The 8th International Conference on Image Formation in X-Ray Computed Tomography



CT MEETING 2024

August 5-9, Bamberg, Germany

Program and Proceedings

This event is supported by:

Canon

DECTRIS detecting the future

GE HealthCare

PHILIPS

QRM A PTW COMPANY

SIEMENS Healthineers

UNITED IMAGING

ziehm imaging

The 8th International Conference on Image Formation in X-Ray Computed Tomography is supported by:

















We thank our sponsors for their important and valuable contributions!

We would like to thank the members of the scientific committee for their great work by taking the opportunity to mention them in alphabetical order:

Scientific Committee

Name	**Institution**
Adam Wang	Stanford University, USA
Bruno De Man	GE Research, USA
Cristian Badea	Duke University, USA
Cyril Riddell	GE HealthCare, France
Emil Sidky	University of Chicago, USA
Frédéric Noo	University of Utah School of Medicine, USA
Ge Wang	Rensselaer Polytechnic Institute, USA
Grace Gang	University of Pennsylvania, USA
Guang-Hong Chen	University of Wisconsin-Madison, USA
Hao Zhang	Memorial Sloan Kettering Cancer Center, USA
Hengyong Yu	University of Massachusetts Lowell, USA
Ioannis Sechopoulos	Radboud University Medical Center, The Netherlands
Jeff Siewerdsen	The University of Texas MD Anderson Cancer Center, USA
Jerome Liang	Stony Brook University, USA
Jing Wang	University of Texas Southwestern Medical Center, USA
John Marian Hoffman	University of California, Los Angeles, USA
Jongduk Baek	Yonsei University, South Korea
Joscha Maier	German Cancer Research Center, Germany
Karl Stierstorfer	Siemens Healthineers, Germany
Ke Li	University of Wisconsin-Madison, USA
Ken Taguchi	Johns Hopkins University School of Medicine, USA
Kirsten Boedeker	Canon Medical Systems Corporation, USA
Lifeng Yu	Mayo Clinic, USA
Ludwig Ritschl	Siemens Healthineers, Germany
Marc Kachelrieß	German Cancer Research Center, Germany
Mats Persson	KTH Royal Institute of Technology, Sweden

Nicole Maaß	Siemens Healthineers, Germany
Pascal Paysan	Varian Medical Systems Imaging Lab, Switzerland
Peter Noël	University of Pennsylvania, USA
Rongping Zeng	U.S. 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
Stefan Sawall	German Cancer Research Center, Germany
Thomas Köhler	Philips Research Hamburg, Germany
Tobias Lasser	Technical University of Munich, Germany
J. Webster Stayman	Johns Hopkins School of Medicine, USA
Wenli Wang	Avant Tomography Consulting LLC, USA
Xiaochuan Pan	University of Chicago, USA
Yuxiang Xing	Tsinghua University, China
Zhicong Yu	Accuray, USA
Zhou Yu	Canon Medical Research, USA
Zhye Yin	GE HealthCare, USA

Conference Chair

Name	Institute
Marc Kachelrieß	German Cancer Research Center (DKFZ)

Organizing Committee

Name	Institute
Andrea Kachelrieß	Society of High Performance Computational Imaging (SHPCI) e. V.
Yannick Berker	Society of High Performance Computational Imaging (SHPCI) e. V.
Achim Byl	German Cancer Research Center (DKFZ)
Andreas Heinkele	German Cancer Research Center (DKFZ)
Edith Baader	German Cancer Research Center (DKFZ)
Elias Eulig	German Cancer Research Center (DKFZ)
Fabian Jäger	German Cancer Research Center (DKFZ)
Jan Magonov	German Cancer Research Center (DKFZ)
Johannes Schneider	German Cancer Research Center (DKFZ)
Joscha Maier	German Cancer Research Center (DKFZ)
Lukas Hennemann	German Cancer Research Center (DKFZ)
Maximilian Hammermann	German Cancer Research Center (DKFZ)
Philip Trapp	German Cancer Research Center (DKFZ)
Sophia Klubertz	German Cancer Research Center (DKFZ)
Stefan Sawall	German Cancer Research Center (DKFZ)

Schedule CT Meeting 2024

	Monday *August 5*	**Tuesday** *August 6*	**Wednesday** *August 7*	**Thursday** *August 8*	**Friday** *August 9*
8:00 - 9:40 CEST		**Recon-struction** General opening remarks + 4 talks *(Page 7)*	**Photon-Counting** Basic short course + 3 talks *(Page 12)*	**AI and Physics** Basic short course + 3 talks *(Page 16)*	**Image Quality I** Basic short course + 3 talks *(Page 20)*
		Coffee	**Coffee**	**Coffee**	**Coffee**
10:00 - 11:40 CEST		**Missing Data** 5 talks *(Page 7)*	**Spectral** 5 talks *(Page 12)*	**Artifacts** 5 talks *(Page 16)*	**Image Quality II** 4 talks + Preview of Fully 3D 2025 and CT Meeting 2026 *(Page 20)*
		Lunch	**Lunch**	**Lunch**	**Lunch**
13:20 - 15:00 CEST		**Poster** Poster FF + 24 Posters *(Page 8)*	**Poster** Poster FF + 24 Posters *(Page 13)*	**Poster** Poster FF + 24 Posters *(Page 17)*	**Visit to EZRT** *(Page 21)*
		Coffee	**Coffee**	**Coffee**	
15:20 - 17:00 CEST	**Registration**	**Motion** 5 talks *(Page 10)*	**Deep and Industrial** 5 talks *(Page 15)*	**Novel CT** 5 talks *(Page 19)*	
17:00 - 17:40 CEST		**Plenary** Stephan Achenbach			
19:00 - 22:00 CEST	**Registration and Reception at Restaurant Plückers**	**Dinner at Residenz-schloss**		**Dinner at Klosterbräu**	

DECTRIS

Fast and **Spectral Photon Counting CT** in **Seconds**



PILATUS 4

NEW

up to **1** Mpixel

up to **4** kHz

4 energy thresholds

150 µm pixels, zero noise

EIGER 2

up to **4** Mpixel

up to **4** kHz

2 energy thresholds

75 µm pixels, zero noise

Tuesday, August 6

Oral Session: Reconstruction
Time : Tuesday, 08:00 – 09:40
Chairs : Simon Rit and Wenli Wang

Time	Author	Title	Page
08:00 – 08:20	Marc Kachelrieß	*Opening remarks*	
08:20 – 08:40	Emil Y. Sidky, Xiangyi Wu, Xiaoyu Duan, Hailing Huang, Wei Zhao, Leo Y. Zhang, John Paul Phillips, Zheng Zhang, Buxin Chen, Dan Xia, Ingrid S. Reiser, Xiaochuan Pan	*A directional total variation minimization algorithm for isotropic resolution in digital breast tomosynthesis*	22
08:40 – 09:00	Yuedong Yuan, Linda-Sophie Schneider, Andreas Maier	*Application of gated recurrent units for CT trajectory optimization*	26
09:00 – 09:20	Shudong Li, Xiao Jiang, Yuan Shen, J. Webster Stayman	*CT reconstruction using nonlinear diffusion posterior sampling with detector blur modeling*	30
09:20 – 09:40	Zheng Zhang, Buxin Chen, Dan Xia, Emil Y. Sidky, Xiaochuan Pan	*Improved accuracy in CT reconstruction from transversely truncated data*	34

Coffee Break

Oral Session: Missing Data
Time : Tuesday, 10:00 – 11:40
Chairs : Xiaochuan Pan and Yuxiang Xing

Time	Author	Title	Page
10:00 – 10:20	Nimu Yuan, Jian Zhou, Jinyi Qi	*KVp-switching sparse-view CT images restoration with an edge prior conditioned DDPM*	38
10:20 – 10:40	Xiaoyue Li, Mark D. Butala, Gaoang Wang	*DiffusionX2CT: zero-shot 3D CBCT reconstruction from sparse-view X-rays with warping-based diffusion model*	42
10:40 – 11:00	Jiacheng Xie, Hua-Chieh Shao, Yunxiang Li, You Zhang	*Prior frequency guided diffusion model for limited angle (LA)-CBCT reconstruction*	46
11:00 – 11:20	Zhizhong Ran, Dimeng Xia, Shusen Zhao , Weibin Zhang, Xiangyuan Liu	*Diffusion model based limited-angle CT reconstruction for flat objects*	50
11:20 – 11:40	Anton Kabelac, Elias Eulig, Joscha Maier, Marc Kachelrieß	*Latent space reconstruction and its application to CT detruncation: latent detruncation*	54

Lunch Break

Poster Session: Poster Session 1
Time : Tuesday, 13:20 – 15:00

	Time	Author	Title	Page
	13:20 – 13:40	**Poster Fast Forward**: 30 s - 60 s presentation for each poster		
P1.1	13:40 – 15:00	Gengsheng L. Zeng	*Morphing from the TV-norm to the l0-norm*	58
P1.2	13:40 – 15:00	Shalini Subramanian, W. Paul Segars, Ting-Yim Lee, Andreia V. Faria, Katsuyuki Taguchi	*CT brain perfusion assessment using X-ray angiography images*	62
P1.3	13:40 – 15:00	Zhennong Chen, Matthew Tivnan, Siyeop Yoon, Rui Hu, Quanzheng Li, Dufan Wu	*Optimizing conditional DDPM for head CT motion artifact reduction: brain vs. skull and 3D vs. 2D*	66
P1.4	13:40 – 15:00	Xiangyuan Liu, Shusen Zhao , Dimeng Xia, Weibin Zhang, Fengwei An	*Ring artifacts removal from CT images using an adaptive patch filter in the frequency domain*	70
P1.5	13:40 – 15:00	Fabian Jäger, Pascal Paysan, Michał Walczak, Marc Kachelrieß	*Rawdata-based higher order deep beam hardening correction (DBHC*	74
P1.6	13:40 – 15:00	Yixing Huang, Fuxin Fan, Ahmed Gomaa, Andreas Maier, Rainer Fietkau, Christoph Bert, Florian Putz	*Task-specific data preparation for deep learning to reconstruct structures of interest from severely truncated CBCT data*	78
P1.7	13:40 – 15:00	Siyuan Mei, Fuxin Fan, Fabian Wagner, Mareike Thies, Mingxuan Gu, Yipeng Sun, Andreas Maier	*Segmentation-guided knee radiograph generation using conditional diffusion models*	82
P1.8	13:40 – 15:00	Gernot Kristof, Achim Byl, Elias Eulig, Marc Kachelrieß	*Noise-augmented deep denoising of CT images*	86
P1.9	13:40 – 15:00	Linda-Sophie Schneider, Gabriel Herl, Andreas Maier	*Integer optimization of CT trajectories using a discrete data completeness formulation*	90
P1.10	13:40 – 15:00	Fuxin Fan, Jingna Qiu, Yixing Huang, Andreas Maier	*Enhancing cross-modality synthesis: subvolume merging for MRI-to-CT conversion*	94
P1.11	13:40 – 15:00	Yuang Wang, Siyeop Yoon, Rui Hu, Baihui Yu, Duhgoon Lee, Rajiv Gupta, Li Zhang, Zhiqiang Chen, Dufan Wu	*Noise controlled CT super-resolution with conditional diffusion model*	98
P1.12	13:40 – 15:00	Yirong Yang, Sen Wang, Grant M. Stevens, Adam S. Wang	*Tube voltage modulation for reducing noise in photon-counting CT*	102

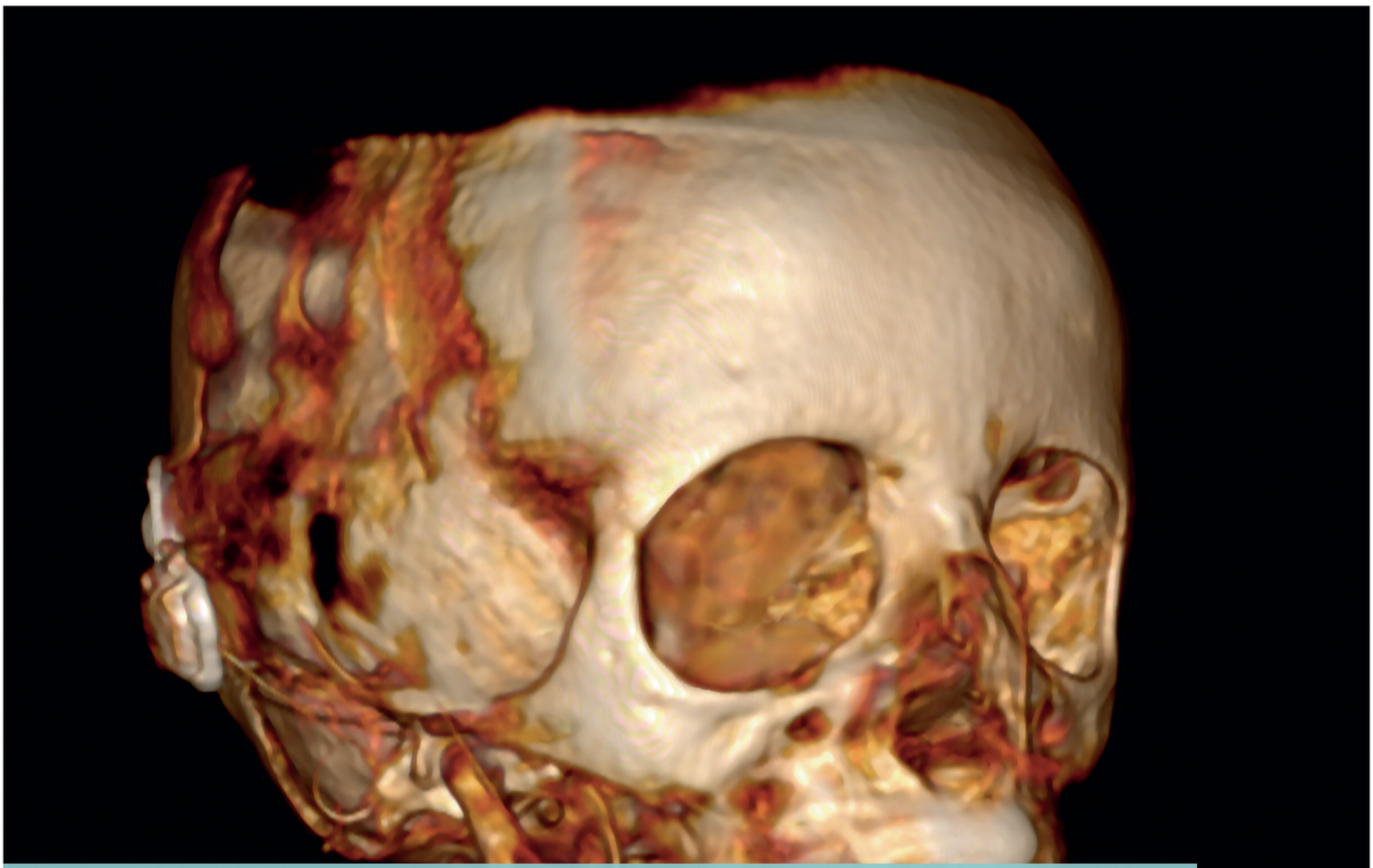
P1.13	13:40 – 15:00	Hongyang Zhu, Xin Lu, Yanwei Qin, Xinran Yu, Tianjiao Sun, Yunsong Zhao	*A dual-domain regularization method for ring artifact removal of X-ray CT*	106
P1.14	13:40 – 15:00	Maximilian Hammermann, Elias Eulig, Joscha Maier, Marc Kachelrieß	*Image-to-image translation for spatial alignment of sequential dual-energy CT acquisitions*	110
P1.15	13:40 – 15:00	Johannes Thalhammer, Manuel Schultheiß, Tina Dorosti, Tobias Lasser, Franz Pfeiffer, Daniela Pfeiffer, Florian Schaff	*Improving automated hemorrhage detection in sparse-view computed tomography via U-Net based artifact reduction*	114
P1.16	13:40 – 15:00	Margarete Kattau, Oliver Pickford Scienti, Ben Glocker, Dimitra Darambara	*Deep learning-based tumour segmentation on spectral photon-counting CT images*	118
P1.17	13:40 – 15:00	Alexander Jung, Gabriel Herl, Andreas Maier	*Towards efficient truncation correction in arbitrary geometry: residual projections for region of interest algebraic reconstruction*	122
P1.18	13:40 – 15:00	Ruifeng Chen, Yang Chen, Dong Liang, Yinsheng Li	*PNIR: a physics-constrained and noise-controlled iterative reconstruction method for photon-counting computed tomography*	126
P1.19	13:40 – 15:00	Yipeng Sun, Linda-Sophie Schneider, Fuxin Fan, Mareike Thies, Mingxuan Gu, Siyuan Mei, Yuzhong Zhou, Siming Bayer, Andreas Maier	*Data-driven filter design in FBP: transforming CT reconstruction with trainable fourier series*	130
P1.20	13:40 – 15:00	Jianpeng Gao, Liang Li, Liang Sheng, Yang Li, Baojun Duan, Yan Song, Dongwei Hei	*Diffusion model-based three-dimensional image reconstruction algorithm from five-view projections*	134
P1.21	13:40 – 15:00	Hsin Wu Tseng, Zhiyang Fu, Srinivasan Vedantham	*Simultaneous radiation dose and scatter reduction in breast CT imaging using a nearly half-sized detector with a self-supervised deep-learning algorithm*	138
P1.22	13:40 – 15:00	Richard N.K. Bismark, Georg Rose	*Preliminary simulation results in model-based iterative DECT with respect to convergence speed, truncation and dose-reduction via double half scan geometry*	142
P1.23	13:40 – 15:00	Min Le, Chunguang Zong, Liang Li	*Performance study of novel silicon photomultiplier devices in scintillator-based photon-counting detector for X-ray CT application*	146
P1.24	13:40 – 15:00	Okkyun Lee, Joonbeom Kim	*Calibration-based statistical framework for material decomposition in photon-counting CT*	150

Oral Session : Motion
Time : Tuesday, 15:20 – 17:00
Chairs : Pascal Paysan and Cyril Riddell

Time	Author	Title	Page
15:20 – 15:40	Abdul Salam Rasmi Asraf Ali, Cristina Sarti, Claudio Landi, Andrea Fusiello	*Partial angle motion compensation for maxillo-facial cone beam CT*	154
15:40 – 16:00	Mareike Thies, Fabian Wagner, Noah Maul, Siyuan Mei, Mingxuan Gu, Laura Pfaff, Nastassia Vysotskaya, Haijun Yu, Andreas Maier	*On the influence of smoothness constraints in computed tomography motion compensation*	158
16:00 – 16:20	Katsuyuki Taguchi, Shalini Subramanian, W. Paul Segars	*Motion-compensated time-enhancement curve estimation for intra-operative brain perfusion assessment using a standard x-ray system with no gantry rotation*	162
16:20 – 16:40	Heyuan Huang, A. Lu, F. Gonzalez, T. Ehtiati, J. H. Siewerdsen, A. Sisniega	*Motion compensation for 4D digital subtraction angiography via deep autofocus and implicit neural motion models*	166
16:40 – 17:00	Hua-Chieh Shao, Tielige Mengke, Tinsu Pan, You Zhang	*Dynamic CBCT reconstruction using prior model-free spatiotemporal implicit neural representation (PMF-STINR)*	170

Plenary Lecture: Cardiac CT: Evolution, Clinical Use and Insight
Time : Tuesday, 17:00 – 17:40
Author : Stephan Achenbach

In his plenary presentation “Cardiac CT: Evolution, Clinical Use and Insight”, Stephan Achenbach will highlight the high technical demands on CT for cardiac imaging as well as the evolution of technology, image quality and clinical applications in the past two decades, including the latest information on photon counting CT. He will also speak about insights that CT has helped provide regarding the mechanisms and epidemiology of coronary atherosclerosis, coronary artery disease and myocardial infarction as well as open questions that remain.



