \sum_{n=1}^{N} b_n \exp(-a_n k) \quad (3)$$

where k is the discrete-time variable representing frame index, N is the number of exponential terms (typically 2-4) [4], [5], a_n controls the decay rate, and b_n represents the residual fraction for each component.

The presence of detector lag effectively alters the effective spectrum, introducing bias into both P_L and P_H, and more critically, creating a mismatch between the paired energy projections. The resulting inconsistency in the projection domain propagates directly into material decomposition errors, as decomposition algorithms assume that P_L and P_H correspond to well-defined, paired spectral measurements. The pattern of this corruption differs across CBCT architectures, as detailed in the following theoretical analysis.

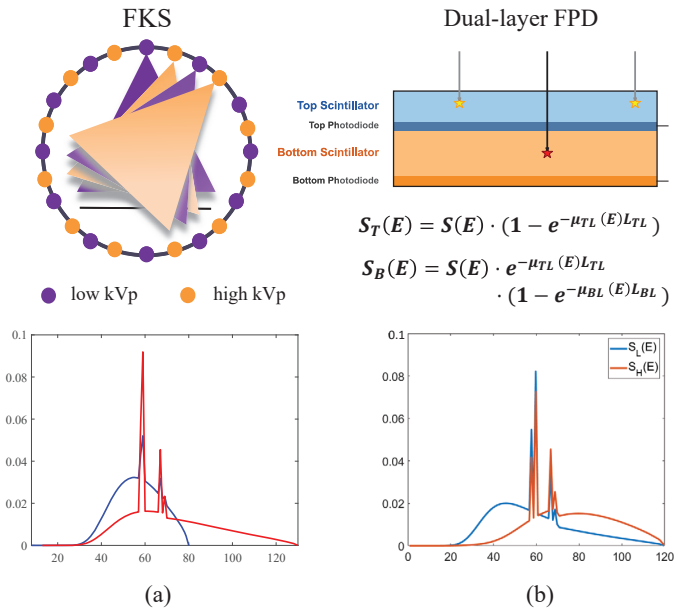


Fig. 1. Schematic diagram of spectral CT data acquisition using fast kV-switching (FKS) and dual-layer flat-panel detector technologies, along with their corresponding low- and high-energy spectra.

B. Theoretical Analysis of Spectral Contamination

To quantitatively elucidate the distinct failure mechanisms of different spectral architectures, we formulated the propagation of lag error based on the discrete convolution model. Let $S_{true}(k)$ denote the ideal incident photon flux at view k, and $S_{meas}(k)$ be the measured signal. The lag-contaminated signal was expressed as

$$S_{meas}(k) = \sum_{n=0}^{\infty} h(n) S_{true}(k-n), \tag{4}$$

where $h(n)$ represents the normalized detector IRF.

1) Fast kV-Switching Acquisition: Let view k be acquired under kVp setting A, where $A \in \{L, H\}$ denotes the low- or high-kVp acquisition state. In an alternating FKS acquisition, previous views alternate between setting A and the opposite setting B, with $B \neq A$. The corresponding impulse responses are denoted by $h_A(n)$ and $h_B(n)$, respectively. The measured signal at view k can then be written into the primary signal and residual lag contributions from prior views:

$$\begin{aligned} S_{meas}^{FKS}(k) = & \underbrace{h_A(0) S_{true}^{A}(k)}_{\text{Primary Signal}} \\ & + \underbrace{\sum_{m=0}^{\infty} h_B(2m+1) S_{true}^{B}(k-(2m+1))}_{\text{Cross-kVp Contamination}} \\ & + \underbrace{\sum_{m=1}^{\infty} h_A(2m) S_{true}^{A}(k-2m)}_{\text{Self-kVp Contamination}}. \end{aligned} \tag{5}$$

The *Cross-kVp Contamination* term represents residual signals from prior views acquired under the opposite tube-voltage setting B being mixed into the current-view measurement acquired under setting A, where $A, B \in \{L, H\}$ and $B \neq A$. This mixing is weighted by the lag response $h_B(n)$ associated with the opposite kVp state. As a result, the effective spectral characteristics of the measured projection are altered, which can introduce bias in material decomposition. In contrast, the *Self-kVp Contamination* term represents temporal mixing of signals from prior views acquired under the same tube-voltage setting A, which occur at non-consecutive views in an alternating FKS acquisition. The asymmetry in lag characteristics between the low- and high-kVp states, i.e., $h_L(n) \neq h_H(n)$, further compounds these errors by introducing kVp-dependent temporal distortions.

2) Dual-Layer and Sequential Acquisitions: In contrast to FKS, the Dual-Layer and Sequential configurations maintain spectral temporal continuity. For a DL detector, the top and bottom layers acquire data simultaneously but are physically separated. Although the layers may exhibit different lag response characteristics (h_{Top} vs. h_{Bot}), the signal history within each layer originates exclusively from its own spectral domain. Thus, for a given layer L (Top or Bottom), the measured signal is

$$S_{meas}^{DL,L}(k) = h_L(0) S_{true}^{L}(k) + \underbrace{\sum_{n=1}^{\infty} h_L(n) S_{true}^{L}(k-n)}_{\text{Self-Contamination Only}}. \tag{6}$$

Similarly, in Sequential acquisition, since the high and low energy scans are performed independently, the lag effect manifests purely as *self-contamination* within each kVp scan. While this avoids the spectral aliasing inherent to FKS, both DL and Sequential modes remain susceptible to lag mismatch errors. If the lag characteristics differ between the two energy levels ($h_{Top}(n) \neq h_{Bot}(n)$ for DL, or $h_{Low-kVp}(n) \neq h_{High-kVp}(n)$ for Sequential), the paired energy data experience unequal temporal corruption, which subsequently compromises material decomposition accuracy.

C. FKS Spectral Calibration

For FKS, we applied a spectral calibration method based on step-wedge measurements. This method utilizes the Expectation-Maximization (EM) algorithm [6] to estimate the dual-energy spectra in the presence of spectral contamination.

D. Material Decomposition

In this study, we employ a polynomial function method to generate the basis material projections, using the following formulas

$$\begin{aligned} M_1 &= \sum_{ij} p_{ij} P_{\mathrm{L}}^{i} P_{\mathrm{H}}^{j}, \\ M_2 &= \sum_{ij} q_{ij} P_{\mathrm{L}}^{i} P_{\mathrm{H}}^{j}. \end{aligned} \tag{7}$$

where M_1 and M_2 are the line integrals of the densities of two basis materials, and $\{p_{ij}\}$ and $\{q_{ij}\}$ can be predetermined by simulations or measurements.

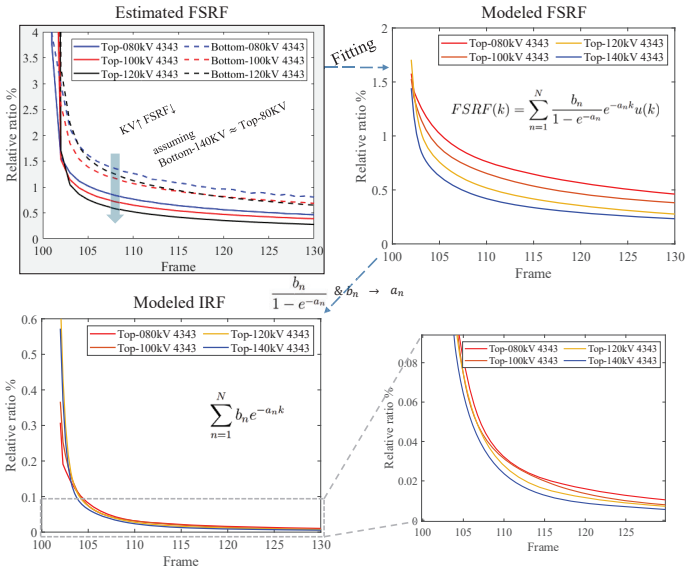


Fig. 2. IRF determination by multi-exponential fitting of experimentally measured FSRF data.

III. Experiment Setup

A. Simulation Conditions

The simulation was configured with a source-to-detector distance of 1000 mm and a source-to-origin distance of 750 mm. A total of 500 projections were acquired over 360°, with Poisson noise added. Spectra at 80 kVp and 140 kVp were simulated using SpekCalc. The unattenuated spectral separation [1] is 15.6 keV for DL and 15.0 keV for FKS. Two phantoms were constructed: a circular cylinder (16 cm diameter) and an elliptical cylinder (axes of 22.5 cm and 13 cm), both containing five 2 cm iodine inserts. As shown in Fig. 3, six ROIs were defined: ROIs 1-6 contain iodine concentrations of 10, 0, 5, 10, 10, and 20 mg/ml, respectively.

B. Detector Lag Model Configuration

The lag coefficients used in this simulation were derived from experimental measurements performed on a FPD (Varex Imaging XRD 4343RF, Salt Lake City, UT, USA) with the acquisition speed of 20Hz. The FSRF curves were measured and used to solve the $\{a_n\}$ and $\{b_n\}$ in IRF in Eq. (3).

Based on the measurement results shown in Fig. 2, which indicated that (1) the bottom layer exhibits greater lag than the top layer and (2) lag decreases with increasing tube voltage, we configured the lag parameters as follows:

Mismatched Lag Condition: The 140 kVp top layer was assigned the minimum lag magnitude, while the 140 kVp bottom layer and the 80 kVp top layer were assumed to exhibit equivalent lag characteristics. This configuration reflects realistic detector behavior where energy-dependent and depth-dependent lag variations coexist.

Matched Lag Condition: As a control scenario, the 140 kVp top layer lag was artificially set equal to the other two conditions, eliminating inter-channel lag discrepancies. This idealized configuration serves to isolate the effects of lag mismatch from other sources of spectral degradation.

IV. Results

Fig. 3 presents a comprehensive comparison of water-iodine MD performance across four spectral CT acquisition modes under both matched and mismatched lag conditions. The first column displays the phantom geometry and ROI definitions. Columns 2-5 show the reconstructed water basis images for DL, Sequential, FKS without spectral calibration with lag (w/o SpecCali), and FKS with spectral calibration with lag (w/ SpecCali), respectively. Columns 6-7 provide quantitative analysis of ROI 6 (ground truth: 20 mg/ml iodine concentration): column 6 presents the mean value bias for both water and iodine concentrations, and column 7 shows the corresponding CNR measurements for both materials.

A. Overall Performance

As shwon in Table I, the DL mode (column 2) achieved highest MD accuracy in most cases, while exhibited the lowest CNR among all modes due to inherent photon starvation: the top layer pre-filters the beam and significantly attenuates photon flux reaching the bottom layer.

The Sequential mode (column 3) was included as a control for FKS, excluding the spectral cross-talk due to kVp-switching.

The FKS (w/o SpecCali) mode (column 4) suffered from severe spectral cross-contamination, resulting in severe artifacts and significant bias in ROI mean values.

Demonstrating effective recovery through spectral calibration, the FKS (w/ SpecCali) mode (column 5) suppressed mean value bias to levels comparable to, and in some cases slightly lower than, the Sequential reference. However, vertical banding artifacts in the elliptical phantom became more pronounced than in other modes, suggesting that the calibration procedure could introduce new systematic errors.

Compared to matched lag conditions, CNRs in all cases slightly improved under mismatch, potentially due to the smaller 140 kVp top layer lag.

B. Sensitivity to Lag Mismatch

For the DL mode (column 2), transitioning from matched to mismatched conditions induced significant degradation. The water basis images Fig. 3 exhibited characteristic shading and cupping artifacts. The quantitative results in columns 6-7 confirmed this observation: the mean value bias increased substantially for both water and iodine.

The FKS modes (columns 4-5) showed relatively smaller sensitivity to the mismatch condition compared to DL.

In particular, in the elliptical phantom, lag mismatch shifted the DL mean deviation from 1.3 to -6.4 mg/mL compared to the matched case, whereas the absolute deviation for FKS (w/ SpecCali) increased from 5.6 to 6.0 mg/mL.

C. Influence of Object Geometry

As shwon in Table I, under lag match conditions, transitioning from the cylindrical to the elliptical phantom increased the absolute mean deviation in the DL water ROI from 0.0 to 1.3 mg/mL, and from 1.6 to 5.6 mg/mL for FKS (w/ SpecCali);

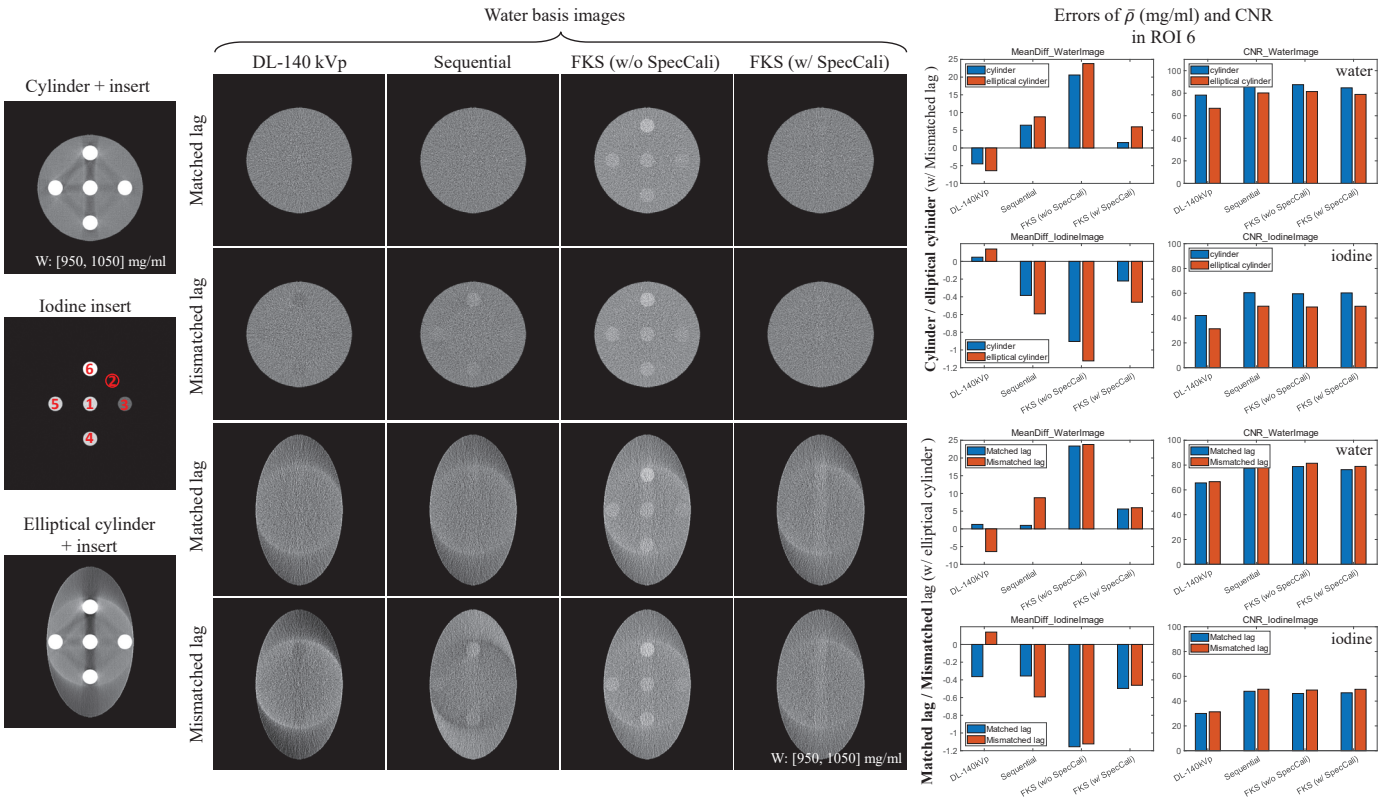


Fig. 3. Water basis images of the cylinder and elliptical cylinder water phantom with five iodine inserts under matched and mismatched lag conditions, along with quantitative analysis of errors of measured mean value and CNR in water/iodine basis images. ROIs 1-6 contain iodine concentrations of 10, 0, 5, 10, 10, and 20 mg/ml, respectively.