WE ARE LOOKING FOR BRIGHT MINDS

INNOVATION LEADER FOR IMAGING SYSTEMS

MOBILE C-ARMS, MADE IN GERMANY

Image represents a prototype. Not for clinical use

To our job vacancies:
www.ziehm.com/career

Ziehm Imaging GmbH, Nürnberg

ziehm imaging

Wednesday, August 7

Oral Session: Photon-Counting
Time : Wednesday, 08:00 – 09:40
Chairs : Ken Taguchi and Mats Persson

Time	Author	Title	Page
08:00 – 08:40	Thomas Flohr	*Basic short course: photon-counting CT*	
08:40 – 09:00	Linying Zhan, Guang-Hong Chen, Ke Li	*Detector deadtime estimation in photon-counting CT: a novel image-domain approach*	174
09:00 – 09:20	Okkyun Lee	*Application of the calibration-based statistical framework for material decomposition to digital mammography: a preliminary study*	178
09:20 – 09:40	Yukang Song, Xuezhi Wang, Zhi Deng, Hewei Gao	*Pileup-proof material decomposition for spectral CT using a simultaneous photon-counting and energy-integrating readout ASIC*	182

Coffee Break

Oral Session: Spectral
Time : Wednesday, 10:00 – 11:40
Chairs : Ke Li and Peter B. Noël

Time	Author	Title	Page
10:00 – 10:20	Altea Lorenzon, Stephen Z. Liu, Xiao Jiang, Grace J. Gang, J. Webster Stayman	*Joint material decomposition and scatter estimation for spectral CT*	186
10:20 – 10:40	Grace J. Gang	*Spectral orbits: combining spectral imaging and non-circular orbits for interventional CBCT*	190
10:40 – 11:00	John Paul Phillips, Emil Y. Sidky, Fatma Terzioglu, Ingrid S. Reiser, Guillaume Bal, Xiaochuan Pan	*Non-unique water and contrast agent solutions in dual-energy CT*	194
11:00 – 11:20	Edith Baader, Marc Kachelrieß	*Tube voltage modulation for contrast-enhanced CT scans*	198
11:20 – 11:40	Buxin Chen, Zheng Zhang, Dan Xia, Emil Y. Sidky, Xiaochuan Pan	*A basis-region data model for reconstructing multiple basis material from DECT data*	201

Lunch Break

Poster Session: Poster Session 2
Time : Wednesday, 13:20 – 15:00

	Time	Author	Title	Page
	13:20 – 13:40	**Poster Fast Forward**: 30 s - 60 s presentation for each poster		
P2.1	13:40 – 15:00	Gengsheng L. Zeng	*Extreme few-view tomography without training data*	205
P2.2	13:40 – 15:00	Jonathan Bollig, Fabian Jäger, Marc Kachelrieß	*Deep bone extraction in X-ray projection domain*	209
P2.3	13:40 – 15:00	Hana Haseljic, Robert Frysch, Vojtech Kulvait, Thomas Werncke, Oliver Speck, Georg Rose	*The impact of arterial input function selection on dynamic liver perfusion imaging with cone-beam CT*	213
P2.4	13:40 – 15:00	Zeyu Yang, Raffael Lutz, Frank G. Zollner	*Development of a reusable 3D-printed phantom for contrast agent evaluation in photon-counting computed tomography*	217
P2.5	13:40 – 15:00	Michael S. Ibrahim, Emil Y. Sidky, Fatma Terzioglu, John Paul Phillips, Ingrid S. Reiser, Guillaume Bal, Xiaochuan Pan	*Variance-bias analysis of dual-energy CT image reconstruction for objects containing metal*	221
P2.6	13:40 – 15:00	Jiadan Song, Shaozhou Pu, Liang Li	*Three mode CT imaging of X-ray transmission, scatter and fluorescence photons*	225
P2.7	13:40 – 15:00	Scott S. Hsieh, Paurakh Rajbhandary	*Properties of an optimized estimator for attenuation at very low count rates*	229
P2.8	13:40 – 15:00	Peiyuan Guo, Longchao Men, Li Zhang, Zhentian Wang	*Dark-field CT reconstruction with a large field of view*	233
P2.9	13:40 – 15:00	David Frank, Tobias Lasser	*curadon: an open-source foundational library for highly efficient differentiable projectors in X-ray computed tomography*	237
P2.10	13:40 – 15:00	Gregory Ongie, Buxin Chen, Zheng Zhang, Dan Xia, Emil Y. Sidky, Xiaochuan Pan	*Simultaneous spectral calibration and image reconstruction in spectral CT with a sum of deltas spectrum model*	241
P2.11	13:40 – 15:00	Paul E. Kinahan, Darrin W. Byrd, Paul Desari, W. Paul Segars, Ehsan Abadi, Katie Olivas, Ehsan Samei	*Impact of patient respiratory motion variability on respiratory-gated PET with CT-based attenuation correction*	245
P2.12	13:40 – 15:00	Lorente Mur, F. C. Del Cerro, M. Desco, M. Abella	*Tomographic reconstruction for chest tomosynthesis based on deep learning*	248

P2.13	13:40 – 15:00	Mikhail Mikerov, Sjoerd A.M. Tunissen, Koen Michielsen, Ewoud J. Smit, Ioannis Sechopoulos	*Adaptation of a 4D noise filter for implementation on GPU*	252
P2.14	13:40 – 15:00	Yanwei Qin, Xin Lu, Hongyang Zhu, Yunsong Zhao, Xing Zhao	*An iterative 3D ring artifacts correction algorithm for CBCT*	256
P2.15	13:40 – 15:00	Xiang Zou, Yuxiang Xing	*A CL scanning with a single spiral trajectory*	260
P2.16	13:40 – 15:00	Ankit Kumar, Johann Kastner, Christoph Heinzl	*Using a patch wise deep U-Net CNN for the segmentation of pores in carbon fibe reinforced polymers*	264
P2.17	13:40 – 15:00	Olivia F. Sandvold, Yinglin Ge, Roland Proksa, Peter B. Noël	*Double bowtie design for high sensitivity pediatric spectral CT*	268
P2.18	13:40 – 15:00	Boris Vervenne, Jens Maebe, Stefaan Vandenberghe	*Feasibility study of a rectangular CT geometry for high throughput total-body PET-CT*	272
P2.19	13:40 – 15:00	Zhilei Wang, Hewei Gao	*Spectral cone-beam artifact reduction for dual-energy cone-beam CT*	276
P2.20	13:40 – 15:00	Faguo Yang, Andriy Andreyev, Matthew Andrew	*Reconstruction methods for scans of X-ray micro tomography with a rotation stage offset*	280
P2.21	13:40 – 15:00	Jannik Dettmer, George S.K. Fung, W. Paul Segars, Karl Stierstorfer	*Faster and more realistic CT projection data generation using digital phantoms and CT simulation software*	284
P2.22	13:40 – 15:00	David Leibold, Jacco van Leenen, Dennis R. Schaart, Marlies C. Goorden	*Initial results for joint reconstruction/decomposition of spectral cone-beam computed tomography data*	288
P2.23	13:40 – 15:00	Y. Boursier, F. Cannet, M. Dupont, S. Anthoine, C. Morel	*Spectral CT reconstruction with ProMeSCT: results on real data from PIXSCAN*	292
P2.24	13:40 – 15:00	Ville-Veikko Wettenhovi, Ari Hietanen, Ville Kolehmainen, Kati Niinimaki, Marko Vauhkonen	*Accelerated iterative reconstruction for cone beam CT using measurement-domain filtering*	296

Oral Session: Deep and Industrial
Time : Wednesday, 15:20 – 17:00
Chairs : Nicole Maaß and J. Webster Stayman

Time	Author	Title	Page
15:20 – 15:40	Aniket Pramanik, Singanallur V. Venkatakrishnan, Obaidullah Rahman, Amirkoushyar Ziabari	*A learnt half-quadratic splitting-based algorithm for fast and high-quality industrial cone-beam CT reconstruction*	300
15:40 – 16:00	Nico Heinz, Andreas Maier, Christopher Syben, Richard Schielein	*Joining photogrammetry and X-ray computed tomography*	304
16:00 – 16:20	Yuzhong Zhou, Linda-Sophie Schneider, Yipeng Sun, Fuxin Fan, Andreas Maier	*A 2D sinogram-based approach to defect localization in computed tomography*	308
16:20 – 16:40	Chengze Ye, Linda-Sophie Schneider, Yipeng Sun, Andreas Maier	*Deep learning computed tomography based on the Defrise and Clack Algorithm*	312
16:40 – 17:00	Alejandro Lopez-Montes, T. McSkimming, W. Zbijewski, H. Huang, A. Skeats, C. Delnooz, B. Gonzales, A. Sisniega	*Latent diffusion reconstruction for stationary CT*	316

Thursday, August 8

Oral Session: AI and Physics
Time : Thursday, 08:00 – 09:40
Chairs : Guang-Hong Chen and Frédéric Noo

Time	Author	Title	Page
08:00 – 08:40	J. Webster Stayman	*Basic short course: combining physics and deep learning*	
08:40 – 09:00	Sen Wang, Maria J. Medrano, Jennie Cao, Wonkyeong Lee, Grant M. Stevens, Justin R. Tse, Adam S. Wang	*Learned tube current modulation from scout images for individualized organ-level CT dose and image quality*	320
09:00 – 09:20	Xiao Jiang, Grace J. Gang, J. Webster Stayman	*CT material decomposition using spectral diffusion posterior sampling*	324
09:20 – 09:40	Arjun Krishna, Ge Wang, Klaus Mueller	*Multi-conditioned denoising diffusion probabilistic model (mDDPM) for medical image synthesis*	328

Coffee Break

Oral Session: Artifacts
Time : Thursday, 10:00 – 11:40
Chairs : Grace Gang and Emil Sidky

Time	Author	Title	Page
10:00 – 10:20	Nele Blum, Thorsten M. Buzug, Maik Stille	*Integration of prior knowledge in application of neural networks for metal artifact reduction in CT*	332
10:20 – 10:40	Frédéric Noo, Z. Guo, K. Stierstorfer, G. S. K. Fung	*Beam hardening correction in diagnostic CT using a convex combination of water and bone*	336
10:40 – 11:00	Leening P. Liu, Kevin M. Brown, Amy E. Perkins, Michael C. Soulen, Peter B. Noël	*Quantitative metal artifact reduction algorithm for spectral CT thermometry*	340
11:00 – 11:20	Shiras Abdurahman, Robert Frysch, Daniel Punzet, Michael Manhart, Hannes Lucking, Stefan Lang, Arnd Dörfler, Georg Rose	*Evaluation of data consistency condition-based beam hardening and scatter corrections for clinical C-arm CBCT*	344
11:20 – 11:40	Yingxian Xia, Zhiqiang chen, Li Zhang, Yuxiang Xing, Hewei Gao	*ComptonNet: a spare-detector-guided deep learning framework for scatter estimation in multi-source dual-ring stationary CT*	348

Lunch Break

Poster Session: Poster Session 3
Time : Thursday, 13:20 – 15:00

	Time	Author	Title	Page
	13:20 – 13:40	**Poster Fast Forward**: 30 s - 60 s presentation for each poster		
P3.1	13:40 – 15:00	Shengzi Zhao, Le Shen, Yuxiang Xing	*Experimental study on charge sharing correction: hardware or software?*	352
P3.2	13:40 – 15:00	Leening P. Liu, Martin V. Rybertt, Pouyan Pasyar, Nadav Shapira, Harold I. Litt, Peter B. Noël	*Spectral quantification in different lumen diameters for cardiovascular applications*	356
P3.3	13:40 – 15:00	Christopher Wiedeman, Chuang Niu, Mengzhou Li, Bruno De Man, Jonathan S Maltz, Ge Wang	*Photon-counting CT using a conditional diffusion model for deblurring and Texture-preservation*	360
P3.4	13:40 – 15:00	Bingxi He, Yongbo Wang, Wei Mu, Shuo Wang, Mengjie Fang, Di Dong, Jianhua Ma, Jie Tian	*Novel medical imaging diagnostics: unleashing artificial intelligence potential through sinogram-driven models*	364
P3.5	13:40 – 15:00	Mingye Wu, Jiayong Zhang, Paul Fitzgerald, James Jobin, Nandan Reddy, Anupama Debnath, Eri Haneda, Ying Fan, Jed Pack, Peter Lorraine, Michael Mahony, Bruno De Man	*An overview of the open-source CT simulation environment XCIST*	368
P3.6	13:40 – 15:00	Giavanna Jadick, Patrick La Riviere	*Accuracy of propagation-based phase-contrast CT under the projection approximation*	372
P3.7	13:40 – 15:00	Matthieu Laurendeau, Frederic Jolivet, Laurent Desbat, Guillaume Bernard, Sebastien Gorges, Simon Rit	*Image reconstruction with local directional total variation regularization using tomographic incompleteness maps*	376
P3.8	13:40 – 15:00	Markus Michen, Rajib Kumar Chanda	*Deep learning based scatter estimation using PYRO-NN: application to real-world objects*	380
P3.9	13:40 – 15:00	J. Guo, G. Khurana, L. Severance, Bahetihazi Maidu, A. Gonzalo, J.C. del Alamo, F. Contijoch	*CT gantry rotation time and reconstructed frame rate impact the accuracy of flow field estimation using a physics-informed neural network*	384
P3.10	13:40 – 15:00	Lauren M. Severance, Andrew M. Kahn, Juan C. del Alamo, Elliot R. McVeigh	*Estimation of blood flow parameters in the left atrial appendage from 4DCT dynamic contrast enhancement*	388
P3.11	13:40 – 15:00	Muge Du, Yinong Liu, Yuxiang Xing	*INFACT: generalizable INR with geometry-aware feature aggregation for CT reconstruction*	392
P3.12	13:40 – 15:00	Martina Nassi, Mikhail Mikerov, Alessandra Tomal, Koen Michielsen, Ioannis Sechopoulos	*Efficient simulation of biological tissue variability using Midgley's decomposition*	396

P3.13	13:40 – 15:00	Zhicong Yu, Georgios Prekas, Daniel Gagnon	*Principal component analysis for assessing spectral performance of photon-counting detectors – spectral distance index*	400
P3.14	13:40 – 15:00	Maria Magnusson, Mikael Sandborg, Asa Carlsson Tedgren, Alexandr Malusek	*The accuracy of CT numbers for DIRA with indata from a real dual-source DECT-scanner*	404
P3.15	13:40 – 15:00	Matthew Tivnan, Dufan Wu, Quanzheng Li	*Pseudo inverse diffusion models for generative CT image reconstruction from low dose data*	408
P3.16	13:40 – 15:00	Guyue Hu, Katharina Niepel, Franka Risch, Christopher Kurz, Matthias Wurl, Thomas Kröncke, Florian Schwarz, Guillaume Landry, Katia Parodi	*Assessment of proton relative stopping power from photon-counting CT data*	412
P3.17	13:40 – 15:00	Nadja Gruber, Johannes Schwab, Elke Gizewski, Markus Haltmeier	*Sparse2Inverse: self-supervised inversion of sparse-view CT data*	416
P3.18	13:40 – 15:00	Maxin Chen, Jonathan Bortfeldt, Daniel Berthe, George Dedes, Franz Pfeiffer, Katia Parodi	*Development of a photon-counting micro-CBCT for a small animal proton irradiation platform*	420
P3.19	13:40 – 15:00	Gregory Ongie	*Continuous domain reconstruction in CT imaging with coordinate-based neural networks*	423
P3.20	13:40 – 15:00	Kai Mei, Shobhit Sharma, Leonid Roshkovan, Steven Ross, Thomas Labno, Zhou Yu, Richard Thompson, Peter B. Noël	*Evaluation of photon-counting detector CT on 3D-printed lung phantom with digital lesions*	427
P3.21	13:40 – 15:00	Gurnoor Singh Khurana, Jinyuxuan Guo, Lauren Severance, Alejandro Gonzalo, Bahetihazi Maidu, Juan C. del Alamo, Francisco Contijoch	*SinoFlow: fluid flow estimation from sinograms using a physics informed neural network*	431
P3.22	13:40 – 15:00	Irene Hernández-Girón	*3D printed anthropomorphic phantoms for CT imaging – Are their attenuation properties consistent for different image reconstructions and CT manufacturers?*	435
P3.23	13:40 – 15:00	Elsa B. Pimenta, Luuk J. Oostveen, Paulo R. Costa	*Evaluation of the detectability index of CT thorax AI-based reconstruction protocols*	439
P3.24	13:40 – 15:00	Xiangyang Tang, Yi Yang	*Dark-field x-ray imaging of anisotropic subdel microstructures: rod-shaped air-bubbles*	443

Oral Session : Novel CT
Time : Thursday, 15:20 – 17:00
Chairs : Bruno De Man and Adam Wang

Time	Author	Title	Page
15:20 – 15:40	Grace J. Gang, Yiqun Q. Ma, Leening Liu, Peter B. Noël, J. Webster Stayman	*Multiple focal spots for high-resolution photon-counting CT*	447
15:40 – 16:00	Olivia F. Sandvold, Roland Proksa, Heiner Daerr, Kevin M. Brown, Thomas Koehler, Amy E. Perkins, Grace J. Gang, J. Webster Stayman, Ravindra M. Manjeshwar, Peter B. Noël	*Hybrid spectral CT: combining rapid kVp-switching and dual-layer detectors for high sensitivity iodine imaging*	451
16:00 – 16:20	Jingyan Xu, Frédéric Noo	*Multi-energy data acquisition using a stationary spatial spectral encoder (S3E)*	455
16:20 – 16:40	Scott S. Hsieh	*Three concepts to improve spatial resolution in next-generation CT scanners*	459
16:40 – 17:00	Linxi Shi, N. Robert Bennett, Adam Wang	*Single-scan quantitative cone-beam CT using a primary modulator and dual-layer flat-panel detector: a phantom study*	463

Friday, August 9

Oral Session: Image Quality I
Time : Friday, 08:00 – 09:40
Chairs : Karl Stierstorfer and Srinivasan Vedantham

Time	Author	Title	Page
08:00 – 08:40	Frédéric Noo	*Basic short course: quantification of image quality*	
08:40 – 09:00	Mathias Anton, Marcel Reginatto, Ulf Mader	*Nonparametric statistics for the quantification of noise and contrast in CT images*	467
09:00 – 09:20	Xi Chen, Tzu-Cheng Lee, Liang Cai	*Accelerating diffusion-based CT image restoration through sampling step sharing*	471
09:20 – 09:40	Jessica Y. Im, Neghemi Micah, Amy E. Perkins, Michael Geagan, Sven Kabus, Kai Mei, Peter B. Noël	*Lifelike and deformable lung phantoms for 4DCT imaging: a three-dimensional printing approach*	475

Coffee Break

Oral Session: Image Quality II
Time : Friday, 10:00 – 11:40
Chairs : Ioannis Sechopoulos and Scott Hsieh

Time	Author	Title	Page
10:00 – 10:20	Leening P. Liu, Pouyan Pasyar, Olivia F. Sandvold, Pooyan Sahbaee, Harold I. Litt, Peter B. Noël	*Long-term quantitative stability of a first-generation dual-source photon-counting CT*	479
10:20 – 10:40	John Hoffman, Morgan Daly	*Feasibility of an open-standard, accessible, reproducible modular insert phantom for quantitative imaging*	483
10:40 – 11:00	Wonkyeong Lee, Sen Wang, Maria Jose Medrano, Jennie Cao, Grant M. Stevens, Justin Ruey Tse, Adam S. Wang	*Organ-level noise estimation in CT images using deep learning*	487
11:00 – 11:20	Md Selim, Prabhat KC, Rongping Zeng	*Image resolution analysis using line-pair and disk-based MTF methods for deep learning CT image denoising*	491
11:20 – 11:40	Larry Zeng, Frédéric Noo, Marc Kachelrieß	*Preview of Fully 3D 2025, preview of CT Meeting 2026, and meeting adjourn*	

Lunch Break

Visit to EZRT

Time : Friday, 13:20 – 17:00

A guided tour to the Fraunhofer Development Center for X-ray Technology (EZRT) will demonstrate various industrial CT systems, including the smallest and the largest CT scanner of the world. Busses will pick you up at the conference center Friday 13:20 and return at about five. This optional post meeting event is free of charge, but it requires preregistration.

Time

13:20 Bus departs from the Welcome Kongresshotel to the EZRT (close to Nürnberg)

14:00 Guided visit of the EZRT:

- Smallest CT scanner in the world
- Largest CT scanner in the world
- Robo CT
- Several other industrial systems

16:20 Bus departs from the EZRT

17:00 Arrival at the Welcome Kongresshotel in Bamberg

A directional total variation minimization algorithm for isotropic resolution in digital breast tomosynthesis

Emil Y. Sidky, Xiangyi Wu, Xiaoyu Duan, Hailing Huang, Wei Zhao, Leo Y. Zhang, John Paul Phillips, Zheng Zhang, Buxin Chen, Dan Xia, Ingrid S. Reiser, Xiaochuan Pan

***Abstract*—An optimization-based image reconstruction algorithm is developed for contrast enhanced digital breast tomosynthesis (DBT) using dual-energy scanning. The algorithm minimizes directional total variation (TV) with a data discrepancy and non-negativity constraints. Iodinated contrast agent (ICA) imaging is performed by reconstructing images from dual-energy DBT data followed by weighted subtraction. Physical DBT data is acquired with a Siemens Mammomat scanner of a structured breast phantom with ICA inserts. Results are shown for both directional TV minimization and filtered back-projection for reference. It is seen that directional TV is able to substantially reduce depth blur for the ICA objects.**

***Index Terms*—Limited angular-range reconstruction, digital breast tomosynthesis, dual-energy, contrast-agent imaging**

I. Introduction

Over the past two decades, X-ray based tomosynthesis has been an emerging imaging modality that has been making great strides in clinical adoption; most notably, digital breast tomosynthesis (DBT) is becoming the workhorse imaging device for breast cancer screening. Although DBT yields partially tomographic volumes, there is still a large gap between it and X-ray computed tomography in terms of achieving isotropically high resolution and in the ability to achieve quantitive imaging for shape and volume features of tissues within the scanned subject. Tomosynthesis imaging is subject to significant depth blur, and this blurring is proportional to the size of the imaged structures. Because of this depth blurring it is not possible to outline the boundaries of structures and estimate their volume. Moreover, the gray scale values in the image do not directly correspond to a physical density.

In this work, an optimization-based algorithm is developed that may enable quantitive imaging with limited angular range scanning such as what is performed in a tomosynthesis acquisition. In order to accomplish this, we exploit sparsity in the gradient magnitude of the scanned subject. For sparse-view imaging over a full scanning angular range, the use of total variation (TV) minimization has proven effective [1]. For limited angular range scanning, however, directional TV constraints show promise [2]. The use of directional TV constraints can be shown to be equivalent to directional TV minimization where the gradient magnitudes in different directions have different penalty weights [3]. The work presented in Ref. [3] showed only simulation studies, and in this work we apply this algorithm to physical DBT transmission data.

The application of interest is contrast enhanced (CE) DBT [4], where the goal is to obtain a quantitative image of the Iodinated contrast agent (ICA) distribution. If such a system is successfully developed it could enable the use of CE-DBT in assessment of breast cancer therapy effectiveness. For this preliminary work, DBT data is obtained for structured breast phantom with ICA inserts. Volumes are reconstructed from scans at 30 and 49 kV, and the weighted image subtraction method is used to isolate the ICA distribution.

II. Methods

A. Image reconstruction

In order to obtain reconstructed DBT volumes that have nearly isotropic resolution for the ICA distribution, we exploit the fact that this distribution is localized with a fairly simple structure and high contrast. In the algorithm implementation directional gradient sparsity regularization is employed and the resolution is reduced by a factor of eight from the native resolution of the DBT system's detector.

The image reconstruction algorithm is developed for the Siemens Mammomat scanner that acquires 25 projections over a 50 degree scanning arc and the detector resolution is 85 microns. The transmission data are processed with the negative logarithm after dividing by a flood-field scan and the relationship between this processed data and the DBT volume is assumed to be standard linear X-ray projection as described by

$$g = Xf,$$

where f is a discrete representation of the DBT volume, X is a matrix encoding X-ray projection, and g is the processed projection data.

For sparsity regularization, image reconstruction is formulated using an optimization model where weighted image directional total variation (TV) is minimized subject to a data-fidelity constraint. Furthermore, an ℓ_1 penalty term is included that encourages pixel sparsity and helps to confine the image to the true object support. Finally, the data fidelity term compares ramp-filtered data with ramp-filtered estimated data, which has

E. Y. Sidky, J. P. Phillips, Z. Zhang, B. Chen, D. Xia, I. S. Reiser, and X. Pan are with the Department of Radiology MC-2026, The University of Chicago, 5841. S. Maryland Ave., Chicago, IL, 60637.

X. Wu, X. Duan, H. Huang, and W. Zhao are with the Department of Radiology, Renaissance School of Medicine, Stony Brook University, 101 Nicolls Road, Health Sciences Center, Level 4, Stony Brook, NY, 11794.

L. Y. Zhang is with the Department of Mathematics, University of Washington, Box 354350, C-138 Padelford, Seattle, WA 98195-4350.

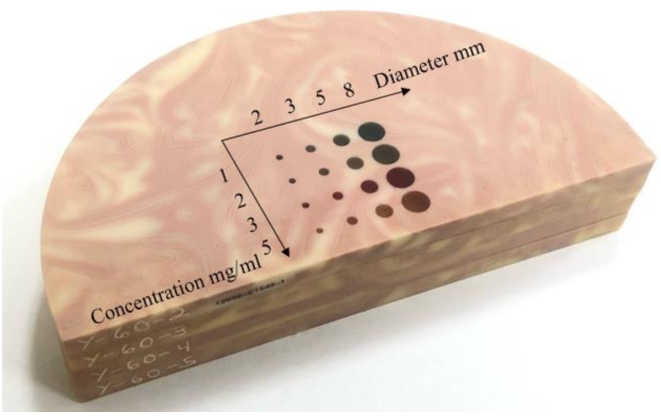


Fig. 1. CIRS phantom with solid ICA inserts. The background is hetergeneous with 50% glandularity.

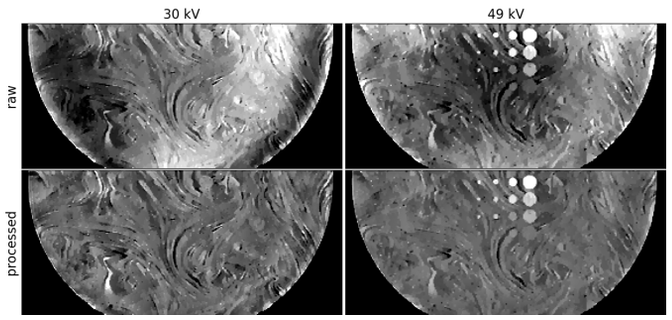


Fig. 2. Raw and processed in-plane images containing all ICA objects for the directional TV minimization images for the 30 kV (Left) and 49 kV (Right) DBT scans. Image processing is performed by polynomial fitting and normalization described in Eqs. (1) and (2), respectively. The gray scale for the raw images is selected to span the range of background variation, and the gray scale for the normalized images is selected to be $\pm$ 30% of background.

two important effects: this filtering deemphasizes low spatial-frequency discrepancy in the sinogram estimation and it also performs preconditioning.

$$f = \arg\min_{f'} \big\{ \alpha \|\nabla_x f'\|_1 + \alpha \|\nabla_y f'\|_1 + (2-\alpha)\|\nabla_z f'\| + \beta \|f'\|_1 \text{ such that } \|R(g - Xf')\|_2 \leq \epsilon \text{ and } f' \geq 0 \big\}.$$

The x and y coordinates are in-plane with y being along the direction of travel for the X-ray source, and the z direction is the depth direction; i.e. perpendicular to the X-ray detector. The symbol ∇ represents finite differencing along the direction indicated by the subscript; α is a weighting parameter that should be chosen between 0 and 2; β is the weighting parameter for the ℓ_1 penalty, R represents ramp filtering and ϵ is a data error tolerance parameter. The posed optimization problem is convex and it can be solved with the primal-dual algorithm of Chambolle and Pock [5].

When $\alpha = 1$, the gradient sparsity regularization term is equivalent to TV and it has been shown that constrained TV minimization is effective for accurate image reconstruction for sparse-view acquisition when the scanning arc is complete. Recently, we have shown that the anisotropic weighting of the directional TV terms can improve image reconstruction accuracy for limited scanning angular ranges such as those used for DBT. This in combination with use of a coarse volume grid allows for nearly isotropic resolution in the reconstructed DBT images using directional TV minimization.

B. Image post-processing

Even when the transmission data preprocessing accurately estimates the image sinogram, Reconstructed DBT images are susceptible to low spatial-frequency artifact due to the limited angular range of the scan. In order to facilitate the dual-energy image subtraction the low spatial-frequency artifacts are estimated and removed by normalization. Specifically, the low spatial-frequency image component is estimated by robust polynomial fitting

$$b = \arg\min_{c} \big\| m \cdot (s - Pc) \big\|_1, \tag{1}$$

where s is an in-plane slice of the image f; P is a matrix of low-order polynomials; c is a vector of coefficients; m is an indicator function which is one on the object support and zero otherwise, where the object support is estimated by thresholding the image slice, s; and b is the estimated background for the slice s. The use of the ℓ_1-norm for the objective function encourages the difference image in its argument to be sparse, and this design should preserve important small structures in the image when the background drift is processed away.

The polynomial fitting is performed slice by slice with the result stacked into the background image f_b. Likewise, the mask image f_m is assembled by stacking the slice masks. The algorithm for solving this robust polynomial fitting is iteratively reweighted least-squares. The final displayed image, f_d, is computed as follows

$$f_d = f_m \cdot \frac{f}{f_b}. \tag{2}$$

After the processed images are computed for low and high kV, a difference image is computed

$$f^{(ICA)} = \left(f_d^{(HE)} - 1 \right) - w \cdot \left(f_d^{(LE)} - 1 \right).$$

For this work, the dual energy data is processed in the image domain by performing a weighted subtraction of low and high kV reconstructed volumes, where the weights are designed to highlight the ICA distribution. Because the low spatial-frequency background variation is reduced by normalization, the background of the processed images center on the value of 1; thus 1 is subtracted from both images before performing the weighted subtraction. The weighting parameter w is selected by visual inspection.