TABLE I
ERRORS OF MEASURED MEAN BASIS MATERIAL DENSITIES IN ROI 6

Water (ground truth of ρ_{water} : 1000 mg/mL)

Lag case	Phantom case	DL-140 kVp	Sequential	FKS (w/o SpecCali)	FKS (w/ SpecCali)
Matched	Cylindrical	0.0	0.5	20.6	1.6
	Elliptical	1.3	1.0	23.4	5.6
Mismatched	Cylindrical	**-4.5**	6.4	20.6	**1.5**
	Elliptical	**-6.4**	8.8	23.8	**6.0**

Iodine (ground truth of ρ_{iodine} in ROI 6: 20 mg/mL)

Lag case	Phantom case	DL-140 kVp	Sequential	FKS (w/o SpecCali)	FKS (w/ SpecCali)
Matched	Cylindrical	-0.21	-0.23	-0.93	-0.25
	Elliptical	-0.36	-0.36	-1.16	-0.50
Mismatched	Cylindrical	0.05	-0.38	-0.90	-0.22
	Elliptical	0.14	-0.59	-1.12	-0.46

under lag mismatch conditions, the absolute mean deviation increased from 4.5 to 6.4 mg/mL for DL, and from 1.5 to 6.0 mg/mL for FKS (w/ SpecCali).

V. CONCLUSIONS

This study compares lag-induced spectral contamination in FKS and DL spectral CBCT technologies. Key findings include: (1) DL is highly sensitive to top–bottom layer lag mismatch, whereas FKS (w/ SpecCali) is more sensitive to phantom geometry variations, causing artifacts and increased MD bias. (2) Given comparable energy separation, DL in most cases achieves MD accuracy comparable to or better than FKS (w/ SpecCali). (3) FKS suffers from spectral cross-contamination. Without spectral calibration, FKS shows severe artifacts and large quantitative errors. EM-based spectral calibration could improve MD accuracy for uniform phantoms but shows reduced effectiveness for non-uniform geometries, and may introduce additional systematic errors.

REFERENCES

[1] H. Zhou, L. Zhang, Z. Wang, and H. Gao, "Fast kv-switching and dual-layer flat-panel detector enabled cone-beam ct joint spectral imaging," Physics in Medicine & Biology **69**(11), 115005 (2024).

[2] Z. Wang, H. Zhou, S. Gu, Y. Xia, H. Liao, Y. Deng, and H. Gao, "Dual-energy head cone-beam ct using a dual-layer flat-panel detector: hybrid material decomposition and a feasibility study," Medical Physics **50**(11), 6762–6778 (2023).

[3] J. Starman, J. Star-Lack, G. Virshup, E. Shapiro, and R. Fahrig, "A nonlinear lag correction algorithm for a-si flat-panel x-ray detectors," Medical Physics **39**(10), 6035–6047 (2012).

[4] J. Hsieh, O. Gurmen, and K. F. King, "Investigation of a solid-state detector for advanced computed tomography," IEEE transactions on medical imaging **19**(9), 930–940 (2000).

[5] N. Mail, D. Moseley, J. Siewerdsen, and D. Jaffray, "An empirical method for lag correction in cone-beam ct," Medical physics **35**(11), 5187–5196 (2008).

[6] E. Y. Sidky, L. Yu, X. Pan, Y. Zou, and M. Vannier, "A robust method of x-ray source spectrum estimation from transmission measurements: Demonstrated on computer simulated, scatter-free transmission data," Journal of applied physics **97**(12) (2005).

Low-Dose CT Image Denoising Using Data-coupled Flow Matching

Bangyan Huang, Jinyi Qi, and Jian Zhou

***Abstract*—Diffusion models have achieved state-of-the-art performance in CT image denoising. However, they suffer from low inference speed due to the need to sample stochastic differential equations (SDE) with hundreds of steps. In this study, we leverage the flow matching model to drive an ordinary differential equation (ODE) for faster sampling. The ODE is constructed using coupled low- and normal-dose image pairs. The resulted sampling paths are shorter and thus enable faster sampling compared to conditional diffusion models. Experiments on 3D images demonstrate that our method achieves superior image quality using only five ODE steps, outperforming diffusion models.**

***Index Terms*—flow matching, ODE, denoising, diffusion model**

I. Introduction

PROBABILITY flow ordinary differential equation (ODE) has recently been established as a fast alternative to diffusion models for image generation [1], [2]. They learn a time-dependent velocity field for sample generation via flow matching (FM). As a result, a probability flow $p_t(\mathbf{x}), t \in [0, 1]$ is constructed with p_0 being the prior distribution and p_1 being the target distribution. FM provides a learning framework through the conditional velocity that bridges the prior and target data samples. The flexibility in constructing the conditional velocity leads to diverse forms of the probability flows. More importantly, distinct from diffusion models, FM models do not assume a particular prior distribution. Thus, for the image generation task, several studies have constructed a coupling between the training samples and the random samples from the prior distribution [3]–[5]. These techniques can provide a straighter sampling path that is easier to solve. Furthermore, Albergo *et al.* proposed a framework for general image restoration [6], which leverages paired corrupted and clean data to learn a velocity field that directly flows from the domain of corrupted data to that of clean data. This method significantly shortens the sampling path compared to previous conditional diffusion models that learn the conditional score function and usually start the generation process from Gaussian noise.

Previously, numerous deep learning-based denoising methods have been proposed under the supervised setting for CT images. CNNs and vision transformers are the primary network backbones [7]. They are mostly trained with a regression loss, e.g., MSE loss, between the network output and the target. This learning paradigm does not guarantee that the final image will reside on the correct manifold. Generative models can be used to address this issue. GAN, as a classical generative model, has been introduced to CT image denoising [8]. The discriminator of a GAN can help to improve the realism of the generated images. With the advent of diffusion models, SDEs have become the predominant technique for CT image denoising [9]. They typically leverage the low-dose CT images (LDCT) as the condition and learn a conditional score function to solve the SDEs for high-quality CT images. One of their limitations is the low inference speed due to the need to perform network inference for each step, with the total number being on the order of hundreds in most applications. Several studies have proposed to accelerate sampling through diffusion bridges [5], [10] and ODEs [11], which have also been applied to CT image denoising [9]. Although they reported a fast sampling with 15 steps, the image quality is significantly lower when compared to the standard DDPM or DDIM with sufficiently many sampling steps.

In this work, we apply FM for CT image denoising. The probability flow is constructed directly from the low-dose CT domain to the high-dose CT domain. Our experiments show that the FM model outperforms the conditional DDPM model, while requiring only five ODE steps.

II. Method

A. Probability flow ODE

Given a velocity field $\mathbf{v}(\mathbf{x}_t, t)$, samples are generated by solving an ODE

$$\frac{d\mathbf{x}_t}{dt} = \mathbf{v}(\mathbf{x}_t, t), \tag{1}$$

where $t \in [0, 1]$. The ODE can be solved by integrating (1): $\mathbf{x}_t = \mathbf{x}_0 + \int_0^t \mathbf{v}(\mathbf{x}_\tau, \tau) d\tau$, where $\mathbf{x}_0$ is drawn from the prior distribution. $\mathbf{x}_t$ assumes a time-dependent density $p_t(\mathbf{x})$.

B. Data-coupled flow matching

To obtain the velocity $\mathbf{v}(\mathbf{x}_t, t)$, one needs to first specify the time-dependent density $p_t(\mathbf{x})$. One way to define the time-dependent random variable $\mathbf{x}_t$ is through the stochastic interpolant (SI) [2]

$$\mathbf{x}_t = I(\mathbf{x}_0, \mathbf{x}_1, t) + \gamma(t)\mathbf{z}, \tag{2}$$

where $I(\mathbf{x}_0, \mathbf{x}_1, t)$ is the interpolant and $\mathbf{z} \sim \mathcal{N}(0, I)$ is an additive Gaussian random variable with intensity $\gamma(t)$. The interpolant function and γ should satisfy the boundary conditions $I(\mathbf{x}_0, \mathbf{x}_1, 0) = \mathbf{x}_0, I(\mathbf{x}_0, \mathbf{x}_1, 1) = \mathbf{x}_1$ and $\gamma(0) = \gamma(1) = 0$.

B. Huang and J. Qi are with the Department of Biomedical Engineering, University of California, Davis.

J. Zhou is with Canon Medical Research USA (CMRU), correspondence: jzhou@mru.medical.canon

The first author conducted the majority of this work while interning at CMRU in 2025.

This ensures that the time-dependent density $p_t(\mathbf{x})$ aligns with the prior and the target distributions at the boundaries.

For the task of image denoising, we can treat the noisy images as being drawn from the prior distribution. Thus $\mathbf{x}_0$ and the clean image $\mathbf{x}_1$ should be correlated. They can be seen as data pairs drawn from a joint distribution $\mathbf{x}_0, \mathbf{x}_1 \sim p(\mathbf{x}_0, \mathbf{x}_1) = p(\mathbf{x}_0|\mathbf{x}_1)p(\mathbf{x}_1)$, where the joint density should marginalize on the prior $p(\mathbf{x}_0)$ and target $p(\mathbf{x}_1)$ distributions. With this correlation, the data pair is said to be *coupled*.

To estimate the velocity in (1) that solves the transport of $\mathbf{x}_t$, one can use the following objective [2]

$$\mathcal{L}(\theta) = \int_0^1 \mathbb{E} \left\| \mathbf{v}_\theta(\mathbf{x}_t, t) - \partial_t I(\mathbf{x}_0, \mathbf{x}_1, t) - \dot{\gamma}(t)\mathbf{z} \right\|^2 dt, \quad (3)$$

which is the squared error between the network prediction $\mathbf{v}_\theta$ and the time derivative of the SI. The expectation is taken over $\mathbf{x}_0, \mathbf{x}_1 \sim p(\mathbf{x}_0, \mathbf{x}_1)$ and $\mathbf{z} \sim N(0, I)$. Since the density of $\mathbf{x}_t$ is $p_t(\mathbf{x})$, which is bounded at $t = 0$ and $t = 1$, the prior distribution is transformed to the target distribution by solving the ODE in (1).

C. Implementation details for CT image denoising

In this study, we used a linear interpolant and omitted $\gamma(t)\mathbf{z}$:

$$I(\mathbf{x}_0, \mathbf{x}_1, t) = (1 - t)\mathbf{x}_0 + t\mathbf{x}_1. \quad (4)$$

The velocity field was parameterized by a standard U-Net widely used in diffusion models, with time steps encoded using positional encoding. It consisted of four levels with a base channel of 64, and a total of 57M parameters. We adopted a sampling schedule of ten discrete timesteps that were evenly distributed between 0 and 1. The learned sinusoidal embedding was employed to enhance the performance. During inference, we used the Euler method to solve the ODE. We also trained another model where the timesteps were continuously sampled. The number of inference timesteps was varied from 1 to 20 to evaluate the effect of sampling steps on the image quality. All the models were trained for 332 epochs (240k steps) until the validation loss reached its minimum. A linearly decaying learning rate was used with a peak rate of 1×10^{-4}.

To construct the training dataset, a total of 11 high-dose helical CT scans were selected from routine clinical examinations, each acquired at 120 kV and a minimum of 220 mAs. Of these scans, ten were utilized for model training and one was reserved for validation. To establish the ground-truth targets, high-dose data were reconstructed using a model-based iterative reconstruction (MBIR) algorithm. Following the method in [12], the corresponding low-dose scans were synthesized by injecting data-dependent Gaussian noise into the high-dose data to simulate 20% and 25% dose levels. These simulated low-dose data were subsequently reconstructed using FBP as network inputs. These two dose levels were mixed with the same probability during the training. The trained models were also tested on real clinical LDCT images acquired at 120 kV and 35 mAs. All images were reconstructed with a 512×512 matrix covering a 500-mm field-of-view.

D. Comparison methods

We compared the proposed FM models with 1) a U-Net and 2) a conditional diffusion model. All the methods share the same U-Net structure, including the number of base channels, levels, and 3D convolutional blocks, with the only difference being the presence of the time embedding module required by the DDPM and the FM models. The denoising U-Net was trained with the MSE loss between the network prediction and the target image. For the DDPM, we conditioned the score function with the LDCT image. The v-parametrization was used due to its superior performance.

III. Results

The denoised images are shown in Fig. 1. The FM model is trained using the discrete 10-step schedule but is sampled with 5 steps by skipping one timestep at each step. The denoising U-Net output appears blurry, despite its high PSNR and SSIM values. Both DDIM and FM produce images that are visually similar to the target images. The proposed FM model has higher numerical accuracy, outperforming the DDIM in terms of PSNR and SSIM.

The denoised images from the testing set are shown in Fig. 2. The FM model consistently outperforms the compared methods, achieving better recovery of fine image details.

We evaluated the FM model using different numbers of sampling steps, and the results are shown in Fig. 3. This FM model is trained with a continuous sampling schedule and only on the 20%-dose data. Although the peak PSNR is achieved with two sampling steps, the resulting image appears blurry, similar to the denoising U-Net result. This is expected, since the inference behaves like residual learning between the image pair when the number of sampling timesteps is low. The SSIM value, however, reaches a plateau after increasing the number of steps to five, indicating a constant visual quality.

IV. Discussion

We developed a data-coupled flow-matching framework for CT image denoising. The image accuracy of the proposed method outperforms the conditional DDPM method and the conventional denoising U-Net. This superior result is achieved using only five sampling steps. Compared to the denoising U-Net, it successfully addresses the blurring effect in the restored images. Compared to the conditional DDPM method, the proposed method also achieves better performance. One contributing factor may be that the stochastic process ($\eta = 0.1$ was used) in DDIM can make it more susceptible to hallucinations.

By varying the number of sampling steps, we observe that the performance gain of increasing steps is limited and even becomes negative after a certain point. The possible reason is that the error in the velocity field accumulates over time. The underlying sampling path is already relatively straight with the use of data coupling. Therefore, there is no need to use excessive time steps.