C. DE-DBT transmission data acquisition

The DE-DBT scanning consists of two DBT acquisitions taken at source potential of 30 and 49 kV using filtration with Al and Ti, respectively. DE-DBT scans of a 4 cm thick BR3D phantom (Model 020, CIRS Inc. Norfolk, VA) with solid ICA inserts, see Fig. 1, are used for testing the proposed image reconstruction algorithm. The ICA inserts vary in size and

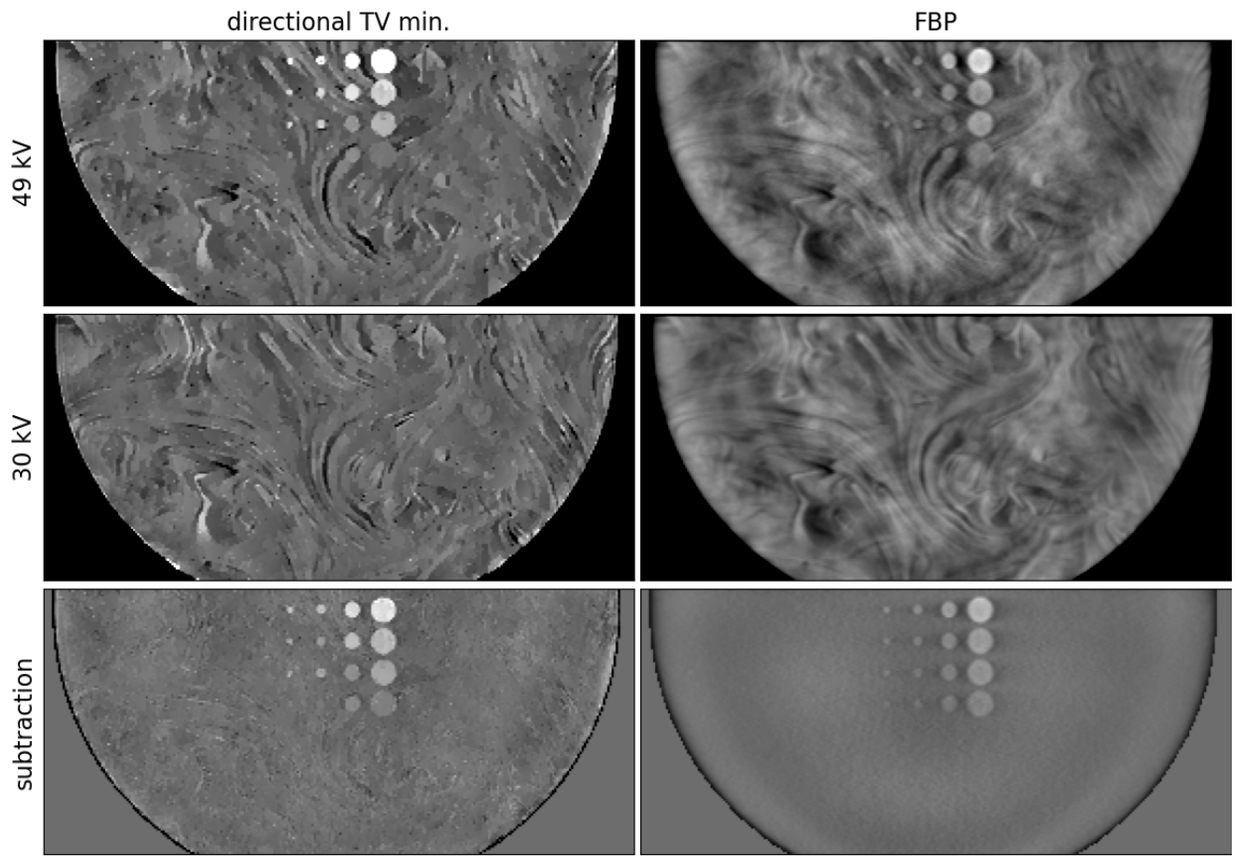


Fig. 3. In-plane images containing all ICA objects for directional TV minimization (Left) and FBP (Right). The top and middle rows show the images corresponding to the 49 and 30 kV acquisitions, respectively. The bottom row shows the weighted subtraction designed to isolate the ICA distributions. The gray scales on the left and right are matched and the images are masked to the object support.

concentration; the diameters are 2, 3, 5, and 8 mm and the Iodine concentrations are 1, 2, 3, and 5 mg/mL.

III. Results

For the present implementation of directional TV minimization, the weighting parameter α is set to 1.75, favoring the in-plane differentiation in the objective function. The ℓ_1 penalty parameter β is set to 0.1, and the data error constraint ϵ is 0.01 in terms of root mean square error. The imaging volume is discretized on a coarse grid of size $188 \times 417 \times 175$ with cubic voxels 680 microns wide.

Raw slice images from applying directional TV minimization to the DBT phantom data are shown in the top panels of Fig. 2, and the low spatial-frequency artifacts are clearly visible. The slice shown cuts through all of the ICA objects. These objects are easily seen in the 49 kV images, and not the 30kV images, because the K-edge of Iodine is 33.2 keV. The spatial dependence of the artifact is different for the low and high kV images, and therefore direct image subtraction of the raw images will not yield good isolation of the ICA distribution. For performing the post-processing, we employed an eighth degree 2D polynomial expansion (45 terms), and resulting images are shown in the bottom row of Fig. 2, where it is seen that the low spatial-frequency artifacts are greatly reduced.

The results of weighted subtraction for directional TV minimization filtered back-projection (FBP) are shown in Fig. 3, where it is seen in the bottom row that it effectively suppresses the background variation and isolates the ICA objects in the image.

Transverse slices, where y is horizontal and z is vertical, for the weighted subtraction of directional TV minimization and FBP images are displayed in Fig. 4. While the FBP images show the usual depth blurring of the ICA objects, it is clear that directional TV minimization can effectively mitigate this artifact. On the other hand, the directional TV difference images have slightly stronger background clutter due to imperfect cancellation of the background glandular structures, which also interferes with the imaging of the 1mg/mL ICA inserts. This imperfect cancellation is likely due to the non-linearity of the sparsity regularization.

IV. Conclusion

In this preliminary study on image reconstruction by directional TV minimization for CE-DBT, the physical phantom results show promise in achieving an accurate ICA distribution estimate with nearly isotropic spatial resolution; it appears that the depth blurring artifact common to DBT imaging is well controlled with directional TV minimization. Because the dual-energy processing, in this work, is image-based the gray values in the ICA images are not quantitive and cannot

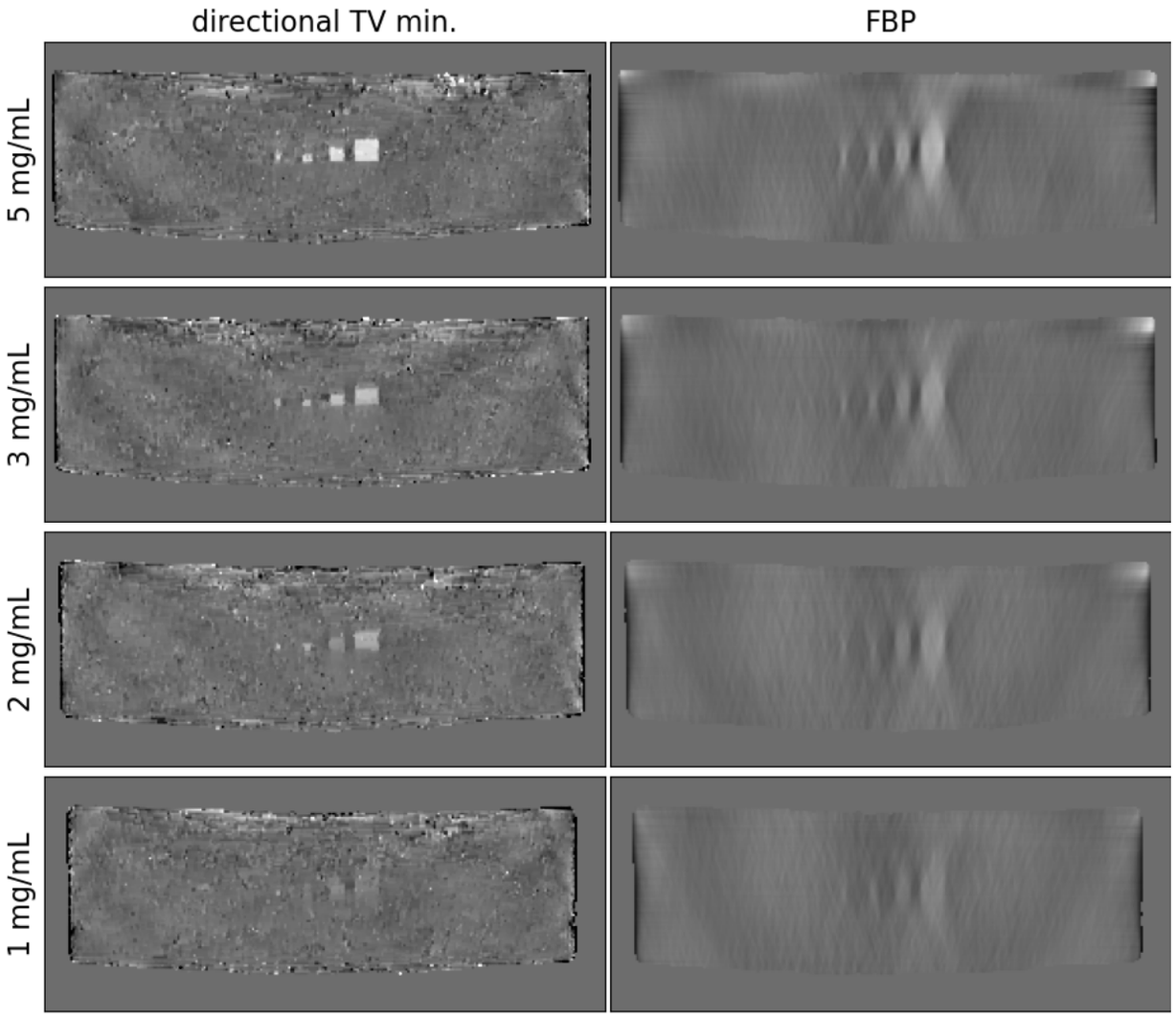


Fig. 4. Transverse plane images cutting through the groups of ICA objects of the same concentration with directional TV minimization and FBP images on the left and right, respectively. Concentration of the ICA decreases going from top to bottom. All gray scales are matched for quantitative comparison and the images are masked to the object support.

be directly tied to a concentration value. Future work will focus on spectral calibration of the DBT transmission data so that both one-step and two-step [6] quantitative DE-DBT image reconstruction can be developed. Adapting directional TV minimization to a quantitive DE-DBT is expected to greatly reduce background clutter in the ICA images from glandular tissue, and this should also provide accurate ICA concentrations.

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 R01EB023968 and R21CA263660. The CEDBT system and clinical research was supported financially by Siemens Medical Solutions, USA.

References

[1] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4808, 2008.

[2] 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*, vol. 70, p. 102030, 2021.

[3] L. Y. Zhang, E. Y. Sidky, J. P. Phillips, Z. Zhang, B. Chen, D. Xia, and X. Pan, "Limited scanning arc image reconstruction with weighted anisotropic TV, minimization," in *17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2023, pp. 491–494.

[4] H. Huang, D. A. Scaduto, C. Liu, J. Yang, C. Zhu, K. Rinaldi, J. Eisenberg, J. Liu, M. Hoernig, J. Wicklein, S. Vogt, T. Mertelmeier, P. R. Fisher, and W. Zhao, "Comparison of contrast-enhanced digital mammography and contrast-enhanced digital breast tomosynthesis for lesion assessment," *J. Med. Imaging*, vol. 6, p. 031407, 2019.

[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. 120–145, 2011.

[6] B. Chen, Z. Zhang, D. Xia, E. Y. Sidky, T. Gilat-Schmidt, and X. Pan, "Accurate image reconstruction in dual-energy CT with limited-angular-range data using a two-step method," *Bioengineering*, vol. 9, no. 12, p. 775, 2022.

Application of Gated Recurrent Units for CT Trajectory Optimization

Yuedong Yuan, Linda-Sophie Schneider, Andreas Maier, *Member, IEEE*

***Abstract*—Recent advances in computed tomography (CT) imaging, especially with dual-robot systems, have introduced new challenges for scan trajectory optimization. This paper presents a novel approach using Gated Recurrent Units (GRUs) to optimize CT scan trajectories. Our approach exploits the flexibility of robotic CT systems to select projections that enhance image quality by improving resolution and contrast while reducing scan time. We focus on cone-beam CT and employ several projection-based metrics, including absorption, pixel intensities, contrast-to-noise ratio, and data completeness. The GRU network aims to minimize data redundancy and maximize completeness with a limited number of projections. We validate our method using simulated data of a test specimen, focusing on a specific voxel of interest. The results show that the GRU-optimized scan trajectories can outperform traditional circular CT trajectories in terms of image quality metrics. For the used specimen, SSIM improves from 0.38 to 0.49 and CNR increases from 6.97 to 9.08. This finding suggests that the application of GRU in CT scan trajectory optimization can lead to more efficient, cost-effective, and high-quality imaging solutions.**

***Index Terms*—Robotic CT, scan trajectory optimization, Tuy condition, data completeness, GRU.**

I. Introduction

COMPUTED Tomography (CT) has transformed non-destructive testing by providing unprecedented insight into the structure of objects. Recent advances in CT technology have introduced dual-robot CT systems, where individual robots control the X-ray source and detector. This innovation provides flexibility by allowing projections from arbitrary views around the object of interest. Robotic CT systems offer the potential for more customized and complex CT trajectories, facilitating scan trajectory optimization [1]. Optimized scan trajectories reduce scan time and radiation exposure while maintaining or improving image quality. This increases scanner efficiency, reduces operating costs, and increases throughput. Furthermore, reconstruction through carefully selected projections can improve image resolution and contrast, which are critical for detecting finer details and subtle anomalies in scanned objects [2].

This paper presents a new approach to CT scan trajectory optimization using Gated Recurrent Units (GRUs), a machine learning technique, based on several projection-based metrics, including absorption, pixel intensities, contrast-to-noise ratio (CNR) [3], and data completeness [4]. Our methodology evaluates and selects projections for data reconstruction using these metrics. This study uses cone-beam CT, focusing on a specific volume of interest (VOI) within the scanned object. We present GRU as a powerful tool for optimizing CT scan trajectories. The approach aims to improve the accuracy and efficiency of CT scanning processes, paving the way for more cost-effective and higher-quality imaging. Our results show that the GRU-optimized scan trajectories are superior to traditional circular CT trajectories in terms of image quality, indicating potential advances in CT imaging techniques.

Y. Yuan, L. Schneider and A. Maier are with the Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany (e-mail:yuedong.yuan@fau.de)

II. Methods

A. Absorption and Pixel Intensity

In X-ray imaging, beam hardening and scattering have a significant impact on image quality, often resulting in artefacts in CT images [5]. These problems arise from the energy-dependent attenuation of X-rays, which is influenced by the composition of the material and the different depths at which they penetrate. The phenomenon of beam hardening occurs when low energy X-rays are absorbed more strongly than their high energy counterparts as they pass through an object [6]. This differential absorption changes the energy profile of the beam. To quantify this effect, we use a metric known as pixel intensity, representing the residual intensity of X-rays captured by the detector. This residual intensity of X-rays is reflected in the pixel values of the projection images. The value of 70-quantiles of pixel values within the region of interest (ROI) is calculated to evaluate this metric. A higher pixel intensity value indicates lower attenuation, which is desirable for optimizing CT scan trajectories. To mitigate the effects of beam hardening, projections with a pixel value within the ROI below a certain threshold α are excluded from the analysis. It is important to note that this threshold varies depending on the specific object being scanned and it is an empirical value.

B. Contrast-to-Noise Ratio

CNR, a core image quality metric [3], helps evaluate the detectability of defects in projection images [7]. The CNR is defined by the equation below.

$$CNR = \frac{|Max - Min|}{\sigma_n} \tag{1}$$

where Max and Min are the highest and lowest pixel values within the ROI, respectively, and σ_n is the standard deviation of the background pixels.

A higher CNR means better contrast differentiation within the ROI, indicating better image quality and projection clarity.

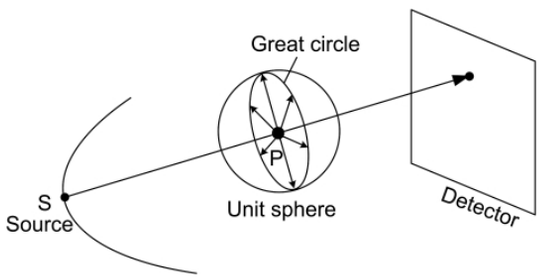


Fig. 1. A great circle formed by normal vectors of all planes containing the X-ray path SP [8].

C. Data Completeness

The data completeness metric, rooted in Tuy's Data Sufficiency Condition, evaluates the scan trajectory of CT systems for accurate volumetric reconstruction of the VOI within scanned objects. Tuy's Data Sufficiency Condition necessitates that each cross-sectional plane of the scanned object intersects the scan trajectory at least once in order to reconstruct the scanned object [9]. A single circular trajectory, for instance, fails to satisfy this condition, since there is no intersection between the single circular trajectory and the cross-sectional planes parallel to this trajectory plane or slightly tilted to this trajectory plane.

Figure 1 illustrates the terms great circle and unit sphere. Each sample point on the unit sphere represents a normal vector of the cross-sectional plane passing through a scanned object point within the VOI, known as the Radon plane. In addition, each scan trajectory point corresponds to a great circle. The plane of the great circle is perpendicular to the viewing direction of the X-ray beam from the source location S to the scanned object point P, for which the data completeness is computed.

To achieve a discrete formulation of the data completeness metric, we build on the concepts from [10] and their extensions in [4]. The procedure involves two main steps. First, we sample M vectors u from the unit sphere as uniformly as possible. Note that only half of the sphere needs to be sampled due to the symmetry of the plane integrals in the upper and lower halves. Second, we assess the unit sphere coverage for each voxel in the VOI. This involves identifying unit sphere vectors u that align with the normal vectors of the Radon planes. Only viewing vectors d that hit the detector are considered.

The Nyquist-Shannon sampling theorem informs the requirement that the angular distance between adjacent planes in Radon space must be less than a maximum gap $\Delta\gamma$. Given that the inner product of the two unit vectors represents the cosine of their angular separation, vectors of interest meet the criterion below.

$$|d^T u| < \sin(\Delta\gamma) \tag{2}$$

This equation is evaluated for each unit vector u against all viewing directions through the VOI, leveraging the constant nature of $\Delta\gamma$ for precomputed sine function values. The complete summary of the steps for calculating the data completeness metric is described in Algorithm 1.

Algorithm 1 Calculation of data completeness matrix for $\mathbf{P}$

```
Source positions S ∈ R^{3×N}
Unit vectors 𝕊_P ∈ R^{3×M}
Initialize data completeness matrix C ∈ 0,1^{N×M}
for s_i ∈ S do
  Calculate viewing direction d := (P − s_i) / ‖P − s_i‖_2
  for u_j ∈ 𝕊_P do
    if |d^T u| < sin(Δγ) then
      C_{i,j} ← 1
    else
      C_{i,j} ← 0
    end if
  end for
end for
```

The data completeness matrix C can be calculated before CT trajectory optimization. This matrix is boolean, with an entry $C_{i,j} = 1$ indicating that the viewing direction associated with the source point s_i falls within the $\Delta\gamma$ threshold for being perpendicular to the unit vector u_j.

As depicted in Figure 1, the more X-ray source locations are employed as CT trajectory points, the more unit sphere points could be hit by one or more great circles. For a subset $S_{opt} \subseteq S$, the coverage c of the unit sphere and thus the region-based data completeness of the optimized CT trajectory S_{opt} can be calculated by counting the columns having at least one non-zero elements for the sampled unit sphere $\mathbb{S}_P$.

D. Gated Recurrent Unit for CT Trajectory Optimization

In the context of data completeness, redundant data occurs when the same point on the unit sphere is covered multiple times by different projections. CT trajectory optimization aims to minimize this redundancy and maximize the coverage of the unit sphere using a given number of projections. This task essentially translates into identifying an optimized combination of projections.

During each iteration of the optimization process, the goal is to select a projection that contributes the most to the unit sphere $\mathbb{S}_P$, considering the influence of previously selected projections. This makes the problem sequential, for which recurrent neural networks (RNNs) are appropriate. RNNs are able to learn from sequential data due to their internal memory and can predict subsequent sequences, making them well suited for this application [11].

A GRU is a recurrent neural network that has become popular for its computational efficiency and effectiveness in sequence-based tasks. As a streamlined alternative to the Long Short-Term Memory (LSTM) network, GRUs were designed to mitigate some of the computational challenges of LSTMs [12]. The distinctive feature of GRUs is their two-gate structure: the update gate and the reset gate, as opposed to the three gates of LSTMs. Figure 2 illustrates a single GRU unit.

The update gate in GRUs controls the flow of information from previous steps to the current step, ensuring that only relevant data is passed on. Meanwhile, the reset gate decides how much past information is forgotten and how much new

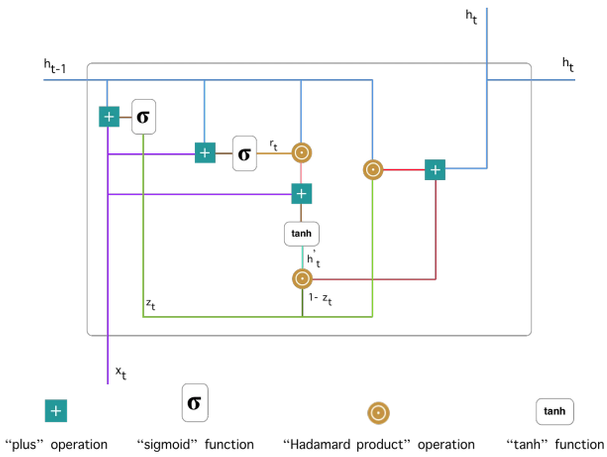


Fig. 2. Scheme of a single GRU unit. x_t denotes the input data at time step t, r_t represents the reset gate at time step t, and z_t corresponds to the update gate at time step t. h_{t-1}/h_t indicates the hidden state at time step $t-1/t$, while h'_t represents the current memory of the unit at time step t [13].

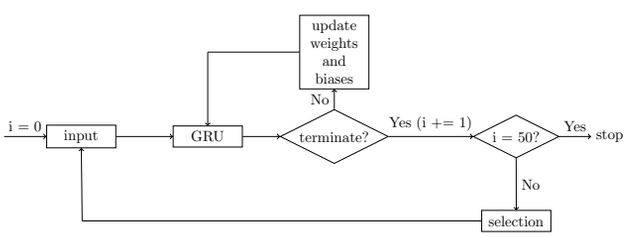


Fig. 3. Diagram of the selection process for the next projection through the GRU network

information is incorporated into the current step. This structure allows GRUs to model long-term dependencies in sequences efficiently. In addition, GRUs have no separate cell state, a simplification that often makes them easier to train, especially with smaller datasets. Given these advantages, GRUs are used in this research to optimize CT scan trajectories. Note that the multilayer GRU network is a sequence of copies of the same GRU unit, and the same units are used for all time steps.

In CT trajectory optimization, pre-computed metrics serve as input to the GRU network. For each projection, these metrics include pixel intensity, CNR, and M numbers for data completeness. Before providing the array of metrics to the network, specific projections are excluded, where the smallest pixel value in the ROI falls below the α threshold.

Figure 3 illustrates the iteration and projection selection process within the GRU network. During each iteration, the weights and biases of the GRU network are updated for a predefined number of training loops. A termination condition is then evaluated, and if it is met, the network selects a projection for that iteration. If the condition is not met, the network continues to be updated. Note that our network does not have a separate validation stage used commonly in traditional neural networks and the GRU network outputs a probability prediction for each batch.

The loss function for the GRU network is defined as the binary cross entropy between the predicted probabilities and the scores of each batch. These scores are calculated using a

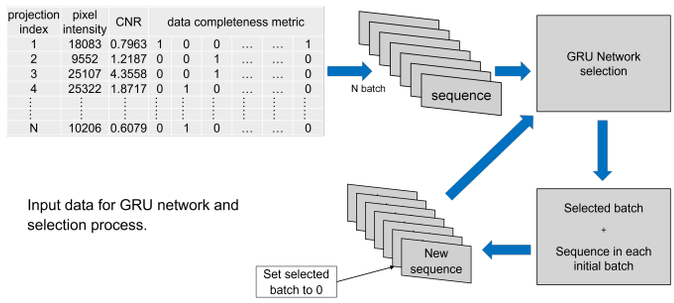
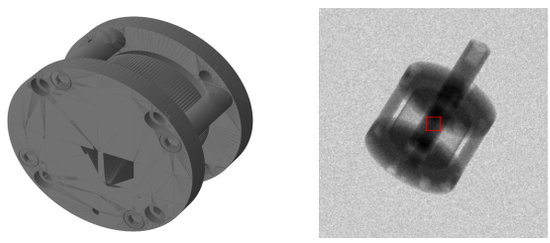


Fig. 4. Illustration of the data processing in GRU network.



(a) (b)

Fig. 5. Test specimen overview. (a) Carbon specimen illustration. (b) Simulated projection highlighting the VOI in red.

weighted average of three normalized metrics: pixel intensity, CNR, and data completeness. Normalizing each metric to a scale of 0 to 1 mitigates the effects of their different value ranges. To emphasize data completeness, a weight factor of 16 is assigned. During the first iteration, the sequence in each batch contains a single element, i.e. a row of the input array. Thus, each one-element sequence contains only one projection. The training process stops when the loss in the current loop exceeds the loss in the previous loop, indicating that the loss value no longer decreases.

After each selection, the data of the selected batch is added to each initial batch. This combination, including the chosen batch and the metrics of an available projection, forms the new input data for the next iteration. As the process is repeated, the sequence in each batch becomes progressively longer. To prevent the network from selecting the same projection twice, batches with an element already selected are assigned a value of 0. Figure 4 illustrates these data processing steps.

III. Experiments and Results

We validated our CT scan trajectory optimization approach using a test specimen as shown in Figure 5, focusing on the center point as the VOI. For this test specimen, the center is complex to reconstruct completely because the X-rays' attenuation of some view directions is significantly strong due to the long transmission length.

We simulated 1000 spherical trajectory projections, filtering out those with high attenuation, reducing the number to 657. The simulation used a polychromatic spectrum corresponding to a tube voltage of $150\,\mathrm{kV}$. A square detector of size $11.52\,\mathrm{cm}$ and isotropic pixel spacing of $450\,\mu\mathrm{m}$ is modeled. The chosen setup leads to a magnification of 2. Noise is simulated, but scatter effects are excluded from the simulation.

We sampled a unit sphere around our VOI, resulting in $|\mathbb{S}_P| = 1000$ unit sphere points and set a maximum angular distance of $\Delta\gamma = 0.573°$.

Our GRU network is trained on a Quadro RTX 8000 GPU and the AdamW algorithm in PyTorch is used to update the weights. Using the Weights & Biases platform, we have explored different hyperparameters for the GRU network to optimize the selection for a set of 50 projections. The final hyperparameters are presented in table I.

TABLE I
OVERVIEW OF OPTIMIZED HYPERPARAMETERS FOR GRU

Parameter	hidden size	number of layers	training loop	learning rate
value	1075	6	34	0.002677

We reconstructed volume data from these 50 projections and compared them to reconstructions from a 50-projection circular trajectory and a 1000-projection spherical reference trajectory (Figure 6). The quantitative evaluation included the calculation of the Structural Similarity Index (SSIM), the Contrast-to-Noise Ratio (CNR) and the Peak Signal to Noise Ratio (PSNR). The GRU-optimized trajectory showed fewer artefacts and better image quality, as demonstrated by higher SSIM, PSNR, and CNR metrics compared to the circular trajectory (table II).

TABLE II
COMPARISON OF QUALITY METRICS FOR 50 PROJECTIONS

Approach	**SSIM↑**	**PSNR↑**	**CNR↑**	**Coverage**
Circular	0.38	120.0db	6.97	45.6%
GRU	0.49	121.0db	9.08	60.2%

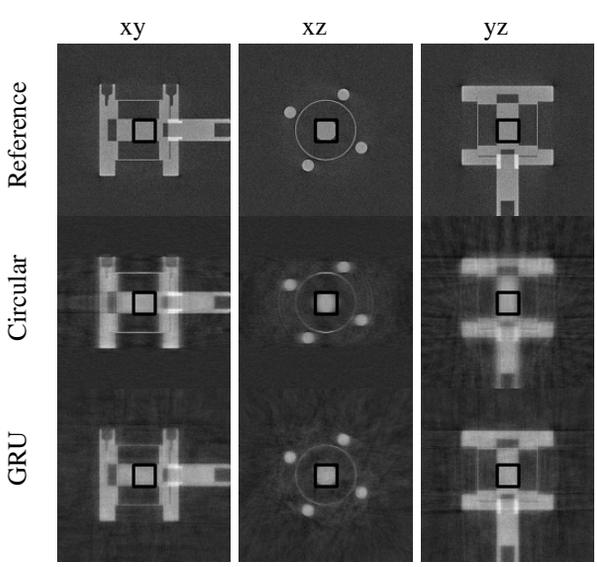


Fig. 6. Optimized CT trajectory for the center voxel. The optimized and circular CT trajectories consist of $k = 50$ projections, while the reference data was a spherical CT trajectory with 1000 projections. All these images are reconstructed using the Algebraic Reconstruction Technique. The black square in each image represents the selected VOI.

IV. CONCLUSION AND OUTLOOK

This study explores the use of GRU networks to optimize CT scan trajectories based on projection metrics. We first compute metrics for each projection of a test specimen, which are then fed into a GRU network to determine an optimized CT trajectory with 50 projections. The results show that this GRU-optimized trajectory outperforms a standard circular trajectory, highlighting the potential of GRUs in trajectory optimization. However, our results are based on a single specimen and simulated data, indicating the need for more extensive testing and application to real CT scan data.

Future research could also focus on the impact of different weights assigned to the projection-based metrics. Currently, we assign a weight of 16 to data completeness and 1 to other metrics. As indicated by our parameter search, different weights could significantly alter the optimized trajectories, warranting further investigation into metric weighting for improved optimization.

REFERENCES

[1] G. Herl, J. Hiller, and A. Maier, *Scanning trajectory optimization using a quantitative Tuybased local quality estimation for robot-based X-ray computed tomography*, *Nondestructive Testing and Evaluation*, vol. 35, no. 3, pp. 287–303, 2020, Taylor & Francis.

[2] L.-S. Schneider, M. Thies, R. Schielein, C. Syben, M. Unberath, and A. Maier, *Learning-based Trajectory Optimization for a Twin Robotic CT System*, *12th Conference on Industrial Computed Tomography (iCT)*, Vol. 28(3), 2023

[3] N. Desai, A. Singh, and D. J. Valentino, *Practical evaluation of image quality in computed radiographic (CR) imaging systems*, in *Medical Imaging 2010: Physics of Medical Imaging*, vol. 7622, pp. 1498–1507, 2010, SPIE.

[4] G. Herl, A. Maier, and S. Zabler, *X-ray CT data completeness condition for sets of arbitrary projections*, *7th International Conference on Image Formation in X-Ray Computed Tomography*, vol. 12304, pp. 67–73, SPIE, 2022.

[5] E. Pettersson, A. Bäck, and A. Thilander-Klang, *Comparison of metal artefacts for different dual energy CT techniques*, *Radiation Protection Dosimetry*, vol. 195, no. 3-4, pp. 232–245, 2021, Oxford University Press.

[6] J. F. Barrett and N. Keat, *Artifacts in CT: recognition and avoidance*, *Radiographics*, vol. 24, no. 6, pp. 1679–1691, 2004, Radiological Society of North America.

[7] R. Tokunaga, H. Suzuki, Y. Ohtake, and T. Yatagawa, *Simulation-Based Detectability Estimation for Artificial Defects on X-ray CT Volume of Metal Components*, *e-Journal of Nondestructive Testing*, vol. 27, no. 6, 2022.

[8] B. Liu, J. Bennett, G. Wang, B. De Man, K. Zeng, Z. Yin, P. Fitzgerald, and H. Yu, *Completeness map evaluation demonstrated with candidate next-generation cardiac CT architectures*, *Medical physics*, vol. 39, no. 5, pp. 2405–2416, 2012, Wiley Online Library.

[9] H. K. Tuy, *An inversion formula for cone-beam reconstruction*, *SIAM Journal on Applied Mathematics*, vol. 43, no. 3, pp. 546–552, 1983, SIAM.

[10] A. Maier, P. Kugler, G. Lauritsch, and J. Hornegger, *Discrete estimation of data completeness for 3D scan trajectories with detector offset*, *Bildverarbeitung für die Medizin 2015: Algorithmen-Systeme-Anwendungen. Proceedings des Workshops vom 15. bis 17. März 2015 in Lübeck*, pp. 47–52, Springer, 2015.

[11] R. Fujii, J. Vongkulbhisal, R. Hachiuma, and H. Saito, *A two-block rnn-based trajectory prediction from incomplete trajectory*, *IEEE Access*, vol. 9, pp. 56140–56151, 2021, IEEE.

[12] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, *Learning phrase representations using RNN encoder-decoder for statistical machine translation*, *arXiv preprint arXiv:1406.1078*, 2014.

[13] S. Kostadinov, *Understanding GRU networks*, *towards data science*, 2017.

CT Reconstruction using Nonlinear Diffusion Posterior Sampling with Detector Blur Modeling

Shudong Li, Xiao Jiang, Yuan Shen, J. Webster Stayman

***Abstract*—There has been a great deal of work seeking to improve image quality in CT reconstruction through deep-learning-based denoising; however, there are many applications where it is spatial resolution that limits application and diagnostics. In this work, we week to improve spatial resolution in CT reconstructions through a combination of deep learning and physical modeling of detector blur. To achieve this goal, we leverage diffusion models as deep image priors to help regularize a joint deblurring and reconstruction problem. Specifically, we adopt Diffusion Posterior Sampling (DPS) as a way to combine a deep prior with a likelihood-based forward model for the measurements. The model we adopt is nonlinear since detector blur is applied after the nonlinear attenuation given by the Beer-Lambert lab. We trained a score estimator for a CT score-based prior, and then apply Bayes rule to combine this prior with a measurement likelihood score for CT reconstruction with detector blur. We demonstrate the approach in simulated data, and compare image outputs with traditional filtered-backprojection (FBP) and model-based iterative reconstruction (MBIR) across a range of exposures. We find a particular advantage of the DPS approach for low exposure data and report on major differences in the errors between DPS and classical reconstruction methods.**

I. Introduction

SOPHISTICATED CT measurement models have allowed model-based iterative reconstruction (MBIR)[1], [2] to achieve improved image quality over traditional approaches. Prior efforts have detailed sophisticated models for various physical effects[3] and noise models for CT [4]. There is also much work that focuses on modeling resolution losses in the detection process and incorporating those models into reconstruction of transmission tomography for nuclear imaging[5], [6] and in flat-panel-based computed tomography[7], [8], where physical effects like light-spread in the scintillator are often a limiting factor for spatial resolution. While previous work in model-based reconstruction has demonstrated an ability to model resolution losses to extend spatial resolution capabilities, MBIR approaches can exhibit unnatural textures[9], [10].

Recently, diffusion models have been widely applied for denoising and deblurring of CT images [11], [12], including the denoising diffusion probabilistic models (DDPM) [13] and score-based generative diffusion models through stochastic differential equations (SDEs) [14]. Such models incorporate sophisticated prior knowledge of CT images and have the potential to provide superior noise control without unrealistic blurring or textures. Diffusion modes have been applied to limited-angle[15], sparse-view[16] and low-dose CT reconstruction[12], [17]. While many or most of those approaches focus on supervised learning to incorporate prior knowledge, recent research has demonstrated the potential of CT reconstruction using diffusion posterior sampling (DPS) - which combines unsupervised training of a generative prior with different measurement models without retraining[18], [19], [16], [20]. The authors of this work have previously combined diffusion posterior sampling (DPS) with a nonlinear physical model consistent with a very general transmission tomography inverse problem[21]. We will refer to this general approach as *DPS Nonlinear* for brevity.

In this work, we apply the DPS Nonlinear approach to CT reconstruction with detector blur. Specifically, the DPS Nonlinear model is sufficiently general to accommodate a linear blur operator that is applied outside the exponential (Beer-Lambert Law). Thus we adopt a physics-driven measurement likelihood model appropriate for modeling detector blur and then solve the blurry CT data inverse problem with score based prior in DPS. The approach has the advantage of other DPS methods including unsupervised training of the generative prior, general application to arbitrary measurement models, and combination of a sophisticated image prior and physics-driven measurement model.