We did not apply conditioning to the velocity field for the purpose of proof-of-concept. Several recent studies highlight the importance of classifier-free guidance (CFG) in image

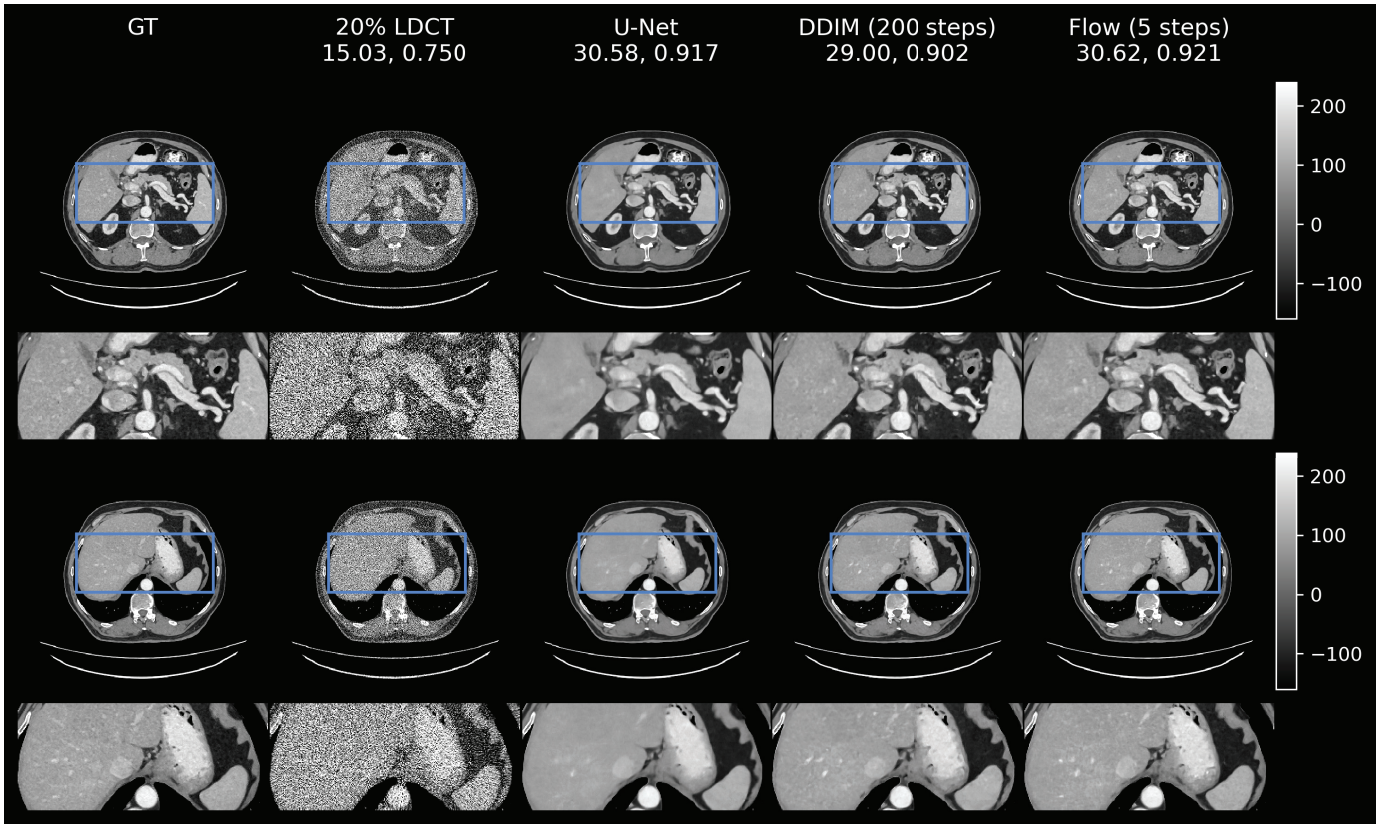


Fig. 1: Denoised results using inputs of a simulated 20%-dose input image. The PSNR (in dB) and SSIM values are listed below the column titles.

generation [13], [14]. In the future, it will be of interest to integrate additional conditions into this framework, e.g., dose levels and structural information from other modalities.

V. Acknowledgment

The authors would like to thank Dr. Marcus Chen at the National Heart, Lung, and Blood Institute, National Institutes of Health, USA for providing the clinical datasets used for model training.

References

[1] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow Matching for Generative Modeling," Sep. 2022. [Online]. Available: https://openreview.net/forum?id=PqvMRDCJT9t

[2] M. S. Albergo, N. M. Boffi, and E. Vanden-Eijnden, "Stochastic Interpolants: A Unifying Framework for Flows and Diffusions," Oct. 2025, arXiv:2303.08797 [cs]. [Online]. Available: http://arxiv.org/abs/2303.08797

[3] A. Tong, K. Fatras, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, G. Wolf, and Y. Bengio, "Improving and generalizing flow-based generative models with minibatch optimal transport," Transactions on Machine Learning Research, Nov. 2023. [Online]. Available: https://openreview.net/forum?id=CD9Snc73AW

[4] X. Liu, C. Gong, and Q. Liu, "Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow," Sep. 2022. [Online]. Available: https://openreview.net/forum?id=XVjTT1nw5z

[5] G.-H. Liu, A. Vahdat, D.-A. Huang, E. Theodorou, W. Nie, and A. Anandkumar, "I^{2}SB: Image-to-Image Schrödinger Bridge," Jun. 2023. [Online]. Available: https://openreview.net/forum?id=WH2Cy3eQd0

[6] M. S. Albergo, M. Goldstein, N. M. Boffi, R. Ranganath, and E. Vanden-Eijnden, "Stochastic Interpolants with Data-Dependent Couplings," in Proceedings of the 41st International Conference on Machine Learning. PMLR, Jul. 2024, pp. 921–937. [Online]. Available: https://proceedings.mlr.press/v235/albergo24a.html

[7] Y. Lei, C. Niu, J. Zhang, G. Wang, and H. Shan, "CT Image Denoising and Deblurring With Deep Learning: Current Status and Perspectives," IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 8, no. 2, pp. 153–172, Feb. 2024. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10354073

[8] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," IEEE Transactions on Medical Imaging, vol. 37, no. 6, pp. 1348–1357, 2018.

[9] W. Xia, Q. Lyu, and G. Wang, "Low-dose ct using denoising diffusion probabilistic model for 20× speedup," 2022. [Online]. Available: https://arxiv.org/abs/2209.15136

[10] L. Zhou, A. Lou, S. Khanna, and S. Ermon, "Denoising diffusion bridge models," in The Twelfth International Conference on Learning Representations, 2024. [Online]. Available: https://openreview.net/forum?id=FKksTayvGo

[11] C. Lu, Y. Zhou, F. Bao, J. Chen, C. LI, and J. Zhu, "DPM-Solver: A Fast ODE Solver for Diffusion Probabilistic Model Sampling in Around 10 Steps," in Advances in Neural Information Processing Systems, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35. Curran Associates, Inc., 2022, pp. 5775–5787.

[12] Y. Fan, A. A. Zamyatin, and S. Nakanishi, "Noise simulation for low-dose computed tomography," in 2012 IEEE Nuclear Science Symposium and Medical Imaging Conference Record (NSS/MIC), 2012, pp. 3641–3643.

[13] Z. Geng, M. Deng, X. Bai, J. Z. Kolter, and K. He, "Mean Flows for One-step Generative Modeling," May 2025, arXiv:2505.13447 [cs]. [Online]. Available: http://arxiv.org/abs/2505.13447

[14] T. Kynkäänniemi, M. Aittala, T. Karras, S. Laine, T. Aila, and J. Lehtinen, "Applying guidance in a limited interval improves sample

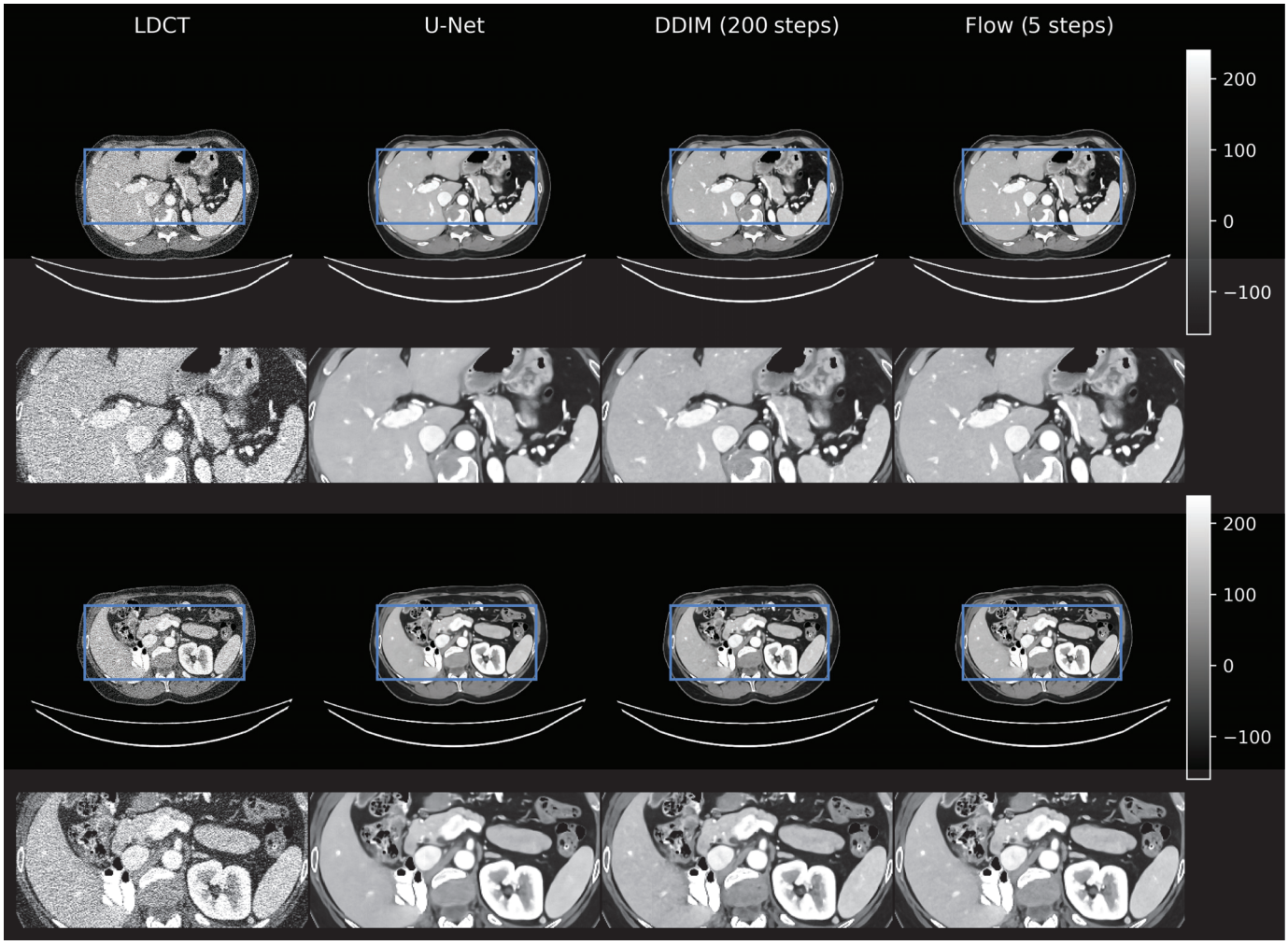


Fig. 2: Denoised results on the real LDCT image. Data were acquired at 120kV and 35 mAs.

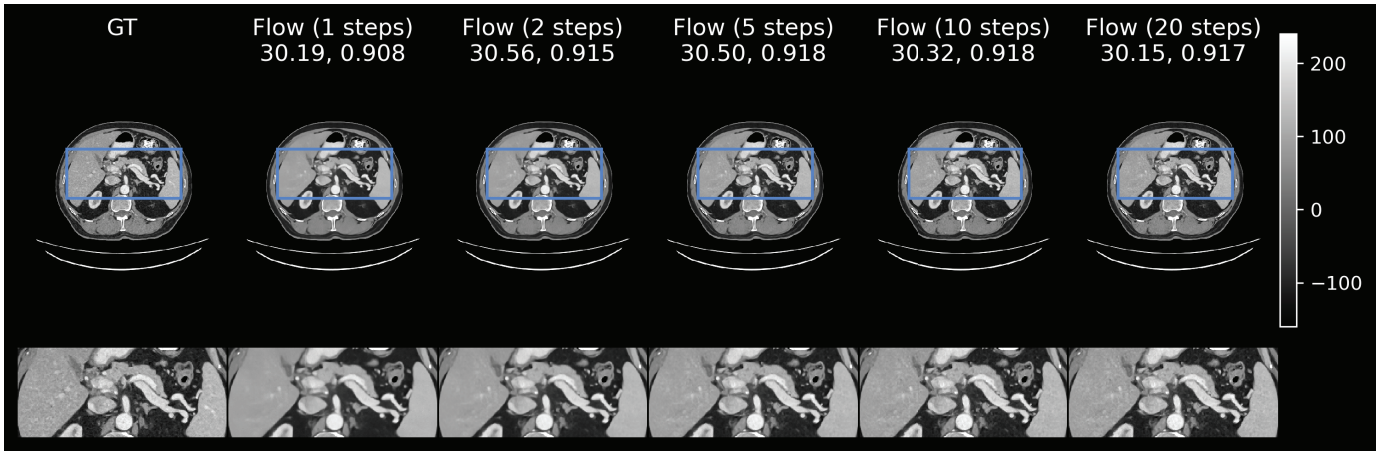


Fig. 3: Denoised images of the FM model with different numbers of sampling steps. Note that the model is trained with the continuous sampling schedule and only uses the 20%-dose data.

and distribution quality in diffusion models," in The Thirty-eighth Annual Conference on Neural Information Processing Systems, 2024. [Online]. Available: https://openreview.net/forum?id=nAIhvNy15T

Application of UNet-Based Network for Projection-Domain Material Decomposition in CBCT, A Simulation Study

Yuedong Yuan, Daniel Punzet and Georg Rose

***Abstract*—Spectral computed tomography(CT) enables advanced material decomposition by registering energy-dependent X-ray attenuation information. Traditional decomposition approaches have drawbacks, such as complex computation and sensitivity to beam-hardening and scatter artifacts. To address these limitations, this study explores the application of a customized UNet-based neural network with a 3D encoder and a 2D decoder for projection-domain material decomposition. The network takes as input four sets of phantom projections at four different energy channels and predicts the corresponding projections for each of the materials, soft tissue, bone, iodine and background water in the phantoms. Experimental results demonstrate the network's capacity to generate high-quality decomposed material projections with mean SSIM 0.995 and mean MSE 9.57×10^{-5} on six test phantoms. For the reconstructed volumes of the decomposed material projections, the mean SSIM was 0.958 and the mean MSE was 1.70×10^{-7}. The developed network architecture offers a data-driven alternative to conventional approaches for the inverse problem of material decomposition in cone-beam CT.**

***Index Terms*—Spectral CT Imaging, UNet-Based Network, Material Decomposition, Cone-Beam CT.**

I. Introduction

SPECTRAL computed tomography (CT) techniques register the energy dependence of X-ray attenuation to enable imaging with more information than just gray values. After acquiring projection data at multiple energy channels, spectral CT can decompose a scanned object into its constituent materials [1]. Furthermore, virtual non-contrast imaging can be obtained through material decomposition, which is an accurate method for differentiating intracranial hemorrhage from iodinated contrast [2]. However, conventional material decomposition methods, which are usually formulated as linear inverse problems, often have issues with noise amplification, inaccurate effective mass attenuation coefficients, and sensitivity to beam-hardening and scatter effects, especially for projection-domain processing [3], [4]. Meanwhile, deep learning methods have shown great potential in solving inverse problems in medical imaging by learning robust, data-driven mappings from measurements to physical quantities. In this work, we investigate a customized UNet-based neural network for projection-domain material decomposition in cone-beam CT (CBCT). By analyzing phantom projections from four energy channels, the proposed neural network aims to overcome the limitations of traditional decomposition algorithms and improve the robustness and precision of material separation. Most existing approaches employ either a purely 2D or a fully 3D convolutional neural network (CNN) design. In contrast, the proposed method introduces a hybrid UNet-based architecture with a 3D-convolution encoder and a 2D-convolution decoder that are specifically tailored for multi-energy data.