II. Method

A. DPS with a Nonlinear Physical Model

We follow prior work on diffusion models [22], [13], where the forward diffusion process is defined by

$$\mathbf{x_t} = \sqrt{\bar{\alpha}_t}\mathbf{x_0} + (1-\bar{\alpha}_t)\epsilon_{\mathbf{noise}} \quad t \in [0,T], \tag{1}$$

where $\mathbf{x_0} \sim \mathbf{p_{data}}$ is sampled from a training dataset, t denotes time steps from 0 to T, $\epsilon_{\mathbf{noise}} \sim N(0,I)$ represents independent Gaussian noise, $\bar{\alpha}_t$ is a variance scheduler defined on $t \in [0,T]$ such that the final stage of x_T is close to a predefined noise distribution $\pi(\mathbf{x})$. Following recent studies[19], [16], [23], [24], the reverse-time diffusion process for posterior sampling is given by

$$\begin{aligned}\mathbf{dx_t} = [\mathbf{f}(\mathbf{x_t},t) + g^2(t)(\nabla_{\mathbf{x_t}} \log p(\mathbf{x_t}) + \\ \nabla_{\mathbf{x_t}} \log p(\mathbf{y}|\mathbf{x_t})]dt + g(t)\mathbf{d\bar{w}_t} \quad t \in [0,T],\end{aligned} \tag{2}$$

where $\mathbf{f}$ and g are functions defined by the forward scheduler $\bar{\alpha}_t$ from T to 0, with $\mathbf{f}(\mathbf{x_t},t) = -1/2\beta_t\mathbf{x_t}$, $g(t) = \sqrt{\beta_t}$; and $\beta_t = -1/\sqrt{\bar{\alpha}_t}\frac{d}{dt}\bar{\alpha}_t$, $\bar{\mathbf{w}}$ is a standard Wiener process when the time flows from T to 0, and dt is an infinitesimal negative timestep. Following previous studies on score matching methods [14], [25], the prior score function $\nabla_{\mathbf{x_t}} \log p(\mathbf{x_t})$ can be

Shudong Li is in the Electronic Engineering Department at Tsinghua University, Beijing, 100084, China e-mail: sli248@jh.edu

Xiao Jiang is in the Biomedical Engineering Department at Johns Hopkins University, baltimore, MD, 21219, USA, email: xjiang43@jhu.edu.

Yuan Shen is in the Electronic Engineering Department at Tsinghua University, Beijing, 100084, China, e-mail: shenyuan_ee@tsinghua.edu.cn

J. Webster Stayman is in the Biomedical Engineering Department at Johns Hopkins University, baltimore, MD, 21219, USA, e-mail: web.stayman@jhu.edu

estimated by a neural network $\mathbf{s}_\theta(\mathbf{x_t}, t)$ using the following loss function:

$$\theta^* = \arg\min_\theta \; E_{t\sim u(\epsilon,1),\mathbf{x_t}\sim p(\mathbf{x_t}|\mathbf{x_0}),\mathbf{x_0}\sim \mathbf{p_{data}}} [||\mathbf{s}_\theta(\mathbf{x_t}, t) - \nabla_{\mathbf{x_t}} \log p(\mathbf{x_t}|\mathbf{x_0})||_2^2]. \quad (3)$$

By substituting the prior score function $\nabla_{\mathbf{x_t}} \log p(\mathbf{x_t})$ with the estimated score model $\mathbf{s}_\theta(\mathbf{x_t}, t)$, the reverse-time diffusion process can be written by:

$$\mathbf{dx_t} = [\mathbf{f}(\mathbf{x_t}, t) + g^2(t)(\mathbf{s}_{\theta^*}(\mathbf{x_t}, t) + \nabla_{\mathbf{x_t}} \log p(\mathbf{y}|\mathbf{x_t}))]dt + g(t)\mathbf{d\bar{w}_t} \quad t \in [0, T]. \quad (4)$$

This continuous structure of Equation (4) allows for the sampling of data from $\mathbf{p_0}(\mathbf{x_0}|\mathbf{y})$ given the posterior $\mathbf{y}$, and the forward process $\mathbf{p_t}(\mathbf{y}|\mathbf{x_t})$. Following Chung et al.[19], the likelihood term $\nabla_{\mathbf{x_t}} \log p(\mathbf{y}|\mathbf{x_t})$ can be approximated according to Tweedie's formula as follows

$$\nabla_{\mathbf{x_t}} \log p(\mathbf{y}|\mathbf{x_t}) = \nabla_{\mathbf{x_t}} \hat{\mathbf{x}}_\mathbf{0}(\mathbf{x_t}) \nabla_{\mathbf{x_0}} \log p(\mathbf{y}|\mathbf{x_0} = \hat{\mathbf{x}}_\mathbf{0}) \quad \text{where } \hat{\mathbf{x}}_\mathbf{0} = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x_t} + (1 - \bar{\alpha}_t)\mathbf{s}_{\theta^*}(\mathbf{x_t}, t)), \quad (5)$$

where $\hat{\mathbf{x}}_\mathbf{0}$ denotes Tweedie's approximation of $\mathbf{x_0}$ from $\mathbf{x_t}$, $\nabla_{\mathbf{x_t}} \hat{\mathbf{x}}_\mathbf{0}(\mathbf{x_t})$ denotes the derivative of $\mathbf{x_0}$ with respect to $\mathbf{x_t}$, and $\nabla_{\mathbf{x_0}} \log p(\mathbf{y}|\mathbf{x_0} = \hat{\mathbf{x}}_\mathbf{0})$ is log-likelihood of $p(\mathbf{y}|\mathbf{x_0})$. Following our previous work on DPS Nonlinear[21], we adopt the following forward model for measurements with Gaussian noise:

$$\mathbf{y} \sim N(\bar{\mathbf{y}}, \mathbf{D}\{\bar{\mathbf{y}}\}) \;\text{ where }\; \bar{\mathbf{y}} = \mathbf{B}\exp(-\mathbf{A}\mathbf{x_0}), \quad (6)$$

where $\mathbf{A}$ is the linear projection matrix defined for the given CT imaging system, $\mathbf{B}$ is a matrix operator that can model various physical effects including detector blur, $\mathbf{x_0}$ denotes the image of attenuation coefficients in transmission tomography, $\bar{\mathbf{y}}$ denotes the mean measurement vector, $\mathbf{D}\{\cdot\}$ represent a diagonal matrix formed from its vector argument, and $\mathbf{y}$ represents the random vector containing noisy CT measurements. Independent Gaussian noise with measurement variance $\sigma_\mathbf{y} = \mathbf{y}$ approximates Poisson noise from x-ray quanta. We may write the measurement log-likelihood as

$$\nabla_{\mathbf{x_0}} \log p(\mathbf{y}|\mathbf{x_0}) = -\mathbf{A}^\mathrm{T}\mathbf{D}\{\exp(-\mathbf{A}\mathbf{x_0})\} \mathbf{B}^\mathrm{T}\mathbf{D}\{1/\mathbf{y}\}(\mathbf{y} - \mathbf{B}\exp(-\mathbf{A}\mathbf{x_0})), \quad (7)$$

where $\{\cdot\}^\mathrm{T}$ denotes a matrix transpose and $\mathbf{A}^\mathrm{T}$ is the back-projection operator.

B. Definition of Detector Blur

As in previous work[26], [7], $\mathbf{B}$ can be model detector blur - e.g. from light-spread in the scintillator, charge spreading effects, etc. We factor this terms explicity as

$$\mathbf{B} = \mathbf{B}_d I_0, \quad (8)$$

where I_0 is a scalar gain term (presuming uniform x-ray fluence as well as detector sensitivity), and $\mathbf{B}_d$ models all detector blur effects. While the operator is general and may be any linear operation, here we focus on a purely convolution blur. Thus, one does not need to explicity form the matrix $\mathbf{B}_d$, and its action may be performed through convolution or Fourier transform operations.

C. Data Preparation

The XCAT phantom[27] was used for the generation of 2D simulated data. We used a volume of $256 \times 256 \times 200$ voxels with $0.8mm \times 0.8mm \times 1.2mm$ voxels. The system geometry used a source-to-axis and source-to-detector distance equal to $800\ mm$ and $1500\ mm$, respectively. The forward model presumed a monoenergetic beam of $60\ keV$ and used 360 projections covering $360°$. Forward projection and backprojection were implemented using the LivermorE AI Projectors(LEAP) [28]. Detector blur is modeled with a Gaussian kernel with $\sigma = 1$ pixel for a full-width, half-maximum blur of ~ 2.4 pixels. A total of 54 human models were generated and divided into training, validation, and test sets with a ratio around 8:1:1. Individual slices were then extracted to create 2D simulated data based on the models to avoid mixing between training, validation, and testing.

D. Unconditional Training of the Diffusion Model

In this work, we adopted a U-Net architecture network [29] with an attention residual and time embedding module [14] for the prior score function in the unconditional score-based diffusion model. This network is trained independently from the measurement model in an unsupervised fashion. For the stochastic process, we predefined a schedule by $\bar{\alpha}_t = 1e^{-5t}$ with $t \in (0, T]$ where $T = 1$. The network was implemented based on the PyTorch framework. We applied the Adam optimizer for neural network training, setting the batch size to 8 and the initial learning rate to 0.001. It is trained on a single NVIDIA RTX A6000 GPU and the total number of epochs was set to 600 with an early stopping criterion (no decrease in loss over 100 epochs). Best weights were chosen based on the validation loss over all epochs.

E. DPS Reconstruction with Detector Blur Modeling

We adopted the Variance Preserving Stochastic Stochastic Differential Equations (VP-SDEs) to discretize the reverse process into N bins. Following Chung et al. [19], the derivative of the prior score function $\mathbf{s}_{\theta^*}(\mathbf{x_t}, t)$ with respect to the input $\mathbf{x_t}$ is computed by backward propagation in the score model. We also applied a time-dependent coefficient scheduler $\lambda(t)$ in front of the likelihood function to balance the weighting between likelihood and score function, which is described as

$$\lambda_t = k\frac{||\mathbf{s}_{\theta^*}(\mathbf{x_t}, t)||_2^2}{||\nabla_{\mathbf{x_t}} \log p(\mathbf{y}|\mathbf{x_t})||_2^2}, \quad (9)$$

where k can be set manually to balance the weighting.

F. Evaluation and Comparison of Reconstruction Methods

For comparison, we compared the DPS Nonlinear approach with traditional reconstruction methods including filtered-backprojection (FBP) and likelihood-based MBIR. For MBIR,

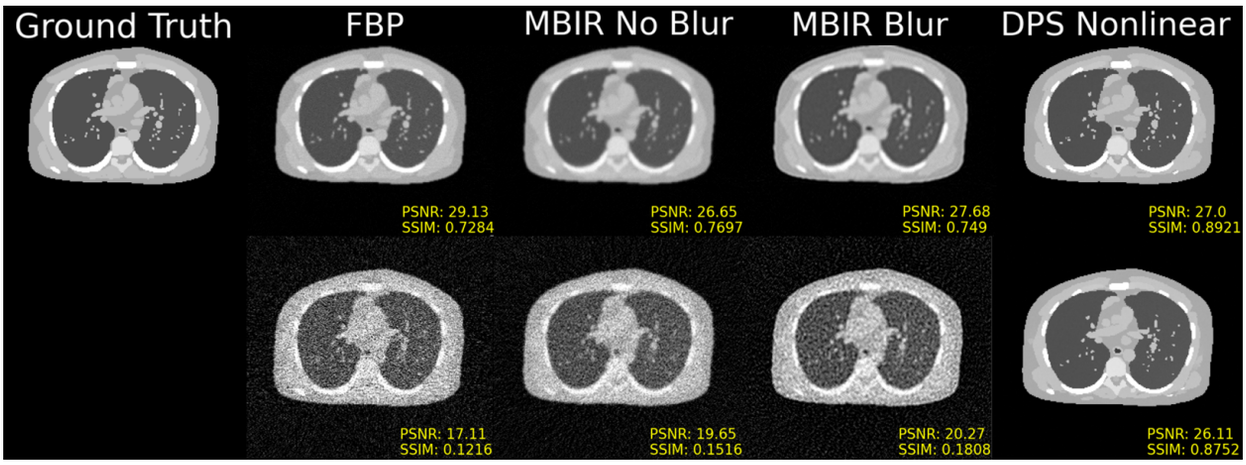


Fig. 1: Reconstructions with detector blur comparing FBP, MBIR, and the DPS Nonlinear approach. Top row: $I_0 = 10^5$ photons/pixel. Bottom row: $I_0 = 10^3$ photons/pixel. All images share a common colormap with W/L = 2000/0 HU.

we investigate two two different variants - one approach that does not have a detector blur model ($\mathbf{B}_d = \mathbf{I}$), and an approach where the reconstruction blur model matches the measurement data. We refer to these strategies with the shorthand notation: "MBIR No Blur" and "MBIR Blur," respectively. For optimization of these approaches we used the same PyTorch implementation and use a simple gradient descent optimization. For this approach, we applied an adaptive step size using the following equation

$$step_size = \frac{g}{||\nabla_{\mathbf{x_t}} \log p(\mathbf{y}|\mathbf{x_t})||_2^2}, \tag{10}$$

where g is a weighting term that can be set manually.

To evaluate different approaches, we considered a range of x-ray exposures from $I_0 = 5x10^2$ to $I_0 = 10^5$ photons/detector element (barebeam). To evaluate the variability, we reconstructed 16 noisy realizations of CT data and computed sample bias and variance images and metrics. We evaluate these metric as a function of exposure. We also show some selected reconstructions and report Peak Signal-to-Noise ratio (PSNR) and Structured Similarity (SSIM).

For MBIR, g is set to 0.01 with 14000 iterations for I_0 = 10^5 and 10^4, and g is set to 0.005 with 28000 iterations for I_0 = 5×10^3, 10^3 and 5×10^2. For DPS Nonlinear, the number of reverse time steps in the reverse-diffusion is set to 1000. The lambda coefficient k is set to 0.24, 0.22, 0.20, 0.18, and 0.16 for five different I_0 levels: 10^5, 10^4, 5×10^3, 10^3, and 5×10^2, respectively.

III. Results

A. CT Reconstruction with Detector Blur

Figure 1 shows the reconstruction results for the $I_0 = 10^5$ and $I_0 = 10^3$ photons/pixel scenarios. For higher exposure scenario, we blur in both the FBP and MBIR No Blur scenarios. The MBIR with blur modeling case shows some resolution enhancement, but also exihibits some of the classic ringing/overshoot associated with imaging that uses deconvolution to enhance spatial resolution. The proposed diffusion posterior sampling (DPS) reconstruction shows sharp anatomical structures, and good general agreement with the ground truth. However, we do note that some fine structures and boundaries are mismatched with the ground truth. Across all cases, the noise levels are relatively low, so that the main errors in estimates are biases (e.g. blur or the as opposed to noise. Quantitatively, the PSNRs are similar, but the SSIM is highest for DPS Nonlinear.

Differences are a bit more evident in the lower exposure scenario. Increased noise is apparent in the FBP reconstruction. This noise is somewhat reduced in the MBIR No Blur case. The MBIR with blur modeling case shows increased sharpness but also significantly increased noise due to the deblurring. radiation dose. The proposed DPS Nonlinear approach yields a sharp and highly denoised image. Again, some small features in DPS Nonlinear different from the ground truth. Quantitatively, there are significant differences between DPS Nonlinear and the other approaches with PSNR and SSIM nearly reaching the values in the $I_0 = 10^5$ case. For the remaining methods, both PSNR and SSIM follow a rank ordering from FBP (worst) to MBIR No Blur to MBIR Blur (best).

B. Investigations of Bias and Variability

We computed the root mean-squared bias and average standard deviation across the entire image for each reconstruction approach using 16 noisy data realizations. A summary of these metrics is shown in Figure 2. For all methods there is a general trend of increasing bias and standard deviation for

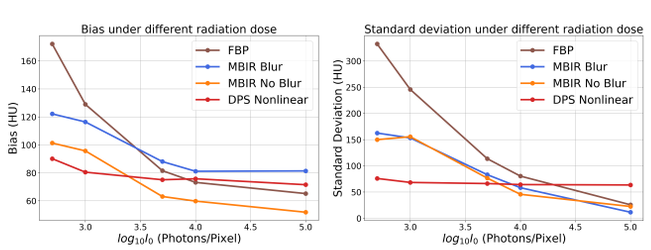


Fig. 2: Root mean-squared bias and standard deviation calculated based on an ensemble of 16 noisy data realizations across different exposure levels.

lower exposure settings. (However, the bias trend is likely due, in part, to the limited size of the data ensemble.) Interestingly, the DPS Nonlinear approach appears to have the greatest advantage at the lowest exposures, and the standard deviation for DPS Nonlinear is very flat. We note that standard deviation as a variability metric as applied here accounts for both the aleatoric uncertainty associated with noisy measurement data as well as the epistemic uncertainty associated with any particular DPS estimate.

A more detailed look at errors in the $I_0 = 10^3$ case is shown in Figure 3. In looking at the bias images, for FBP and the MBIR cases, bias is concentrated around the edges - consistent with a resolution loss in the reconstruction. (We also note a slight increase in bias outside the phantom - likely due to a nonnegativity constraint in reconstruction.) The bias image for DPS Nonlinear is also generally concentrated at boundaries and edges, though is nearly zero for the uniform regions in this phantom. The variance images show the relatively high noise levels of FBP and reduced noise in MBIR. The variance image for DPS Nonlinear is highly nonuniform - concentrating most variability at the edges and boundaries (and for small features).