Y. Yuan, D. Punzet and G. Rose are with Research Campus STIMULATE, Otto-von-Guericke University, 39106 Magdeburg, Germany(e-mail:yuedong.yuan@ovgu.de).

Y. Yuan, D. Punzet and G. Rose are with Chair in Healthcare Telematics and Medical Engineering, Faculty of Electrical Engineering and Information Technology, Otto-von-Guericke University, 39106 Magdeburg, Germany.

II. Materials and Methods

A. Material decomposition

Material decomposition is based on the assumption that the attenuation coefficient can be formulated as a linear combination of the mass attenuation coefficients of the selected basis materials [1], expressed as

$$\mu(E, \vec{r}) = \sum_{m=1}^{M} c_m(\vec{r})\, f_m(E), \tag{1}$$

where $\mu(E, \vec{r})$ is the attenuation coefficient as a function of photon energy E and voxel position $\vec{r}$, M is the number of basis materials, $c_m(\vec{r})$ is the concentration of material m ($m \in \{1, 2, \ldots, M\}$) at the position $\vec{r}$. The term $f_m(E) = \mu_m(E)/\rho_m$ represents the mass attenuation coefficient of material m at the photon energy E, where μ_m and ρ_m are the attenuation coefficient and density of material m, respectively.

Each software phantom is scanned using four energy channels: $[45, 55]$ keV, $[55, 68]$ keV, $[68, 85]$ keV, $[85, 120]$ keV. Consequently, four sets of projection measurements are acquired. The goal of material decomposition is to solve for the material concentrations c_m based on the measured attenuation coefficients μ_E and the corresponding mass attenuation coefficients $f_m(E)$. However, due to noise in the measurements and inaccuracies in the estimated effective mass attenuation coefficients for each energy channel, results obtained by directly inverting the coefficient matrix are often inaccurate. To address this inverse problem, a UNet-based neural network was developed and trained. The UNet's encoder extracts high-level features from the input projections, and the decoder reconstructs spatial details [5]. It is therefore adopted to separate materials in the scanned phantom. The material decomposition result is evaluated by two figures of merit, i.e. the Structural Similarity Index Measure (SSIM) and the Mean Squared Error (MSE) of the entire decomposed material images.

B. Data Generation and Preparation

We simulated a C-arm CBCT system using the open-source framework CTL [6]. The system's geometry was based on a generic C-arm CBCT, and it was configured with a flat-panel detector. The detector consisted of 256 × 240 pixels, each with a size of 4.0 × 4.0 mm². The distance from the source to the isocenter was set to 750 mm, and the C-arm span was 1000 mm. The X-ray source was modeled as a polychromatic source with 120 kV tube voltage. A 1.0-mm aluminum filter was fitted before the X-ray source. The resulting X-ray spectrum was then divided into the four energy channels as mentioned in section II-A, and the four subspectra were calculated using the TASMIP algorithm with an energy bin width of 0.25 keV. Projection data were calculated using the CTL standard forward projection pipeline with spectral effects and Poisson noise. A total of 400 projection views were acquired using a short-scan trajectory. Corresponding ground-truth projections at the monochromatic reference energy of 75 keV were simulated without noise for each material of soft tissue, bone, iodine, and water. Volumes were reconstructed using a filtered back projection (FBP) algorithm followed by 5 iterations of an algebraic reconstruction technique (ART) to reduce artifacts.

To promote robust learning and generalization of the proposed UNet-based neural network, two diverse training datasets of synthetic software phantoms with high variability in geometry and material density were generated. Each phantom was a 128^3-voxel volume with a water-equivalent background and contained a randomized number of objects, ranging from 35 to 50 in the first set and from 80 to 100 in the second. Each object within the phantom consisted of a single material (soft tissue, bone or iodine) with varying concentrations and was represented by one of three shapes (blob, cube, or cylinder) with randomly sampled sizes and spatial distributions. Object dimensions varied within predefined ranges to ensure substantial geometric diversity across the training datasets. Moreover, objects within each phantom were designed to avoid overlap. Material density variability was introduced by randomly scaling the density of each material within material-specific ranges: 0.875–1.0 for bone, 1.0–1.253 for soft tissue, and 0.048–0.073 for iodine, corresponding to the clinically relevant iodinated contrast agent OMNIPAQUE with a concentration of 180–300 mg I/mL. This randomized phantom generation strategy yields a wide distribution of appearances and material configurations and helps reduce overfitting to fixed phantom designs.

C. UNet-Based Network for Material Decomposition

A customized UNet-based neural network with a 3D-convolution encoder and a 2D-convolution decoder was developed to learn a data-driven mapping from four sets of CBCT projections to material-specific projections for a phantom. The input data consists of four projection images at the same projection view stacked along the energy (depth) dimension and jointly processed by 3D convolutions to enable fusion of spectral information and learn the correlation between the energy channels [7]. Each encoder block consists of two consecutive 3D convolutional layers followed by a SiLU activation function, with kernel size of (4,3,3), padding of (0,1,1) and stride of (1,1,1), which collapses the depth (energy) dimension into 2D feature maps while preserving in-plane spatial resolution. As encoding proceeds, spatial resolution is gradually reduced and the number of feature channels is increased, enabling the extraction of increasingly abstract features.

The decoder operates purely in 2D convolution and generates high-resolution material-specific projections using UNet-style skip connections to fuse low-level spatial features from the encoder with high-level semantic information. Each decoder block consists of a transposed 2D convolution with kernel size 2, stride 2 and padding 0 for upsampling, followed by two 2D convolutional layers with kernel size 3, padding 1, stride 1, and a SiLU activation function. This design enables effective spatial resolution recovery while maintaining stable gradient flow during training. A Sigmoid activation function is applied at the output layer (1×1 convolution) to generate normalized decomposed projection images for each material, including soft tissue, bone, iodine, and water (background). Fig. 1 illustrates the architecture of the customized UNet-based network and input data as well as output data.

D. Network Training

The proposed network was trained on a high-performance computing cluster using two NVIDIA GeForce RTX 2080 Ti GPUs. The total number of trainable parameters in the model is approximately 2.79×10^6, allowing efficient training while maintaining sufficient model capacity for projection-domain material decomposition.

Prior to network training, two various training sets of 200 software phantoms were created using software CTL. For each phantom, four series of projection data and ground truths were simulated with the CT system configuration described in section II-B. The complete training dataset amounts to approximately 314.6 GB and was randomly split into training and validation subsets, with 85% of the samples (340 phantoms) used for training and the remaining 15% (60 phantoms) reserved for validation. The test data were generated separately.

The input data for the network were preprocessed to ensure numerical stability and consistent scaling. For each phantom, noisy projection data acquired using four energy channels were first stacked along the energy channel and then normalized using min-max normalization to map the values to the range $[0, 1]$. After that the data were cast into single-precision floating-point format. Ground-truth projections for each basis material were normalized independently using min-max normalization to account for differences in attenuation magnitude of different materials.

The network parameters were optimized using the optimizer AdamW [8] with learning rate of 10^{-4} and a weight decay of 10^{-4}. A learning rate scheduler based on ReduceLROnPlateau in the deep learning framework PyTorch was employed to adaptively reduce the learning rate when the validation loss ceased to improve, thereby stabilizing convergence during training. To supervise network training, a Smooth L_1 loss

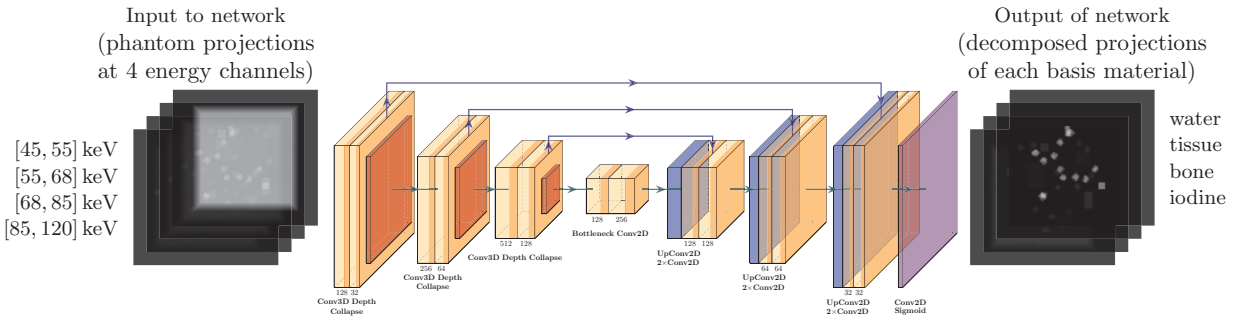


Fig. 1: Designed UNet-based Network, each encoder block consists of Conv3D-Conv3D-SiLu-MaxPool2d, each decoder block consists of UpConv2D-Conv2D-Conv2D-SiLu, and the bottleneck consists of Conv2D-Conv2D.

function was adopted due to its robustness to outliers and noise in projection data. The loss was computed separately for each material projection and combined using a foreground-weighted scheme. The final loss was obtained by summing the weighted pixel-wise losses across all decomposed material projections, encouraging accurate reconstruction of both low-contrast and sparse material objects. The calculation of training and evaluation loss is formulated as below.

Let $\hat{\mathbf{Y}} \in \mathbb{R}^{N\times C\times H\times W}$ denote the network output and $\mathbf{Y} \in \mathbb{R}^{N\times C\times H\times W}$ the corresponding ground truth, where N is the batch size, C denotes the number of material projections, and $H \times W$ represents the spatial resolution of the projection images.

The Smooth L_1 loss (ℓ_{SL1}) is defined element-wise as

$$\ell_{\mathrm{SL1}}(x) = \begin{cases} \frac{1}{2\beta}x^2, & |x| < \beta, \\ |x| - \frac{\beta}{2}, & \text{otherwise}, \end{cases} \tag{2}$$

where $x = \hat{Y}_{n,c,i,j} - Y_{n,c,i,j}$ and β is a threshold parameter, set to 0.1 in this study.

To compensate for the imbalance between background and foreground pixels, a material-specific foreground weighting scheme is applied. The weighting mask for channel c is defined as

$$W_{n,c,i,j} = 1 + \alpha_c \, \mathbb{I}(Y_{n,c,i,j} > 0)\,, \tag{3}$$

where α_c is a channel-dependent weighting factor, set to the empirical values 10, 8, 3, 2 for materials soft tissue, bone, iodine and water respectively and $\mathbb{I}(\cdot)$ denotes the indicator function.

The weighted loss for each channel is computed as

$$\mathcal{L}_c = \frac{\sum_{n,i,j} W_{n,c,i,j}\, \ell_{\mathrm{SL1}}\left(\hat{Y}_{n,c,i,j} - Y_{n,c,i,j}\right)}{\sum_{n,i,j} W_{n,c,i,j}}, \tag{4}$$

and the total training loss is obtained by summing over all channels:

$$\mathcal{L}_{\mathrm{total}} = \sum_{c=1}^{C} \mathcal{L}_c. \tag{5}$$

III. Results and Discussion

Six test phantoms consisting of soft tissue, bone, iodine with different concentrations and background water were generated

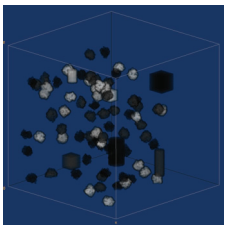


Fig. 2: Sketch of the test phantom 6.

to evaluate the trained model. One of them is depicted in Fig. 2. The projections of each test phantom at four energy channels and projections of each material in the phantom as ground truths were simulated. Then the projection-domain decomposition was carried out with the trained model. After that, the volumes for each material using the decomposed projections were reconstructed in the same way as ground truth reconstructions.

The evaluations are based on the normalized, decomposed material projections and reconstructions ranging from 0 to 1. Table I shows the evaluation of the decomposed material projections from the trained model with ground truth of material projections, and Table II shows the evaluation of the whole reconstructed volumes from the decomposed material projections with ground truths of each material. Fig. 3 presents cross-sections (central slice) of the reconstructed volumes derived from the decomposed material projections and their corresponding ground truths, as well as the differences between them of test phantom 6.

Evaluations of the decomposed projections and reconstructions indicate that the designed UNet-based network architecture can adequately decompose materials in CBCT using four energy channels and denoise projection images. However, the current model is trained by simulation data and must be transferred to real data. Additionally, the weighting factors for the materials should be optimized to adress the remained artifacts inside the decomposed material projections in the future.

IV. Conclusion

This simulation study investigates the application of a UNet-based network adapted with a 3D-convolution encoder and

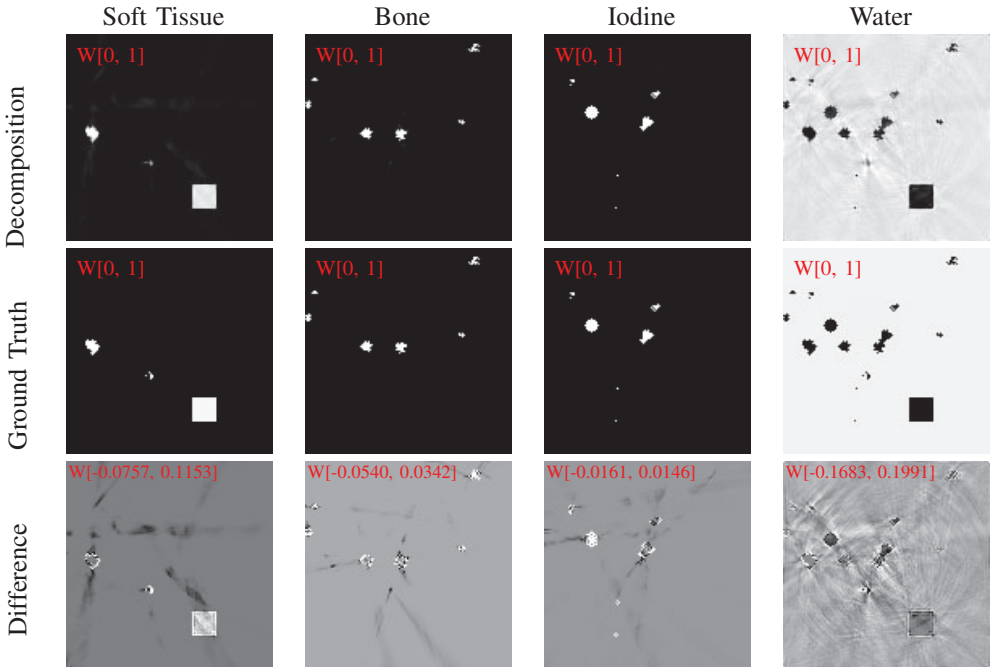


Fig. 3: Slices at the same index from reconstructed decomposed-material volumes for test phantom 6. From top to bottom, the rows show decomposed materials, ground truth, and difference images computed as ground truth minus decomposition. Display window range is indicated at the upper-left corner of each image by W.

TABLE I: Projection-Based Evaluation of Decomposed Materials for Test Phantoms

Phantom	Metric	Tissue	Bone	Iodine	Water
Phantom1	SSIM	0.9860	0.9957	0.9998	0.9983
	MSE	3.35×10^{-4}	1.27×10^{-4}	4.37×10^{-6}	1.12×10^{-5}
Phantom2	SSIM	0.9899	0.9979	0.9999	0.9982
	MSE	1.92×10^{-4}	5.08×10^{-5}	3.26×10^{-6}	1.43×10^{-5}
Phantom3	SSIM	0.9883	0.9935	0.9998	0.9979
	MSE	2.72×10^{-4}	2.85×10^{-4}	7.25×10^{-6}	1.36×10^{-5}
Phantom4	SSIM	0.9851	0.9975	0.9998	0.9981
	MSE	1.72×10^{-4}	4.89×10^{-5}	4.87×10^{-6}	1.01×10^{-5}
Phantom5	SSIM	0.9745	0.9972	0.9998	0.9976
	MSE	3.19×10^{-4}	4.79×10^{-5}	5.11×10^{-6}	1.21×10^{-5}
Phantom6	SSIM	0.9883	0.9916	0.9999	0.9984
	MSE	1.53×10^{-4}	1.97×10^{-4}	2.82×10^{-6}	7.79×10^{-6}

TABLE II: Evaluation of Reconstructed Volumes of Decomposed Materials for Test Phantoms

Phantom	Metric	Tissue	Bone	Iodine	Water
Phantom1	SSIM	0.9732	0.9869	0.9988	0.8713
	MSE	5.86×10^{-7}	1.50×10^{-7}	5.55×10^{-9}	2.74×10^{-8}
Phantom2	SSIM	0.9667	0.9948	0.9993	0.8823
	MSE	2.60×10^{-7}	9.89×10^{-8}	5.04×10^{-9}	3.19×10^{-8}
Phantom3	SSIM	0.9667	0.9817	0.9989	0.8736
	MSE	3.02×10^{-7}	2.55×10^{-7}	8.21×10^{-9}	3.24×10^{-8}
Phantom4	SSIM	0.9736	0.9908	0.9992	0.8628
	MSE	4.14×10^{-7}	8.04×10^{-8}	6.15×10^{-9}	3.31×10^{-8}
Phantom5	SSIM	0.9620	0.9922	0.9991	0.8705
	MSE	7.58×10^{-7}	9.90×10^{-8}	6.65×10^{-9}	3.72×10^{-8}
Phantom6	SSIM	0.9753	0.9922	0.9993	0.8786
	MSE	3.48×10^{-7}	5.14×10^{-7}	3.72×10^{-9}	2.82×10^{-8}

a 2D-convolution decoder for material decomposition in the projection-domain in CBCT using four energy channels. This network architecture can effectively and robustly separate materials, denoise projection images, and can be efficiently trained. In the future, the real data need to be applied to retrain the network and the hyper-parameters of the network should be refined to further improve the model's capability.

Acknowledgment

This work was funded by the Federal Ministry of Research, Technology, and Space (BMFTR) as part of the Research Campus STIMULATE under grant number 13GW0835A. The authors would like to thank Hana Haseljic and Fatima Saad for their help.

References

[1] B. J. Heismann, B. T. Schmidt, and T. G. Flohr, *Spectral Computed Tomography*. Springer, 2012.
[2] A. A. Postma, M. Das, A. A. R. Stadler, *et al.*, "Dual-Energy CT: What the Neuroradiologist Should Know," *Current Radiology Reports*, vol. 3, no. 16, 2015.
[3] W. Cong, B. De Man, and G. Wang, "Projection Decomposition via Univariate Optimization for Dual-Energy CT," *Journal of X-Ray Science and Technology*, vol. 30, no. 4, pp. 725–736, 2022.
[4] J. Schaefer *et al.*, "Simulation Study of Material Decomposition in Dual-Energy Radiography for Bone Removal," in *Proc. IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*, 2022, pp. 1–7.
[5] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, 2015.
[6] T. Pfeiffer, R. Frysch, R. N. K. Bismark, and G. Rose, "CTL: Modular Open-Source C++ Library for CT Simulations," 2019.
[7] K. Han, C. H. Ryu, C.-L. Lee, and T. H. Han, "Deep learning-based material decomposition of iodine and calcium in mobile photon counting detector CT," *PLOS ONE*, vol. 19, no. 7, p. e0306627, 2024.
[8] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv:1711.05101*, 2017.

Enhancing Reproducibility for CT-based Radiomics: A Pragmatic Approach to Mitigate the Impact of Reconstruction Kernel Variations

Yun-Huan Lyu, Wei Wu, Yuzheng Zhang, Paul Kinahan
Department of Radiology, University of Washington, Seattle, USA
Fred Hutchinson Cancer Center, Seattle, USA

Abstract

Introduction: When CT images are used in quantitative radiomics analysis, changes in the image reconstruction kernel are known to be a confounding factor. Our goal was to evaluate a pragmatic method of harmonizing resolution across images reconstructed with different kernels and analyze the value change of individual radiomic features, and also the downstream impact on a predictive model for lung cancer classification.

Methods: We derived the ratio of the frequency spectra between pairs of images reconstructed with a reference (standard) kernel and a sharper (i.e. higher frequency) kernel to estimate the kernel-pair frequency transfer function. We used the ratios of the radial frequency magnitude from axial CT image slices under different reconstruction kernels to guide a Gaussian Process regression that was then used to smooth the sharper image of each pair. Images included phantoms and patients. We determined the optimal Gaussian Process hyperparameter by fitting the frequency ratio between the sharp and reference images of a Corgi phantom. For pairs of patient images, we evaluated the impact on radiomics feature values and also on predictions from a risk model of malignancy.

Results: The Gaussian Process fit was shown to closely match the empirically-measured transfer function. The variance of the difference of the predicting model across a patient subset was substantially reduced using these optimized smoothing factors. Although use of the optimized Gaussian Process fit did not eliminate the bias and variance of the risk prediction model, both were substantially reduced.

Conclusion: We demonstrated an explainable and robust approach to pre-process CT images reconstructed with different kernels that are being used in a radiomics-based risk prediction model. The pre-processing step can be calibrated across varying reconstruction settings types using phantoms.

Introduction

This study focuses on radiomics, the extraction of quantitative values, called features, from CT and other types of medical images to provide diagnostic, prognostic, and predictive information [Lambin]. Radiomics has shown substantial promise in supporting disease diagnosis and prognosis [Hannequin, Zhang]. However, variations in CT acquisition and reconstruction parameters (pitch, tube current, voltage, slice thickness, reconstruction kernel, and others) reduce feature reproducibility, which is one of the factors limiting clinical translation [Berenguer]. Among these, the reconstruction kernel is a major factor affecting radiomic features variability [Peng , Cherezov].

Variations in the choice of kernel occur for multiple reasons, including manufacturer and scanner dependent options, site practices, and physician preferences. This complicates cross-scanner implementation of quantitative analysis methods such as radiomics. Methods have been developed for pre-processing or harmonizing reconstructed CT images reconstructed with different kernels including restricting the choice of allowable kernels [e.g. Jordan], using generative adversarial networks (GANs) as an image preprocessing step prior to radiomics feature calculation [Krishnan], and using statistical methods to compensate for batch effects in the extracted feature values [Mali].

We previously conducted a study using radiomics-based risk model to predict the malignancy status of indeterminate pulmonary nodules found on chest CT images of 284 patients with known outcomes [Lastwika]. While the model achieved state-of-the-art performance levels, it was based on locally-acquired CT images with a high level of consistency in the choice of kernel. To use this model with external cohorts, we explored a pragmatic, explainable and robust approach to pre-process CT images using Gaussian Process regression fitting. Part of the motivation was the ability to use phantom scans to calibrate the degree of smoothing.

Methods

Analytical derivation: Since reconstruction kernels are applied directly to the projections, we can calculate the conversion from a source kernel to a target kernel by using the inverse ratio of the two kernels:

$$P_{target}(u,v,w) = \frac{K_{target}(u,v,w)}{K_{source}(u,v,w)} \times P_{source}(u,v,w)$$

By the central-section theorem, the frequency-domain ratio of reconstruction kernels equals the ratio of corresponding image pairs from the same patient:

$$\frac{K_{target}(u,v,w)}{K_{source}(u,v,w)} = \frac{I_{target}(u,v,w)}{I_{source}(u,v,w)}$$

where I represents the patient images in the frequency domain. This equivalence enables empirical estimation of the kernel ratio from paired phantom or clinical images, eliminating the need to derive each individual reconstruction kernel explicitly.

Empirical derivation: From our prior study [Lastwika], we analyzed patient chest CT images reconstructed using different pairs of reconstruction kernels. Each image contained at least one lung nodule measuring 6–30 mm. Corresponding kernel pairs were matched with images of the Corgi phantom reconstructed under the same settings, enabling comparison in the frequency domain. For each pair, we computed ratios of the transfer functions and assessed their fit using Gaussian Process regression. The model employed a combination of constant and radial basis function (RBF) kernels as the covariance function, with hyperparameters empirically optimized using the Corgi phantom data.

Pre-processing optimization: From our prior study there was a subset of 15 patient acquired on GE CT scanners who had images reconstructed with both smooth(standard) and sharp(lung) reconstruction kernel. For each patient image we applied the Gaussian Process fitted curve to harmonize the sharp image and quantitatively evaluated the Root Mean Squared Error (RMSE), Structural Similarity Index Measure (SSIM) between the smooth images and harmonized sharp images, and qualitatively visualize the image alignment improvement.

Impact on radiomic features and risk model: We evaluated the impact of Gaussian Process harmonization on both radiomic features and downstream risk model predictions. A pre-existing lung cancer prediction model [Lastwika], incorporating nine radiomics features and three demographic variables, defined as $P(cancer) = \frac{1}{1+e^{-z}}$, $where\ z = -2.299 + \sum_{i=1}^{12} \beta_i \times \frac{(X_i - mean_i)}{std_i}$,with the specific features and standardized coefficients detailed in Table1. This model was applied to smooth, sharp, and harmonized sharp images with corresponding demographic data. To visualize radiomics changes induced by reconstruction kernel variation, we analyzed features used in the model while excluding texture-independent features (e.g., "original_shape_Flatness" and "original_shape_LeastAxisLength"), and plotted patient-level scatter plots in feature space before and after harmonization. To quantify prediction probability shifts

Table 1. Model features and standardized coefficients. The mean and standard deviation of each feature were calculated from the training dataset.

Feature (X_i)	Coefficient (β_i)	$mean_i$	std_i
original_shape_Flatness	0.4065	0.5774	0.1456
original_shape_LeastAxisLength	-0.3071	11.93	5.228
wavelet-LLH_glcm_Imc1	-0.7283	-0.2344	0.0916
wavelet-LLH_glcm_Imc2	-1.114	0.8044	0.1063
wavelet-LHL_glcm_Imc1	0.9193	-0.1452	0.0596
wavelet-HLL_ngtdm_Busyness*	0.9272	-0.6754	1.05
wavelet-HLH_glcm_Imc2*	1.244	-1.081	0.3117
wavelet-HLH_gldm_DependenceVariance	-0.6328	23.49	5.425
wavelet-LLL_firstorder_Energy*	1.02	-21.25	1.259
age	0.02419	n/a	n/a
sex	-0.619	n/a	n/a
Pack-years of smoking	0.03088	n/a	n/a

* denotes log-transformed variables

due to kernel variation, we evaluated individual radiomics features and computed the median change in predicted malignancy probability attributing to the remaining seven features.

Overall harmonization performance was assessed by comparing bias, variance, concordance correlation coefficient (CCC), and discriminative performance of model predictions before and after harmonization.

Results

Empirical derivation: Figure 1 shows strong concordance in transfer function ratios for the patient and phantom data, and the corresponding Gaussian Process fit (R^2 = 0.99).

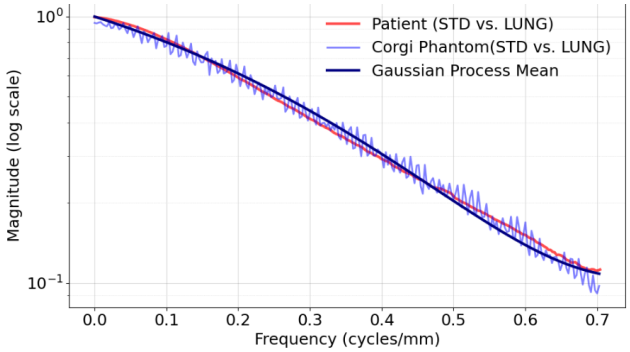


Figure 1. Frequency distribution ratio curves for patient, Corgi phantom images, and the Gaussian Process fit (R^2 = 0.99).

Pre-processing optimization: Gaussian Process hyperparameters, optimized using the Corgi phantom, were applied to 15 patient datasets to harmonize sharp images. To evaluate harmonization, we computed RMSE and SSIM within lung regions segmented using a pre-trained U-Net model (LTRCLobes_R231) [Hofmanninger]. As shown in Figure 2, white contours delineate lung boundaries, and the difference maps indicate significant improvement in both RMSE and SSIM after harmonization.

Impact on radiomic features and risk model: Figure 3 illustrates improved alignment of radiomics feature values following harmonization. After harmonization, data points cluster more closely to their original positions in feature space, indicating that variability induced by reconstruction kernel differences is substantially reduced. Figure 4 illustrates the contribution of individual radiomics features to prediction probability shifts associated with reconstruction kernel variation. Notably, the feature "wavelet-HLL_ngtdm_Busyness" and "wavelet-HLH_glcm_Imc2" induce shifts exceeding 10% when switching to the lung kernel. These shifts are substantially reduced with Gaussian Process harmonization. Figure 5 shows that the prediction probability differences between soft and sharp images decrease after harmonization, with bias reduced from 6.8% to 2.0%, the 95% confidence interval narrowing from 42.9% to 11.9%, and the concordance correlation coefficient (CCC) increasing from 0.925 to 0.9935. Figure 6 presents the area under the ROC curve (AUC) for soft, sharp, and harmonized sharp patient images. Although the harmonized sharp images do not achieve the highest AUC, their performance is closer to that of the soft images. This result highlights the goal of harmonization, not to improve absolute performance, but to reduce sensitivity to reconstruction differences and achieve greater consistency across imaging conditions.

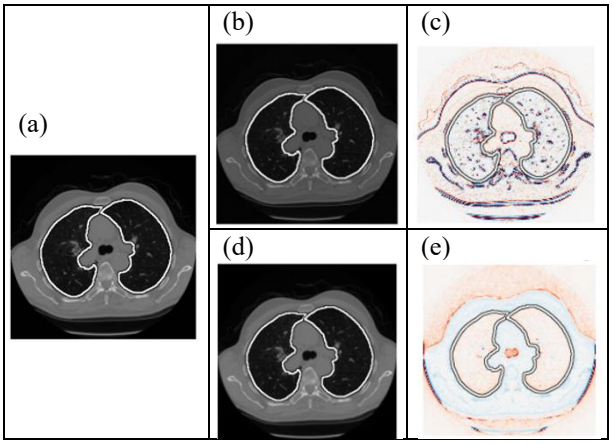


Figure 2. Patient chest CT images are acquired with (a) soft, (b) sharp, and (d) harmonized sharp kernels. Difference maps are presented in (c) and (e), with red indicating positive differences and blue indicating negative differences, and they both use the same color scale. The white contour delineates the ROI used for RMSE and SSIM calculations. Following Gaussian Process harmonization, the difference maps show reduced intensity variation, accompanied by a decrease in RMSE (91.0 to 11.9) and an increase in SSIM (0.64 to 0.99).