IV. Discussion and Conclusion

We have extended our previous work on diffusion posterior sampling (DPS) with a nonlinear physics-driven CT model to incorporate detector blur. This technique allows unsupervised training of a generative prior image model and application to arbitrary system designs/protocols without retraining. We demonstrated this approach in a simulated CT example with detector blur and show potential advantages of the approach - particularly at low exposure levels. We observe that DPS Nonlinear has very different bias and variance patterns compared with standard reconstruction approaches, and that the method is susceptible to distortions/hallucinations of boundaries and small features. This has been observed in related DPS work, and, in ongoing work, we are looking at approaches to further stabilize the DPS Nonlinear process.

Acknowledgment

Work was supported, in part, by NIH grant R01CA249538.

References

[1] Thibault, J.-B *et al.*, *Medical physics*, vol. 34, no. 11, pp. 4526–4544, 2007.
[2] Yoon, M. A *et al.*, *Journal of computer assisted tomography*, vol. 36, no. 5, pp. 596–601, 2012.
[3] Nuyts, J *et al.*, *Physics in Medicine & Biology*, vol. 58, no. 12, p. R63, 2013.
[4] Whiting, B. R, in *Medical Imaging 2002: Physics of Medical Imaging*, vol. 4682. SPIE, 2002, pp. 53–60.
[5] Yu, D. F *et al.*, *IEEE transactions on medical imaging*, vol. 19, no. 11, pp. 1094–1105, 2000.
[6] Feng, B *et al.*, *IEEE transactions on medical imaging*, vol. 25, no. 7, pp. 941–949, 2006.
[7] Stayman, J. W *et al.*, in *Medical Imaging 2014: Physics of Medical Imaging*, vol. 9033. SPIE, 2014, pp. 908–913.
[8] Tilley, S *et al.*, *IEEE transactions on medical imaging*, vol. 37, no. 4, pp. 988–999, 2017.
[9] Park, C *et al.*, *European radiology*, vol. 31, pp. 3156–3164, 2021.
[10] Kaga, T *et al.*, *Clinical Radiology*, vol. 76, no. 9, pp. 710–e15, 2021.
[11] Kazerouni, A *et al.*, *Medical Image Analysis*, p. 102846, 2023.
[12] Lei, Y *et al.*, *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2023.
[13] Ho, J *et al.*, *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
[14] Song, Y *et al.*, *arXiv preprint arXiv:2011.13456*, 2020.
[15] Liu, J *et al.*, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 498–10 508.
[16] Chung, H *et al.*, *Advances in Neural Information Processing Systems*, vol. 35, pp. 25 683–25 696, 2022.
[17] Tivnan, M *et al.*, *arXiv preprint arXiv:2303.13285*, 2023.
[18] Song, Y *et al.*, *arXiv preprint arXiv:2111.08005*, 2021.
[19] Chung, H *et al.*, *arXiv preprint arXiv:2209.14687*, 2022.
[20] Lopez Montes, A *et al.*, in *Int'l Mtg. on Fully 3D Image Recon in Rad. & Nuc. Med.*, 2023.
[21] Li, S *et al.*, *arXiv preprint arXiv:2312.01464*, 2023.
[22] Yang, L *et al.*, *arXiv preprint arXiv:2209.00796*, 2022.
[23] Feng, B. T *et al.*, *arXiv preprint arXiv:2304.11751*, 2023.
[24] Meng, X and Kabashima, Y, *arXiv preprint arXiv:2211.12343*, 2022.
[25] Hyvärinen, A and Dayan, P, *Journal of Machine Learning Research*, vol. 6, no. 4, 2005.
[26] Tilley II, S *et al.*, in *Medical Imaging 2016: Physics of Medical Imaging*, vol. 9783. SPIE, 2016, pp. 194–199.
[27] Segars, W. P *et al.*, *Medical physics*, vol. 37, no. 9, pp. 4902–4915, 2010.
[28] Kim, H and Champley, K, *arXiv preprint arXiv:2307.05801*, 2023.
[29] Ronneberger, O *et al.*, in *MICCAI 2015*. Springer, 2015, pp. 234–241.

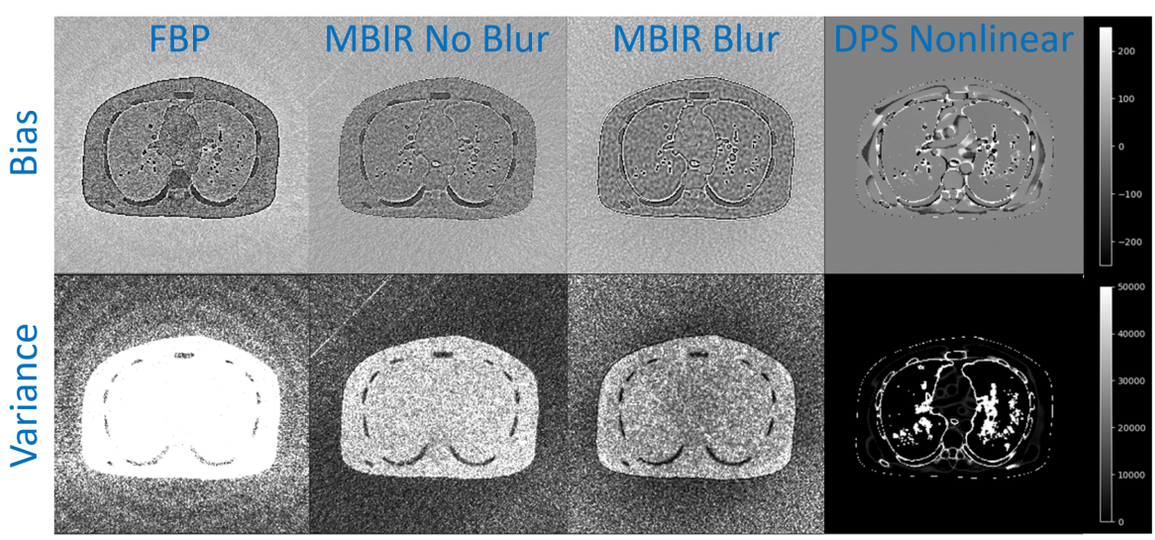


Fig. 3: Bias and variance image for different reconstruction approaches for the I0 = 10^3 photons/detector element scenario.

Improved accuracy in CT reconstruction from transversely truncated data

Zheng Zhang, Buxin Chen, Dan Xia, Emil Y. Sidky, and Xiaochuan Pan

***Abstract*—In fan-beam CT, when the field-of-view (FOV) of the X-ray beam received by the detector is smaller than the support of the subject scanned, it yields data with transverse truncation. Interest remains in image reconstruction from truncated data, and algorithms have been investigated for image reconstruction within the FOV. In this work, we focus on investigating accurate image reconstruction not only within the FOV but also, more importantly, within a region significantly larger than the FOV from data collected with truncation configurations. We formulate the reconstruction problem from truncated data as a convex optimization program with hybrid constraints on image-region total-variation (TV) and ℓ_1-norm, and develop a primal-dual-based algorithm, referred to as TV-L1 algorithm, to solve the optimization problem for achieving image reconstruction within the subject support. Numerical studies are conducted with noiseless and noisy data of varying truncation degrees for verifying and demonstrating the algorithm performance and robustness. Results of the studies reveal that the TV-L1 algorithm can accurately reconstruct images within the FOV, and that it may also yield accurate images beyond the FOV within the subject support for truncation extents of practical interest. The approach and TV-L1 algorithm developed in the work can be extended to cone-beam CT.**

***Index Terms*—computed tomography, truncation, image reconstruction, primal-dual algorithm, hybrid image constraints**

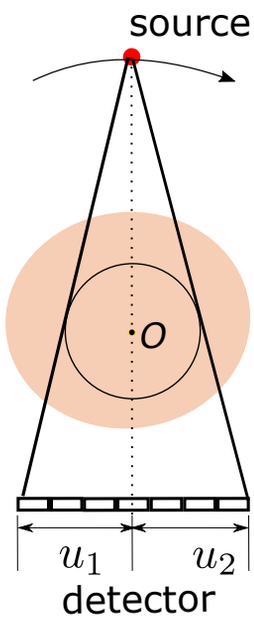


Fig. 1. Scan configurations with symmetrical truncation with FOV highlighted by the dark circles, smaller than the subject support (orange). The center-of-rotation is denoted by "O", whereas the X-ray source is depicted by the red dot.

I. Introduction

In CT applications, it is not uncommon that the field of view (FOV) of the X-ray beam received by the detector is smaller than the support of the subject scanned, thus resulting in truncated data. Addressing the challenge of image reconstruction from truncated data is a subject of interest, with most efforts focused on algorithms tailored for reconstructing images within the FOV [1]. However, there remains interest, both theoretically and practically, to delve deeper into achieving accurate image reconstruction not only within the FOV but also extending beyond it, while mitigating truncation artifacts and improving image quality.

In this work, we investigate accurate image reconstruction from data collected with varying truncation degrees. Our approach aims to reconstruct images not only within the FOV but also within a region substantially larger than the FOV. To tackle this challenge, we devise the reconstruction problem from truncated data as a convex optimization program. In particular, we seek to minimize the fidelity error of data ℓ_2-norm while incorporating hybrid constraints encompassing image-region total-variation (TV) [2], [3] and ℓ_1-norm. These constraints are strategically imposed across different regions, improving image accuracy within the FOV and potentially enabling accurate reconstructions beyond it, within the subject's support.

To realize this innovative approach, we tailor and develop a primal-dual (PD)-based algorithm [4]–[6], referred to as the TV-L1 algorithm, to solve the optimization problem and achieve accurate image reconstruction. To validate the efficacy and robustness of our proposed method, numerical studies are conducted using both noiseless and noisy data sets with varying degrees of truncation. Through these investigations, we aim to verify and evaluate the performance and robustness of the TV-L1 algorithm in image reconstruction from truncated CT data.

II. Material and Methods

A. Optimization program

We formulate the reconstruction problem from truncated data as a convex optimization program with a data-ℓ_2-minimization term and hybrid, image-region TV and ℓ_1-norm constraints:

$$\mathbf{f}^\star \quad = \quad \Phi(\mathcal{H}\mathbf{f}, \mathbf{g}^{[\mathcal{M}]}) \quad \text{s.t.} \quad \Psi(\mathbf{f}), \tag{1}$$

Z. Zhang, B. Chen, D. Xia, and E. Y. Sidky are with the Department of Radiology, The University of Chicago, Chicago, IL 60637, USA.

X. Pan [email:xpan@uchicago.edu] is with the Departments of Radiology & Radiation and Cellular Oncology, The University of Chicago, Chicago, IL 60637, USA.

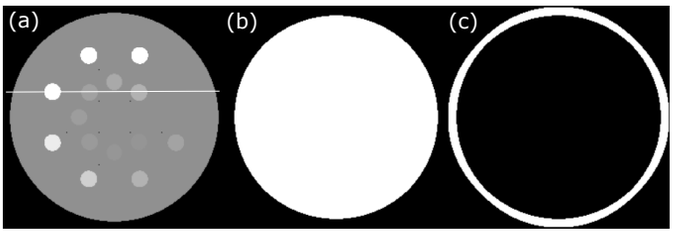


Fig. 2. (a) Truth image, (b) inner region, and (c) outer region selected. The profiles of reconstructions below are plotted along the white line in (a).

where $\Phi(\mathcal{H}\mathbf{f}, \mathbf{g}^{[\mathcal{M}]})$ denotes the ℓ_2-norm between $\mathcal{H}\mathbf{f}$ and $\mathbf{g}^{[\mathcal{M}]}$; vector $\mathbf{f}$ is the image; $\mathcal{H}$ denotes a discrete form of the 2D X-ray transform; $\mathbf{g}^{[\mathcal{M}]}$ represents measured data. The term $\Psi(\mathbf{f})$ encapsulates the image-constraint components, which consist of three distinct parts: an image-total-variation (TV) constraint within an inner region M_i, an image-ℓ_1 constraint within an outer region M_o, and an image non-negativity constraint. It's important to note that the inner region M_i and outer region M_o are complementary to each other, and their combination covers the image support or extends beyond it. For a visual representation of the inner and outer regions, please refer to Fig. 2.

B. The TV-L1 algorithm for image reconstruction

We then derive a PD-based algorithm [4], [6], referred to as the TV-L1 algorithm in the work, to solve the optimization program in Eq. (1). Obviously, the final reconstruction of the TV-L1 algorithm can be affected by constraint parameters, as well as the selection of inner and outer regions M_i and M_o, in Eq. (1). In a simulated-data study, we compute these parameters and select M_i and M_o based upon the truth phantom, whereas in a real-data study, we perform reconstructions for a slew of these parameters selected and determine empirically the final parameters that yield visually and quantitatively best images among the slew of parameters. In this work, we include below results of both simulated noiseless and noisy data. It is worth noting that the TV-L1 algorithm becomes the TV algorithm [2], [3] when the inner region M_i covers the subject support and the outer region M_o is empty in Eq. (1). For comparison, we also reconstruct images by using the TV and FBP algorithms.

C. Study design

1) Scan configurations: We consider a fan-beam X-ray scanning setup employing a one-dimensional linear detector, configured in a symmetrical truncation arrangement, as depicted in Fig. 1. The distances from the source to rotation and from the source to the detector are 100 cm and 150 cm, respectively. The linear detector comprises 750 bins, each with a width of 0.82 mm. A total of 720 views are evenly distributed across an angular span of 2π. In Fig. 1, the "central line" at each view is defined as the line connecting the X-ray source and the center of rotation denoted as O.

TABLE I
Configurations used in the numerical study below. Note that the full detector is of size 750 bins completely covering the subject support.

Type	S1	S2
(u_1,u_2)	(100, 100)	(80, 80)

2) Phantom: We have conducted studies with numerical and physical phantoms of varying anatomy and report below the study result of a numerical disk phantom. The numerical phantom mimicks the standard disk phantom used widely in CT imaging for scan calibration and reconstruction evaluation, that consists of multiple inserts of solid water, calcium, and iodine. The phantom is discretized on a grid of 256×256 square pixels, each measuring 1.6 mm, as depicted in Fig. 2a.

3) Selection of M_i and M_o: The selection of inner and outer regions, represented by M_i and M_o, respectively, can significantly influence image reconstruction outcomes. We have explored various sizes for M_i and M_o to assess their impact on reconstruction quality. However, for clarity and consistency, this work exclusively presents reconstructions using a specific configuration of M_i and M_o. The corresponding inner and outer regions utilized in our study are illustrated in Figs. 2b and 2c, respectively, with the outer region encapsulating a narrow band along the phantom boundary, spanning a width of 10 pixels. It's crucial to emphasize that the selection of these regions remains independent of the imaging FOV. Furthermore, in our investigation, the outer region intentionally maintains a significantly smaller size compared to the corresponding truncation extent.

4) Data generation: We first generate noiseless full data of the phantom without any truncation at each view. Subsequently, we create truncated data sets with different degrees of truncation by removing the appropriate proportions of full data from both sides of the detector for all of the views. As illustrated in Fig. 1, the two end detector bins are positioned at distances $u_1 \neq 0$ and $u_2 \neq 0$, respectively, to the line connecting the source and the center-of-rotation, forming a symmetrical truncation configuration with $u_1 = u_2$. By varying the values of u_1 and u_2, we obtain noiseless truncated data from the noiseless full data for each of the scanning configurations outlined in Table I. The ratios between detector sizes $u_1 + u_2$ of pairs S1 and S2, respectively, and the full detector size (750 bins) are 27% and 21%.

Additionally, we generate noisy data by introducing Poisson noise to the noiseless data, corresponding to an equivalent of 10^5 photons per ray.

III. Numerical studies

A. Verification study of the TV-L1 algorithm

We first verify the TV-L1 algorithm and its implementation by conducting a numerical verification study from noiseless full data. Results show that the TV-L1 algorithm yields reconstruction that is identical quantitatively with the ground truth at the level of computer floating-point precision. The reconstruction, the truth, and their difference images are

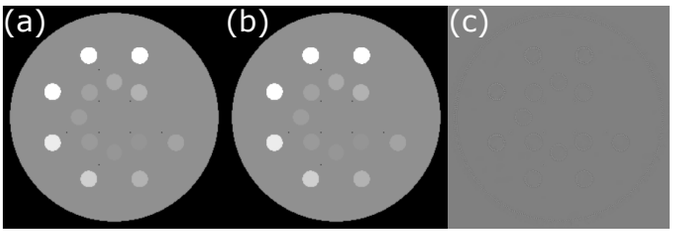


Fig. 3. Truth image (a), TV-L1 reconstruction from noiseless full data (b), and difference image (c) between (a) and (b). Display window: [0, 0.3] cm^{-1} for (a) and (b), and [-10^{-5}, 10^{-5}] cm^{-1} for (c).

shown in Fig. 3, revealing that the reconstruction of the TV-L1 algorithm is visually and quantitatively identical to the truth. All algorithm's convergence conditions are achieved, indicating the correctness of the TV-L1 algorithm and its implementation.

Following the successful verification of the TV-L1 algorithm and its implementation, we proceed to apply it to investigate numerical reconstructions from truncated data.

B. Numerical studies with noiseless truncated data

We then utilize the verified TV-L1 algorithm to reconstruct images from noiseless data obtained with the symmetrical truncation configurations listed in Table I. In the first row of Fig. 4, we present images reconstructed from data collected with the symmetrical truncation configurations S1 and S2. For comparison, images reconstructed from the same truncated data using the TV and FBP algorithms are displayed in the second and third rows of Fig. 4, respectively.