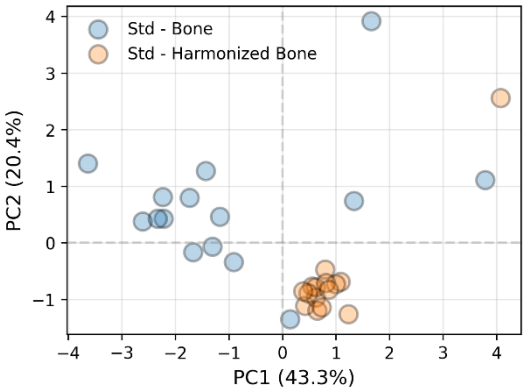


Figure 3. Alignment of radiomics feature values before and after harmonization. After harmonization, samples concentrate nearer their reference positions in feature space, indicating a marked reduction in variability associated with reconstruction kernel differences.

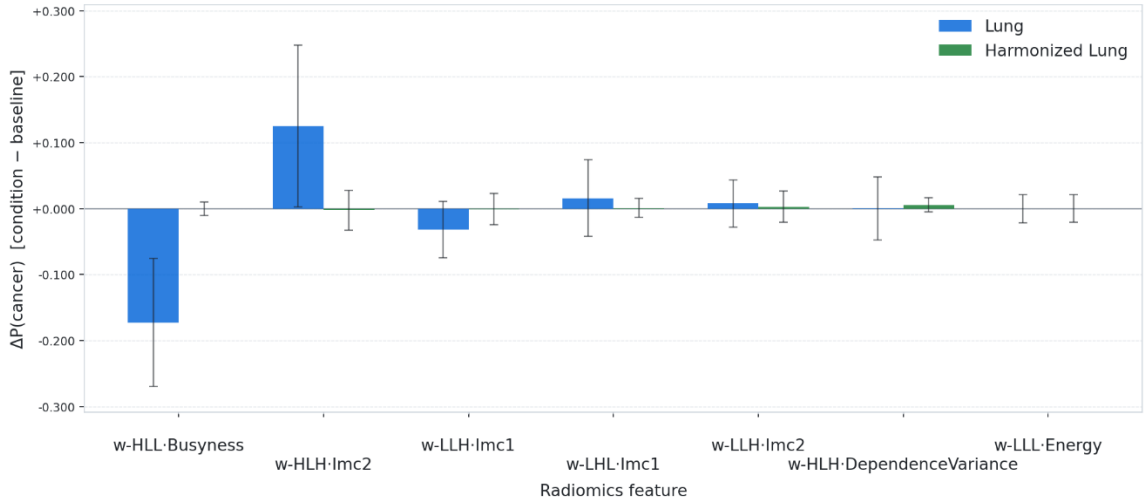


Figure 4. Feature-wise effects on prediction probability under reconstruction kernel variation. Blue bars represent shifts observed with the lung kernel, while green bars correspond to harmonized lung kernel images. The bar lengths (deviations) are computed from the median radiomics feature values across the patient cohort. Large deviations (>10%) are observed for "wavelet-HLL_ngtdm_Busyness" and "wavelet-HLH_glcm_Imc2" with the lung kernel, and these deviations are markedly mitigated after Gaussian Process harmonization.

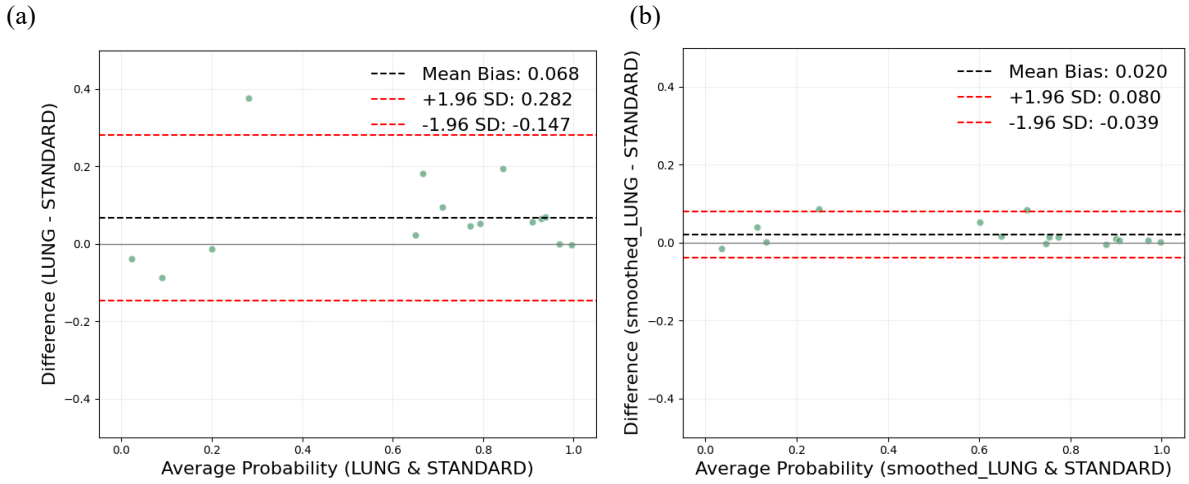


Figure 5. Comparison of prediction probability differences between soft and sharp images before (a) and after (b) harmonization. Harmonization reduces bias from 6.8% to 2.0%, narrows the 95% confidence interval from 42.9% to 11.9%, and improves the concordance correlation coefficient (CCC) from 0.925 to 0.9935, indicating enhanced agreement.

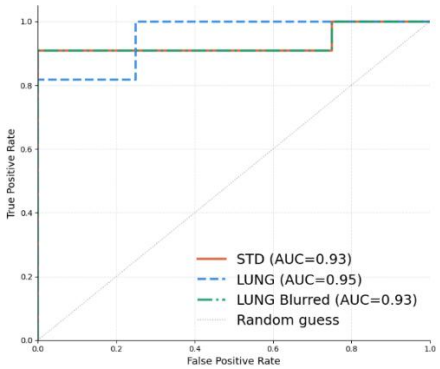


Figure 6. Area under the ROC curve (AUC) for soft, sharp, and harmonized sharp patient images. While harmonized sharp images do not yield the highest AUC, their values are more closely aligned with those from soft images. This reflects the objective of our harmonization method, improving consistency across reconstruction settings rather than maximizing absolute performance.

Conclusion

Transfer function ratio for different kernel pairs can be estimate using paired phantom scans. From this we can fit a Gaussian Process regression filter for pre-processing CT images reconstructed with sharper kernels. This provides an explainable and robust approach for radiomics-based risk prediction models. The pre-processing step, which can be calibrated across varying reconstruction settings using phantoms, reduced the bias and variance while robustizing the discriminative performance in risk prediction.

Acknowledgements

Supported by NIH grants P50CA228944 and R01CA304272

References

Berenguer R, et al.. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. Radiology. 2018;288:407-415. PMID: 29688159.

Cherezov D, et al.. Rank acquisition impact on radiomics estimation (AcquIRE) in chest CT imaging: A retrospective multi-site, multi-use-case study. Comput Methods Programs Biomed. 2024;244:107990. PMID: 38194767.

Hannequin P, et al. FDG PET and CT radiomics in diagnosis and prognosis of non-small-cell lung cancer. Transl Lung Cancer Res. 2022;11:2051-2063. PMID: 36386457.

Jordan DW and Misseldine RE. Quantitative CT reconstruction kernel harmonization for multi-site lung cancer screening. J Appl Clin Med Phys. 2025;26:e70325. PMID: 41204416.

Krishnan AR, et al. Multipath cycleGAN for harmonization of paired and unpaired low-dose lung computed tomography reconstruction kernels. Med Phys. 2025;52:e70120. PMID: 41206352.

Lambin P, et at. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14:749-762. PMID: 28975929.

Lastwika et al.. Multi-Omic Biomarkers Improve Indeterminate Pulmonary Nodule Malignancy Risk Assessment. Cancers (Basel). 2023;15:3418. PMID: 37444527.

Mali SA, et al.. Harmonizing CT scanner acquisition variability in an anthropomorphic phantom: A comparative study of image-level and feature-level harmonization using GAN, ComBat, and their combination. PLoS One. 2025;20:e0322365. PMID: 40344028.

Peng X, et al. Repeatability and Reproducibility of Computed Tomography Radiomics for Pulmonary Nodules: A Multicenter Phantom Study. Invest Radiol. 2022;57:242-253. PMID: 34743134.

Zhang Y, et al.. Radiomics-based Prognosis Analysis for Non-Small Cell Lung Cancer. Sci Rep. 2017;7:46349. PMID: 28418006.

Hofmanninger J, et al. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. Eur Radiol Exp. 2020;4:50. PMID: 32814998

Table 1. Model features and standardized coefficients. The mean and standard deviation for each feature is calculated

Streak Artifact Reduction in Human-scale Dark-field CT Using 3D Gaussian Splatting

T. Dorosti[*1,2,3], D. Frey[*1,2], J. B. Thalhammer[1,2,3,4], J. F. Hilmer[1,2], P. Bleuel[1,2], S. Peterhansl[1,2], J. McGinnis[5,6], D. Pfeiffer[3,4], F. Pfeiffer[1,2,3,4], D. Rueckert[5,6,7], F. Schaff[1,2]

***Abstract*—The clinical dark-field CT scanner (DFCT) enables attenuation- and small-angle scatter-based imaging of human-scale phantoms. Due to the scanner's vibrating Talbot-Lau grating interferometer, images suffer from inherent streak artifacts. In this work, DFCT data is processed by re-purposing the 3D Gaussian Splatting (3DGS) method, designed for rendering sparsely sampled X-ray attenuation data, and is compared against conventional filtering. A ventilated ex vivo porcine lung placed inside an anthropomorphic phantom was measured at 20 randomly selected table positions, and reconstructed with the filtered back-projection (FBP) algorithm at the clinical DFCT setup. The re-purposed 3DGS model was compared with Gaussian, bilateral, and total variation (TV) filters for post-processing DFCT data. For each conventional filter, the hyperparameters were optimized via grid search, maximizing the structural similarity index between 3DGS and the filtered image at the global image level. The contrast-to-noise ratio (CNR) and the full-width-at-half-maximum (FWHM) of manually-selected 20×20 pixel regions of lung tissue and air were calculated. The resulting mean and standard deviation values indicate that 3DGS outperforms conventional filtering in DFCT streak reduction, preserving edge sharpness and a high CNR, a necessity for diagnostic accuracy in future clinical trials.**

I. Introduction

Dark-field X-ray imaging is a technique sensitive to ultra-small-angle X-ray scattering of materials below the attenuation-based X-ray imaging spatial resolution, providing information about microstructures, such as pulmonary alveoli [1], [2]. With a Talbot–Lau interferometer, which consists of a set of periodic grating, attenuation, differential phase, and dark-field signals can be simultaneously measured [3]. Dark-field imaging has shown promising results in radiological applications, in particular as a complimentary functional imaging modality to X-ray attenuation-based radiography [4]–[6].

The Dark-field clinical CT scanner (DFCT) prototype at the Munich Institute of Biomedical Engineering allows for attenuation- and small-angle scatter-based imaging of human-scale phantoms. The system includes a vibrating Talbot–Lau grating setup embedded in the CT gantry to obtain the phase stepping values, as opposed to a discrete translational grating movement [7]. This clinical design choice challenges phase retrieval for DFCT reconstruction as it groups neighboring projection frames within a single gantry rotation to estimate the phase information, resulting in incomplete angular sampling for each reconstructed volume and leading to sparsely sampled streaking artifacts [8].

With a so-called multi-rotation scan, phase information is no longer sparsely sampled as multiple gantry rotations are acquired and utilized for phase retrieval and reconstruction of a single CT volume. Although multi-rotation scans come at the cost of increased acquisition time and radiation dose, limiting factors for practicality in clinical settings, they allow for DFCT processing and reconstructions with improved image quality and reduced streak artifacts [9].

Data-driven reconstruction and post-processing methods have been explored to tackle these limitations. Namely, convolutional neural networks have been shown to successfully reduce such streaks by learning a supervised mapping from single- to multi-rotation data [10]. While this method improves image quality, it depends on availability of high-quality ground truth (GT) multi-rotation data, which itself is affected by residual artifacts. Another limitation is posed by the domain gap between phantom data used for training and real-world clinical scans, which can hinder generalization.

Recently, self-supervised learning has gained popularity due to its domain-independent nature. Representation learning proposes alternative approaches to encoding image information, for example by replacing discrete voxels with parameterized Gaussians. Inspired by concepts from photogrammetry, 3D Gaussian splatting (3DGS) has emerged as a rendering technique capable of representing inconsistently sampled scenes [11].

The self-supervised approach of 3D R^2-Gaussian splatting has shown promise in reconstructing sparsely sampled X-ray data [12], effectively reducing streaking artifacts while preserving contrast and edge sharpness. By forward-projecting existing CT volumes and then reconstructing them with 3DGS, we effectively employ 3DGS as a post-processing step. In this work, we apply the 3DGS method to multi-rotation DFCT data and compare the results with those obtained using traditional post-processing filters. This aims at enhancing DFCT image quality to provide improved surrogate GT for supervised learning approaches.

II. Methods

A. Dataset

A ventilated ex vivo porcine lung placed inside an anthropomorphic Lungman phantom (Kyoto Kagaku, Tokyo, Japan)

*Equal contribution.

No live experiments were performed. Ethical approval was not required.

The authors declare no conflicts of interest.

[1]Chair of Biomedical Physics, Technical University of Munich (TUM), Germany
[2]Munich Institute of Biomedical Engineering, TUM, Germany
[3]Institute for Diagnostic and Interventional Radiology, TUM University Hospital, TUM, Germany
[4]TUM Institute for Advanced Study, TUM, Germany
[5]Chair for AI in Healthcare and Medicine, TUM, Germany
[6]Munich Center for Machine Learning (MCML), TUM, Germany
[7]Department of Computing, Imperial College London, U.K.

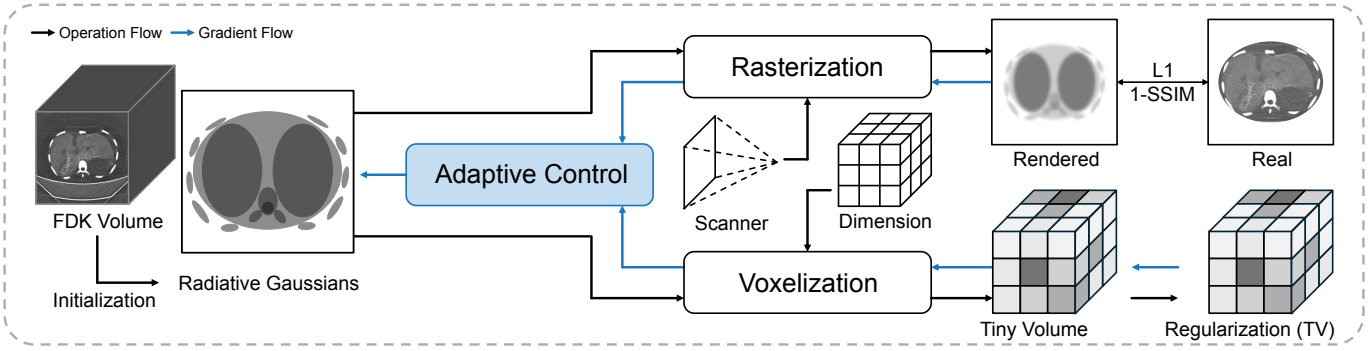


Fig. 1: Pipeline for the 3D Gaussian Splatting post-processing method for X-ray dark-field CT data. The volume is first rasterized with a differentiable renderer, computing pixel-wise projections through integration of the sum of 3D Gaussian density kernels along the X-ray path. Transformed ray-space 2D Gaussian kernels are used to calculate the L_1 and structural similarity index (SSIM) composite loss. The differentiable voxelizer tiles the 3D volume to accumulate neighboring radiative Gaussian contributions for total variation (TV) regularization of the discrete density volumes. Throughout the training process, the Gaussian density representations are dynamically adjusted with the adaptive control, such that important representations are densified, and others are removed. The pipeline is adapted from [12].

was measured at 20 randomly selected patient table positions. CT acquisitions were performed at $80\,\mathrm{kVp}$ and $550\,\mathrm{mA}$ at $1.5\,\mathrm{s}$ rotation time, with 2400 shots per revolution. The lung was ventilated at $10\,\mathrm{mbar}$. Multi-rotation scans (up to 20) were reconstructed from the full stack of recordings using virtual phase-stepping for phase retrieval and the filtered back-projection (FBP) algorithm onto a 800×800 pixel image matrix in the axial plane [9]. The multi-rotation acquisition increases the effective angular sampling for each reconstructed volume, thereby reducing streak artifacts at the expense of increased radiation dose.

B. Conventional Filtering

Streak reduction via post-processing with conventional filters was investigated to determine whether the same effect can be achieved by classical means. Namely, Gaussian, bilateral, and total variation (TV) filters were explored. For each conventional filter, the hyperparameters were optimized on the global axial slice level through a grid search to maximize the structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) between the proposed method (3DGS) and the conventional post-processing filtering methods. The resulting finalized hyperparameters and corresponding metrics are provided in Table I. Filters were implemented with Python's (version 3.12.10) Scikit-learn library (version 0.25.2).

TABLE I: Image-Level Comparison of 3DGS with Conventional Post-processing Filtering Methods ($n = 20$, mean $\pm$ std)

Metric	Gaussian	Bilateral	TV
SSIM	0.93 ± 0.02	0.96 ± 0.01	0.97 ± 0.02
PSNR (dB)	39.06 ± 0.92	39.47 ± 1.01	40.50 ± 0.82
Parameter	$\sigma = 1.00$	$\sigma_{Spatial} = 5.60$	$\lambda = 0.04$

C. 3D Gaussian Splatting

The 3DGS model from [12] was adapted and applied as post-processing to the DFCT data, as depicted in Fig. 1. First, all the data were normalized to a range of zero to one and zero-padded to include 800 axial slices, resulting in a final shape of $800 \times 800 \times 800$ pixels. Next, the 3DGS model was trained using the Tomographic Iterative GPU-based Reconstruction Toolbox (TIGRE version 3.0.0). Training took 75 minutes on an NVIDIA A100 GPU with $80\,\mathrm{GB}$ of VRAM for each volume. With fewer iterations and smaller volume sizes, training time can be substantially reduced.

Gaussian point clouds were initialized with the Feldkamp–Davis–Kress (FDK) algorithm. Throughout training, adaptive control was adjusted such that Gaussians with a size larger than $0.05\,\%$ of the volume were densified from the initial iteration onward, and those with density smaller than $1 \cdot 10^{-5}$ were pruned. Each volume was trained for a total of $30\,\mathrm{k}$ iterations with the Adam optimizer. Position, density, scale, and rotation learning rates were initially set to $2 \cdot 10^{-4}$, $1 \cdot 10^{-2}$, $5 \cdot 10^{-3}$, and $1 \cdot 10^{-3}$, respectively, and exponentially decayed at a rate of 0.1. The composite loss $\mathcal{L}_{total} = \mathcal{L}_1 + \lambda_{SSIM}(1 - \mathcal{L}_{SSIM}) + \lambda_{TV}\mathcal{L}_{TV}$ was utilized such that $\lambda_{SSIM} = 0.25$ and $\lambda_{TV} = 0.05$. The TV regularization volume size was set to 32 voxels. Optimization, learning rate, and loss parameters were unchanged from the values reported in [12], as additional grid-search fine-tuning of the control parameters did not improve training.

D. Evaluation Metrics

20×20 pixel regions of interest for lung tissue and air were manually selected for all scans. The mean and standard deviation of the contrast-to-noise ratio (CNR) and the full-width-at-half-maximum (FWHM) of the tissue-to-air edge spread function derivative were calculated as measures of contrast and sharpness, respectively.

III. Results

Evaluation of contrast (CNR) and sharpness (FWHM in pixels) over all samples is provided in Table II. Although the

TV filter results in a higher CNR in comparison to the 3DGS method ($CNR_{TV} = 8.09 \pm 3.47$, $CNR_{3DGS} = 6.74 \pm 2.51$), the comparative sharpness preserved with the 3DGS method is noticeable ($FWHM_{TV} = 15.80 \pm 16.21$, $FWHM_{3DGS} = 10.06 \pm 8.83$).

A comparison of all methods is also visualized with an example axial slice in Fig. 2. The lung tissue and air regions depicted allow for both qualitative and quantitative comparison of the dark-field signal contrast and sharpness between the two material edges. Gaussian and bilateral filters show poor performance for both CNR and FWHM, and the high CNR achieved by TV filtering is offset by over-smoothing at the lung tissue-air interface edge.

TABLE II: 20×20 Pixel Region-of-interest Comparison for Lung Tissue-to-Air Contrast and Sharpness Metrics with Various Post-processing Methods ($n = 20$, mean $\pm$ std)

Method	CNR ↑	FWHM ↓
Reference	2.44 ± 0.93	40.25 ± 11.32
Gaussian Filter	4.14 ± 1.48	31.99 ± 16.22
Bilateral Filter	6.44 ± 2.60	23.36 ± 16.33
TV Filter	8.09 ± 3.47	15.80 ± 16.21
3DGS	6.74 ± 2.51	10.06 ± 8.83

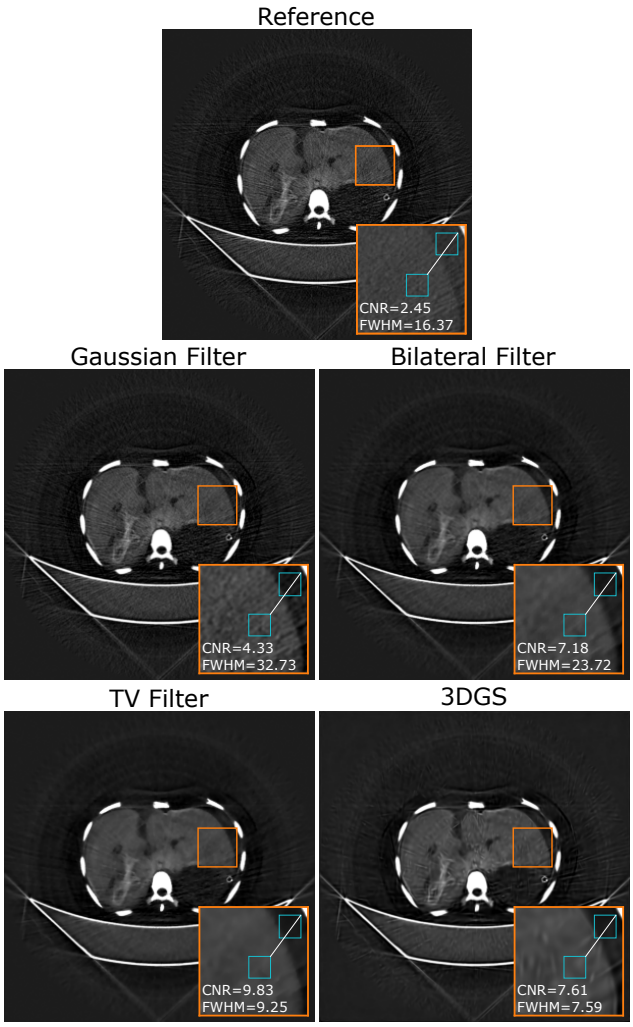


Fig. 2: Example axial DFCT slice, without (reference) and with streak-reduction post-processing. The contrast-to-noise ratio (CNR) and full-width-at-half-maximum (FWHM) values for the 20×20 pixel lung tissue and air regions (blue) are provided. The white line is used for pixel-wise FWHM calculation.

IV. Discussions and Outlook

The results suggest the potential of utilizing the 3DGS method for direct streak-free reconstruction of single-rotation DFCT data. Unlike the suggested 3DGS post-processing method, providing the raw projection data to the model ensures that no inherent FDK-reconstructed streaks will be forward-projected, further enhancing the CNR and FWHM for DFCT. However, a direct reconstruction of DFCT data using the 3DGS method is currently challenged by the intricate dark-field signal processing and phase retrieval of single-rotation data.

Applications of 3DGS for both post-processing and direct reconstruction of DFCT data, as well as the lack of domain constraints given the method's unsupervised nature, highlight its versatility and value. The proposed 3DGS post-processing explored in this work can be applied to further improve reference multi-rotation DFCT data, which is valuable for streak reduction in clinically relevant single-rotation DFCT scans within the supervised-learning approach introduced in [10].

The importance of dark-field image quality for diagnostic purposes is prevalent in the context of prospective clinical trials. To this end, future work should also validate the quality control and assessment of the various processing methods discussed in a clinically oriented reader study. Additionally, the 3DGS method is tailored for attenuation-based X-ray CT, and a fine-tuned adaptation of the proposed method to the dark-field imaging physics and CT processing should be investigated in more detail.

V. Conclusion

In this work, we compared post-processing of DFCT volumes with various conventional filtering and 3DGS methods. Our findings indicate that 3DGS outperforms conventional filtering in DFCT streak-reduction, preserving edge sharpness and a high CNR, a necessity for diagnostic accuracy in future clinical trials. While the study was limited to ex vivo phantom data, the findings suggest that 3DGS-based post-processing is a promising tool for improving DFCT image quality.

The availability of volumes reconstructed using classical methods, such as FBP, enables further extension of the proposed approach by fusing original and post-processed images, for example, via linear combination. In future work, enhanced 3DGS reconstructions could serve as improved GT for training supervised learning models.

Acknowledgment

Financial support through the European Research Council (ERC Smart Detectors for Darkfield X-ray Imaging, SyG

101167328), and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the States, as well as by the Technical University of Munich – Institute for Advanced Study.

REFERENCES

[1] Pfeiffer, F., Bech, M., Bunk, O., Kraft, P., Eikenberry, E. F., Brönnimann, C., Grünzweig, C., and David, C., "Hard-x-ray dark-field imaging using a grating interferometer," *Nature Materials* **7**(2), 134–137 (2008).

[2] Bech, M., Jensen, T. H., Bunk, O., Donath, T., David, C., Weitkamp, T., Le Duc, G., Bravin, A., Cloetens, P., and Pfeiffer, F., "Advanced contrast modalities for x-ray radiology: Phase-contrast and dark-field imaging using a grating interferometer," *Zeitschrift Fur Medizinische Physik* **20**(1), 7–16 (2010).

[3] Pfeiffer, F., Weitkamp, T., Bunk, O., and David, C., "Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources," *Nature Physics* **2**(4), 258–261 (2006).

[4] Urban, T., Gassert, F. T., Frank, M., Willer, K., Noichl, W., Buchberger, P., Schick, R. C., Koehler, T., Bodden, J. H., Fingerle, A. A., Sauter, A. P., Makowski, M. R., Pfeiffer, F., and Pfeiffer, D., "Qualitative and quantitative assessment of emphysema using dark-field chest radiography," *Radiology* **303**(1), 119–127 (2022).

[5] Urban, T., Sauter, A. P., Frank, M., Willer, K., Noichl, W., Bast, H., Schick, R., Herzen, J., Koehler, T., Gassert, F. T., Bodden, J. H., Fingerle, A. A., Gleich, B., Renger, B., Makowski, M. R., Pfeiffer, F., and Pfeiffer, D., "Dark-field chest radiography outperforms conventional chest radiography for the diagnosis and staging of pulmonary emphysema," *Investigative Radiology* **58**(11), 775–781 (2023).

[6] Gassert, F. T., Bast, H., Urban, T., Schick, R., Lochschmidt, M. E., Kaster, L., Koehler, T., Karrer, A., Keppler, A., Steinhardt, M., Marka, A. W., Steinhelfer, L., Sauter, A. P., Makowski, M. R., Pfeiffer, F., and Pfeiffer, D., "Dark-field chest radiography for pneumothorax detection: A prospective study," *Radiology. Cardiothoracic Imaging* **7**(6), e240560 (2025).

[7] Viermetz, M., Gustschin, N., Schmid, C., Haeusele, J., von Teuffenbach, M., Meyer, P., Bergner, F., Lasser, T., Proksa, R., Koehler, T., and Pfeiffer, F., "Dark-field computed tomography reaches the human scale," *Proceedings of the National Academy of Sciences* **119**(8), e2118799119 (2022). Publisher: Proceedings of the National Academy of Sciences.

[8] Schmid, C., Viermetz, M., Gustschin, N., Noichl, W., Haeusele, J., Lasser, T., Koehler, T., and Pfeiffer, F., "Modeling Vibrations of a Tiled Talbot-Lau Interferometer on a Clinical CT," *IEEE Transactions on Medical Imaging* **42**(3), 774–784 (2023).

[9] Haeusele, J., Schmid, C., Viermetz, M., Gustschin, N., Lasser, T., Koehler, T., and Pfeiffer, F., "Advanced phase-retrieval for stepping-free x-ray dark-field computed tomography," *IEEE Transactions on Medical Imaging* **42**(10), 2876–2885 (2023).

[10] Kumschier, T., Thalhammer, J., Schmid, C., Haeusele, J., Koehler, T., Pfeiffer, F., Lasser, T., and Schaff, F., "Streak artefact removal in x-ray dark-field computed tomography using a convolutional neural network," *Medical Physics* **51**(10), 7404–7414 (2024).

[11] Kerbl, B., Kopanas, G., Leimkuehler, T., and Drettakis, G., "3D Gaussian Splatting for Real-Time Radiance Field Rendering," *ACM Transactions on Graphics* **42**(4), 1–14 (2023).

[12] Zha, R., Lin, T. J., Cai, Y., Cao, J., Zhang, Y., and Li, H., "R²-gaussian: Rectifying radiative gaussian splatting for tomographic reconstruction," (2024).

Alphabetical Name Index

Names consolidated from the supplied page-number listings

Name	Page(s)
Abbey, Craig K.	42
Ahlers, Jannis	210
An, Yunqiang	381
Andrew, Matthew	327
Andreyev, Andriy	131
Araujo, Stephen	345
Aydogan, Bulent	91
Baek, Jongduk	20, 314, 331, 502
Baer-Beck, Matthias	190, 241
Baffour, Francis	274
Bai, Chuanyong	117
Bai, Yunkun	319
Berry, Sean	104
Bert, Christoph	190
Bian, Zhaoying	55, 200, 397, 491
Bleuel, Paulina	227, 524
Boedeker, Kirsten	246, 282
Boiset, Gisell	195, 265
Boone, John M.	42
Brown, Kevin M.	452, 470
Cagnon, Daniel	117
Cai, Liang	350
Cao, Wenjing	373
Carter, Rickey	67
Chang, Shaojie	364
Chang, William	402
Chen, Buxin	416
Chen, Changyu	81, 295, 381, 475
Chen, Guang-Hong	157, 420, 479, 487
Chen, Hong	51

Name	Page(s)
Chen, Li	140, 144
Chen, Qiule	63
Chen, Ruifeng	144, 369, 385
Chen, Yang	144, 369, 377
Chen, Yuting	507
Chen, Zhiqiang	81, 381, 475
Chiang, Jing-Tzyh	38
Cho, Hyosung	126
Cho, Seungryong	113, 424
Colbeth, Richard E.	42, 174, 178
Cormode, David	148
Costa, Paulo	195, 265
Costello, Lucy	210
Cross, Andrew	51
Curtis, Tyler	157
D'Amico, Lorenzo	210
Daerr, Heiner	470
Dai, Dongrui	381
Davis, Aeden	274
Delorme, Guillaume	483
Demianova, Sofia	232
Dettmer, Jannik	190
Dhanaliwala, Ali	148, 186
Dhanantwari, Amar	246
Din, Huay	71
Dinh, Mark	174
DiNitto, Julie	393
Donnelley, Martin	210
Dorosti, Tina	210, 227, 232, 524
Du, Muge	381
Duan, Xinhui	182
Eo, Yoonha	314, 331
Erath, Julien	25, 86, 190
Fair, Ryan	148
Fan, Shuyi	497

Name	Page(s)
Fan, Yue	77
Fang, Changsheng	497
Fang, Kaylee	269
Feng, Juan	144
Fletcher, Joel G.	67
Foley, Thomas	364
Fonseca, Jordan	291
Fournié, Eric	86
Frach, Thomas	174
Frey, Daniel	227, 232, 524
Fu, Lin	359
Fu, Zhiyang	373
Fujii, Katsuya	406
Fung, George	205
Gang, Grace J.	71, 77, 108, 355
Gao, Hewei	29, 81, 475, 507
Gao, Yang	381
Gao, Yu	428, 466
Gao, Yuanxiang	33, 63, 491
Ge, Yinglin	470
Geagan, Michael	95, 100
Glinz, Jonathan	340
Gong, Hao	67, 274
Gonsalves, Wilson	274
Grasruck, Michael	241
Grewar, Murdock	305
Guo, Zijia	393
Hagio, Tomoe	246, 282
Han, Shuo	497
Haneda, Eri	345
Haneishi, Hideaki	323
Harris, Wendy	104
He, Yutong	443
Heinkele, Andreas	25, 86

Name	Page(s)
Hennemann, Lukas	25, 86
Hernandez, Andrew M.	42
Hilmer, Josepha	227, 232, 524
Hiu, Theresa	227
Hoelzer, Steve	174
Hoffman, John	256
Hofmann, Christian	25
Holland, Kyle	174, 178
Holt, Kevin	174
Hoppel, Bernice	246
Hou, Haomin	33, 200, 491
Hsiao, Albert	345
Hsieh, Scott S.	59, 67, 166, 169
Hu, Haoxiang	140
Hu, Hui	51
Huang, Bangyan	511
Hyun, Subong	424
Im, Jessica	95, 100, 246
Imae, Toshikazu	448
Inoue, Akitoshi	67
Iyadomi, Taro	402
Jain, Amit	117
Jansen, Isabelle Heukensfeldt	345
Jeon, Kiwan	310
Ji, Xu	144
Jiang, Lily	402
Jiang, Xiao	77, 108, 237
Jiang, Zilin	420, 479
Jin, Shujie	420, 479
Jin, Yannan	470
Joen, Duhee	223
Jolivet, Frédéric	483
Juhé, Aniol Serra	252, 287
Jung, Hojin	20
Kabelac, Anton	121

Name	Page(s)
Kabongo, Luis	335
Kachelrieß, Marc	25, 86, 121, 241, 252, 287
Kaji, Shizuo	448
Kalra, Mannudeep	460
Kastner, Johann	340
Kayser, Lennard	232
Kharat, Shravani	274
Kim, Anna	402
Kim, Grace Hyun	71, 182
Kinahan, Paul	519
Kitchen, Marcus	210
Knaup, Michael	252
Knobloch, Catherine	241
Koehler, Thomas	227, 232, 470
Kong, Deyang	33, 200, 491
Kranioti, Elena	340
Kudo, Hiroyuki	406
Lai, Ruiran	157
Lee, Eunmin	331
Lee, Okkyun	278
Lee, Seoyoung	274, 424
Lee, Tzu-Cheng	350
Lell, Michael	252, 287
Leng, Shuai	67, 274, 364
Lesaint, Jérôme	218
Létang, Jean Michel	218
Li, Ke	157, 479
Li, Limin	33, 55
Li, Ou	282
Li, Quanzheng	456
Li, Shudong	355
Li, Yinsheng	140, 144, 369, 377, 385
Li, Yixin	411
Li, Yule	33, 55, 63, 200, 491

Name	Page(s)
Li, Zhang	389
Liang, Dong	140, 144, 369, 385
Liao, Jingyi	33, 55, 63, 200, 397, 491
Liao, Suyu	355
Lin, Zhencheng	299
Lingel, Konrad	327
Liu, Chang	287
Liu, Jiaming	42
Liu, Leening	148, 182, 186, 246, 470
Liu, Xiangyuan	135
Liu, Yang	377
Liu, Yi	385
Liu, Yue	397
Lochschmidt, Maximilian	232
Lohman, Henrik	174, 178
Lorenzon, Altea	237
Lu, Bin	381
Lyu, Yun-Huan	519
Ma, Canzhen	385
Ma, Jianhua	33, 55, 63, 200, 397, 491
Maier, Andreas	252, 287
De Man, Bruno	157, 345, 359
Manhart, Michael	393
Mao, Hui	223
Mao, Yifu	373
Mathew, Manoj	186
McClung, George	186
McCollough, Cynthia H.	67, 169, 274, 364
McGinnis, Julian	227, 524
McNitt-Gray, Michael	71, 182
McVeigh, Elliot	345
Medrano, Maria Jose	269
Mei, Kai	95, 100

Name	Page(s)
Meng, Xiao	397
Meyer, Sebastian	95
Miao, Donghang	214
Michielsen, Koen	282, 434
Mirzaian, Nicole	148
Missert, Andrew D	67
Mo, Wenxin	29
Mollov, Ivan	174
Mondal, Parmita	67
van Duijnen Montijn, Meo	218
Moran, Jean	104
Morgan, Kaye	210
Morimoto, Yuki	323
Morovati, Bahareh	497
Mosig, Daniel	252, 287
Multari, Gauthier	483
Najaf, Mahrokh	439
Nakagawa, Keiichi	448
Narum, Solveig	169
Nassi, Martina	434
Nawa, Kanabu	448
Nepelius, Lukas	340
Nilsen, Roy	157
Noël, Peter	95, 100, 148, 182, 186, 246, 470
Noo, Frédéric	305, 393
Nowak, Tristan	148, 186
Oelze, Philipp	241, 252, 287
Okamoto, Takayuki	323
Oktay, Mehmet Bugrahan	393
Ongie, Greg	439
Oostveen, Luuk	265, 282
Ozaki, Sho	448
Pan, Xiaochuan	91, 416
Pantsar, Tuomas	174

Name	Page(s)
Papadakis, Nicolas	335
Parada, Ricardo	402
Park, Hyoung Suk	310
Park, Subin	314
Pasyar, Pouyan	100
Pearson, Erik	91
Peng, Jiangjun	397
Perkins, Amy	470
Peterhansl, Sebastian	210, 227, 524
Petersilka, Martin	25, 86
Pfeiffer, Daniela	210, 227, 232, 524
Pfeiffer, Franz	210, 227, 232, 524
Phillips, JP	416
Pimenta, Elsa	265
Plank, Bernhard	340
Polster, Christoph	205
Prokein, Andreas	241
Proksa, Roland	470
Punzer, Daniel	515
Qi, Jinyi	511
Qi, Zhihua	479
Quan, Guotao	373
Rajendran, Kishore	182
Reinholdsen, Fredrik	174, 178
Reiser, Ingrid	416
Ren, Liqiang	182
Riehl, Patrick	166
Ripoche, Xavier	335
Rit, Simon	218, 335, 483
Robinson, Vance	42
Rose, Georg	515
Roshkovan, Leonid	100
Ru, Xudong	135
Rueckert, Daniel	227, 524

Name	Page(s)
Rybertt, Martin	148, 182, 186
Sadeghian, Maryam	364
Sahbaee, Pooyan	148, 186
Sandvold, Olivia	470
Santos, Renan	195
Savaliya, Devang	174, 178
Schaff, Florian	210, 227, 524
Schuele, Johannes	327
Schwoebel, Paul	42
Sechopoulos, Ioannis	265, 282, 434
Senck, Sascha	340
Shao, Muhan	359
ShaoYang, Zhang	389
Shea, Jacob	117
Shen, Le	381
Sheng, Jiabing	200
Shi, Gengxin	42
Shi, Wuliang	295
Shi, Yu	497
Shim, Jiyong	126
Shin, Daniel	246, 282
Sidky, Emil	91, 402, 411, 416
Siewerdsen, Jeffrey H.	42
Silva, Gustavo	195
Silverberg, Noah	104
Sisniega, Alejandro	42
Smith, Ronan	210
Song, Lei	369
Specovius, Andreas	121
Star-Lack, Josh	174, 178
Stayman, J. Webster	71, 77, 108, 237, 350, 355
Stevens, Grant	269
Stewart, Alex	174, 178
Stierstorfer, Karl	25, 86, 162, 190, 205

Name	Page(s)
Sun, Liyan	269
Sun, Zexi	377, 385
Sun, Zhongwei	369
Sunnegaardh, Johan	25
Szkitsak, Juliane	190
Taguchi, Katsuyuki	166, 205
Tang, Qiang	140, 144
Tang, Xiangyang	223
Tatachak, Idris	335
Teng, Peiqing	350, 355
Thalhammer, Johannes	210, 227, 524
Tian, Jiao	373
Tivnan, Matthew	456
Tomal, Alessandra	195, 265
Trieb, Klemens	340
Tse, Justin	269
Tseng, Hsin Wu	38
Ullberg, Christer	174
Urech, Mattias	174, 178
Vedantham, Srinivasan	38
Veijola, Simo	174, 178
Villasenor, Katie	148
Wang, Adam	269, 502
Wang, Dayang	497
Wang, Ge	359, 460
Wang, Hao	33, 55, 63, 200, 397, 491
Wang, Jianxu	460
Wang, Sen	269
Wang, Shuo	381
Wang, Wenying	152, 373
Wang, Xin	350
Wang, Yankun	33, 55, 63, 200, 491
Wang, Yuang	381, 456
Wei, Jia	205

Name	Page(s)
Weinberger, Patrick	340
Wellinghoff, Jarod	169
Wiedeman, Christopher	359
Williamson, Eric	364
Wilson, Benjamin	364
Wu, Dufan	456
Wu, Junhao	33, 63, 200, 491
Wu, Pengwei	345
Wu, Wei	519
Wu, Xiangyi	411
Wu, Yijing	479
Xia, Dan	91
Xia, Dimeng	443
Xia, Wenjun	460
Xia, Yinxian	29
Xiang, Jun	140, 144, 369, 385
Xiao, Meng	397
Xie, Huiqiao	223
Xing, Lei	104, 428
Xing, Yuxiang	81, 214, 295, 319, 475
Yamashita, Hideomi	448
Yang, Faguo	131
Yang, Sun	389
Yang, Yi	223
Yang, Yirong	152
Yanoff, Brian	157
Ye, Bonghong	55
Ye, Siqi	428
Yu, Hengyong	497
Yu, Lifeng	67, 169, 274
Yu, Minwoo	502
Yu, Mitchell	104
Yu, Xiaopeng	157, 487
Yu, Zhicong	117, 299

Name	Page(s)
Yuan, Mengxuan	369
Yuan, Yijie	71
Yuan, Yuedong	515
Yun, Sungho	113
Zabic, Stanislav	452
Zbijewski, Wojciech	42
Zeng, Dong	33, 55, 63, 200, 397, 491
Zeng, Larry	47, 261
Zhan, Linying	157
Zhang, Di	182
Zhang, Hao	104, 428
Zhang, Jiaying	299
Zhang, Li	81, 381, 475
Zhang, Ran	157, 420, 479, 487
Zhang, Ruoqiao	350
Zhang, Xi	373
Zhang, Xun	475
Zhang, Yikun	377
Zhang, You	466
Zhang, Yuzheng	519
Zhang, Zheng	91, 416
Zhang, Zhipeng	55
Zhao, Bingshuai	377
Zhao, Shengzi	214
Zhao, Shusen	135, 443
Zhao, Wei	411
Zhou, Hao	29, 507
Zhou, Jian	350, 511
Zhou, Li	497
Zhou, Zeyu	131
Zhou, Zhongxing	169
Zhu, Guoxi	81
Zhuang, Tingliang	466
Ziskin, Vitaliy	51

Name	Page(s)
Zou, Xiang	319

Sunday, May 31	Monday, June 1	Tuesday, June 2	Wednesday, June 3
	8:00 – 8:10 **Opening remarks**	7:30 – 10:30 **Photon counting CT** **Refresher course:** physics, clinical use, future directions (7:30 – 8:30) **Oral session T1** (8:50 – 10:30)	8:00 – 8:10 **Future meetings**
	8:10 – 9:50 **Oral session M1: Novel CT imaging concepts**		8:10 – 9:50 **Oral session W1: Angiographic, cardiac & high-resolution imaging**
	9:50 – 10:10 **Coffee break**	10:30 – 10:50 **Coffee break**	9:50 – 10:10 **Coffee break**
	10:10 – 11:30 **Oral session: M2** Top Four Abstracts	10:50 – 11:50 **Oral session T2 : Beyond-absorption X-ray imaging**	10:10 – 11:30 **Oral session W2: Limited angle & sparse view imaging**
	11:35 – 12:10 **Poster Teasers**		11:35 – 12:10 **Poster Teasers**
	12:10 – 13:30 **Lunch break**	11:50 – 13:30 **Poster exchange & Lunch break**	12:10 – 13:30 **Lunch break**
14:00 – 16:00 **Workshop on CT-omics** (J. Liang, Granite Ballroom)	13:30 – 15:30 **Poster Session I** (posters to be installed by 8am and removed the next day during lunch break)	13:30 – 15:50 **Image quality, radiomics & dose** **Refresher course :** Diffusion models and applications (13:30-14:15) **Oral session T3** (14:20-16:00)	13:30 – 15:30 **Poster Session II** (posters to be installed during Tuesday's lunch break and removed by 6pm)
	15:30 – 16:00 **Coffee break**	16:00 – 16:20 **Coffee break**	15:30 – 16:00 **Coffee break**
16:00 – 18:30 **Registration:** University of Utah Guesthouse – Granite Ballroom lobby	16:00 – 17:40 **Oral session M3: Handling non-ideal conditions (incomplete data, metal, motion, geometry, rings, scatter)**	16:20 – 17:40 **Oral session T4: NDT & Dental Imaging**	16:00 – 17:40 **Oral session W3 Addressing noise, low dose and material decomposition**
18:30 – 21:30 **Welcoming reception:** University of Utah Guesthouse Plaza and Granite Ballroom	18:30 – 21:30 **Dinner** Ken Garff University Club at the Rice-Eccles Stadium	18:30 – 21:30 **Dinner** Rodizio Grill, Trolley Square	18:30 – 21:30 **Dinner** University of Utah Guesthouse Plaza and Granite Ballroom