It is evident that the TV-L1 algorithm accurately reconstructs the image within the entire subject support, extending beyond the FOV. Additionally, we observe that the TV algorithm also achieves accurate reconstruction within the entire subject support. However, as demonstrated subsequently, the accuracy of the TV reconstruction deteriorates more rapidly than that of the TV-L1 algorithm when the truncated data contain noise or other inconsistencies. Due to significant artifacts and distortions evident in the FBP reconstructions, we opt not to include FBP results in the subsequent noisy-data studies.

C. Numerical studies with noisy symmetrically truncated data

In an attempt to demonstrate the robustness of the TV-L1 algorithm in yielding accurate reconstruction from data containing inconsistencies, we apply it to reconstructing images from noisy data collected with symmetrical truncation configurations S1 and S2 in Table I. In Fig. 5, images reconstructed by use of the TV-L1 algorithm are displayed in rows 1 and 3, respectively, with wide and narrow display windows. It can be observed that the images reconstructed appear of little visible artifacts except for some minor artifacts for S2 with the largest truncation extent. The observation is corroborated by the profile results of the images reconstructed, as shown in the top panel of Fig. 6. The image profiles appear to agree largely with the truth, while slight discrepancies over some of the insert regions can be observed.

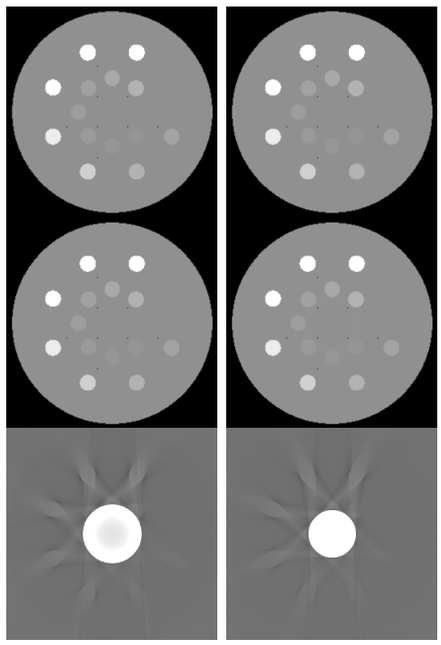


Fig. 4. Images reconstructed from truncated data of configurations S1 (left) and S2 (right) in Table I by use of the TV-L1 (row 1), TV (row 2), and FBP (row 3) algorithms, respectively. Display window: [0, 0.3] cm^{-1}.

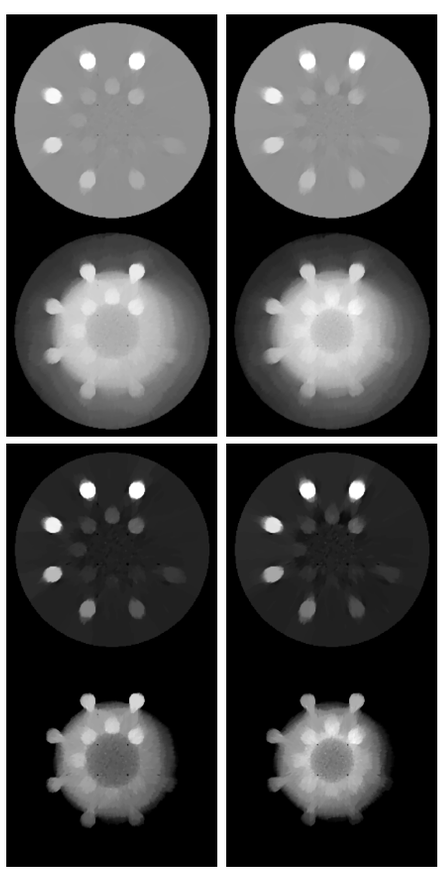


Fig. 5. Images reconstructed from data of configurations S1 (left) and S2 (right) in Table I by use of the TV-L1 (rows 1 & 3) and TV (rows 2 & 4) algorithms. Wide window: [0., 0.3] cm^{-1} for rows 1 and 2, whereas narrow window [0.15, 0.3] cm^{-1} for rows 3 and 4.

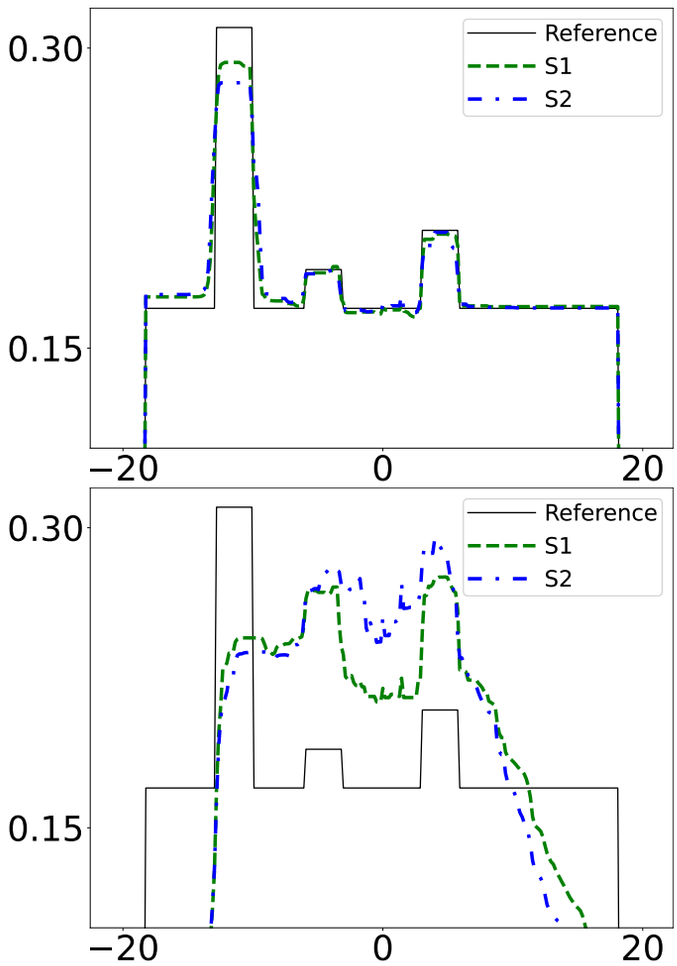


Fig. 6. Profiles of images in Fig. 5 over the white line depicted in Fig. 2a: TV-L1 (top) and TV (bottom) reconstructions from data of configurations S1 (green, dash line) and S2 (blue, dashdotted line). Solid black line indicates the truth profile.

For comparison, we also apply the TV algorithm to reconstructing images from the same noisy truncated data and display them in rows 2 and 4 of Fig. 5, respectively, with wide and narrow display windows. The images reconstructed contain significant visual artifacts that are similar to those observed often in images reconstructed from truncated data by use of other algorithms in an interior problem. This observation is corroborated also by the profile results in the bottom panel of Fig. 6 in which the profiles of the TV reconstructions significantly deviate from the truth.

We have conducted reconstructions with additional data of varying truncation levels, and the observations similar to those discussed above can also be made.

IV. Discussion

In this work, we investigate and develop an optimization-based algorithm, referred to as the TV-L1 algorithm, for image reconstruction within the subject support that encompasses the system FOV from truncated data collected with symmetrical truncation configurations in fan-beam CT. Firstly, we conduct a verification study to verify the correctness of the TV-L1 algorithm and its implementation, ensuring its reliability in subsequent studies. With verification established, we proceed to reconstruct images within the subject support from noiseless data featuring varying levels of truncation, providing insights into the upper-bound performance of the TV-L1 algorithm. To further evaluate its robustness, we apply the TV-L1 algorithm to reconstruct images from noisy data with varying truncation levels. Our findings indicate that the TV-L1 algorithm effectively yields images of improved accuracy within a region significantly larger than the FOV, showing a high degree of performance robustness. For comparison, we also employ the TV algorithm to reconstruct images from the same truncated data. While the TV algorithm accurately reconstructs images within the subject support from noiseless data, its performance declines rapidly when applied to noisy truncated data, resulting in reconstructions with visual artifacts and quantitatively inaccurate grey values.

The comparison study of the TV-L1 and TV algorithms clearly suggests that the key factor for the reconstruction robustness of the TV-L1 algorithm is the design and use of the hybrid constraints on image-region TV and ℓ_1-norm. Moving forward, we plan to conduct additional studies to demonstrate the algorithm's robustness concerning the selection of constraint parameter values and its performance using real data.

Furthermore, we note that when the detector size is fixed, asymmetrical truncation configurations may offer advantages over symmetrical ones, particularly when the detector covers less than one half of the subject support. Thus, investigating the TV-L1 algorithm's performance and robustness for image reconstruction from data collected with asymmetrical truncation configurations is of considerable interest in future studies. Lastly, the methodology and TV-L1 algorithm developed in this study can be extended to 3D image reconstruction in cone-beam CT (CBCT) imaging.

Acknowledgment

This work was supported in part by NIH Grant Nos. R01EB026282 and 5R21CA263660-02. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] B. Ohnesorge, T. Flohr, K. Schwarz, J. Heiken, and K. Bae, "Efficient correction for CT image artifacts caused by objects extending outside the scan field of view," *Med. Phys.*, vol. 27, no. 1, pp. 39–46, 2000.

[2] E. Y. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *J. X-Ray Sci. and Technol.*, vol. 14, pp. 119–139, 2006.

[3] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, pp. 4777–4807, 2008.

[4] 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.

[5] 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.

[6] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, Princeton, New Jersey, 1999.

KVp-switching Sparse-View CT Images Restoration with an Edge Prior Conditioned DDPM

Nimu Yuan, Jian Zhou, and Jinyi Qi

Abstract— **This study introduces an innovative approach for restoring sparse-view CT (SVCT) images in kVp-switching Dual-Energy CT (DECT) systems using an edge prior conditioned Denoising Diffusion Probabilistic Model (DDPM). DECT systems with kVp-switching acquire CT images at two energies in a single rotation, thus minimizing mis-registration between the two images, but suffers high noise and sparse view artifacts. To address these issues, we employed a deep neural network, namely, DDPM with an edge prior, to restore the full view image at each energy. We found that the edge map obtained from the mixed kVp projections is very robust to the kVp-switching and it can guide the deep network in reducing noise and artifacts while restoring fine structures. Our method uses training data generated from existing single-kVp CT scans and circumvents the need for paired kVp-switching and reference full-view CT scans, a major hurdle in conventional deep-learning approaches. The results demonstrate significant improvements in image quality, particularly in streaky artifact reduction, texture and structure restoration, highlighting the potential of our method in clinical applications. This research not only addresses a critical challenge in kVp-switching DECT imaging but also opens new avenues for effective utilization of existing single-energy CT data in advanced imaging techniques.**

Index Terms—**SVCT, kVp-switching, DECT, Deep Learning, DDPM, edge prior**

I. INTRODUCTION

The advent of kVp-switching CT systems [1] represents a significant advancement in this field, demonstrating the potential application of conventional single-source CT while providing dual-energy imaging capabilities that enhance tissue characterization and material differentiation. Unfortunately, such switching introduces unique challenges, primarily the manifestation of sparse-view (SV) streaky artifacts inherent to both subset of high- and low-kVp data, which may deteriorate the diagnostic value of CT images particularly in the context of fine structures and details and render the subset images unusable for further material decomposition algorithm.

Numerous deep learning-based methods [2], [3] have been proposed to restore SVCT images and have shown impressive results. However, these networks typically require paired high-quality full-view (FV) CT scans as reference labels for training, posing a substantial obstacle due to the scarcity of such paired data in kVp-switching contexts. While it is possible to train these networks using only single-kVp SVCT images, this approach inherently omits all information from the missing views, potentially limiting the network's ability to restore details accurately. On the other hand, employing both high- and low-kVp SVCT images as inputs necessitates the availability of corresponding paired high- and low-kVp FVCT images as labels. This requirement significantly complicates the data collection process, posing a substantial hurdle in the development and optimization of these deep learning methods.

In this study, we propose a novel approach utilizing an edge prior conditioned denoising diffusion probabilistic model (DDPM) to address the challenges. Our method uses existing single-kVp CT scans to generate SV training data. To provide additional information, an edge map is provided to the DDPM as a prior. We found that the edge map is robust with respect to kVp switching, so we can use edge maps from full view single-kVp image for training and those from mixed energy scans during network deployment. Our results demonstrate that the edge maps, representing detailed structural information free of SV streaky artifacts, can effectively guide the network in reducing noise and artifacts while restoring fine structures. The proposed method provides a new approach for effective utilization of existing single-kVp data in dual-energy CT (DECT) imaging.

II. MATERIALS AND METHODS

A. The kVp-switching system

We simulate a kVp-switching DECT system, where an acquisition cycle consists of 3 high-kVp projections, followed by 1 transition projection, then 3 low-kVp projections, and another transition projection, i.e., a 3-1-3-1 pattern [1]. Consequently, both the high-kVp and low-kVp datasets can be categorized as SVCT data. These datasets are prone to significant sparse-view streaky artifacts and high noise when images are reconstructed using the Filtered Back Projection (FBP) algorithm. The primary objective of this study is to enhance the image quality of these SVCT images, transforming them into their corresponding single-kVp FVCT counterparts.

B. DDPM Network

A 3D conditional DDPM was utilized to enhance the SVCT images (Fig. 1). The architecture of this network is based on the framework outlined in our previous study [4], but with modifications to adapt it into a 3D patch-based model. As illustrated in Fig. 1, the conditional input of the network was enriched by including original SVCT I_{SVCT}, sinogram-domain

Manuscript received January 23, 2024.

N. Y. is with Biomedical Engineering Department, University of California, Davis, Davis, CA 95616 USA (e-mail: nmyuan@ucdavis.edu).

J. Z., is with Canon Medical Research USA, INC, Vernon Hills, IL 77005 USA. (e-mail: jzhou@mru.medical.canon).

J. Q. is Biomedical Engineering Department, University of California, Davis, Davis, CA 95616 USA (e-mail: qi@ucdavis.edu).

[1] Our proposed method can also be applied to other kVp switching patterns.

interpolated SVCT I_{interp}, and the corresponding edge map I_{edge}. The loss function is as follows:

$$L = E_{x_0,n,k}[\|n - n_\theta(x_k, k, I_{SVCT}, I_{interp}, I_{edge})\|] \quad (1)$$

where x_0 is the original image sample; n is the added noise; k is the timestep in the diffusion process; x_k is the noise-added image at timestep k; and $n_\theta(x_k, k, I_{SVCT}, I_{interp}, I_{edge})$ is the network prediction of the noise.

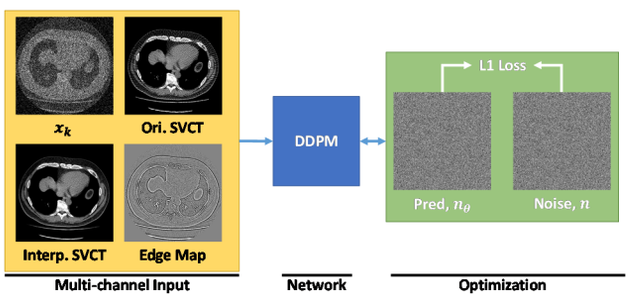


Fig. 1. The schematic diagram of the DDPM reverse process at timestep k. In this example, the network will be trained to predict X_{k-1} based on X_k and the conditional input.

In the experimental setup, we set the numbers of root features and the time steps to 64 and 1000, respectively. The SVCT was simulated using FVCT by keeping only the projections corresponding to either high- or low-kVp projections, and the edge maps were generated from FVCT data (see Section II.C for more details). The FVCT were used as the labels. Image patches with a size of 128x128x32 were randomly extracted from training and label images during the training.

C. Edge Map

Despite the variations in beam penetration and contrast inherent to high- and low-kVp settings, the capacity of CT scanners to capture high-frequency information, such as edges, textures, and fine structures, remains relatively consistent across different kVp settings. These high-frequency details, which are indicative of abrupt changes in tissue density, are less influenced by the energy of the X-ray beam. Therefore, to simulate edge maps from kVp-switching scan, we can simply extract edge maps from single-kVp FVCT images rather than actual kVp-switching images. It is worth noting, however, that adjusting the dose level is advisable when simulating low-kVp projections, as these scans may exhibit higher noise levels due to reduced beam penetration.

In this study, we opted to use "high- & low-dose" projections in place of "high- & low-kVp" for the extraction of edge maps. Specifically, projections at various low-dose levels were generated to mimic the noise characteristics of edge maps at low-kVp CT projections by employing the count-domain dose splitting method proposed in the Half2Half technique [5]. These "high- & low-dose" projections were then ordered in the same manner as a high- and low-kVp switching scan. To extract an edge map, we adopted a technique inspired by 'local tomography' for the FBP reconstruction. Specifically, we applied a second-order derivative filter on the radial samples, which was then followed by back-projection. This process effectively produced edge maps. These resulting edge maps were subsequently included as one channel of the conditional input to the DDPM network.

Fig. 2 shows a set of edge maps generated from a pair of 135 kVp and 80 kVp full scans of an abdominal phantom. The edge maps, from left to right, are obtained from the original 135 kVp scan (high dose), low-dose 135 kVp scan, 80 kVp scan, mixture of high- and low-dose 135 kVp scans, and mixture of 135 kV and 80 kV scans. The first four are various edge maps that we can generate during training and the last one is the edge map during deployment.

We can see that the edge map from the mixture of high- and low-dose 135 kVp scans is consistent with the mixed-kVp one, validating our approach of using lower-dose projections to simulate low-kVp for edge map generation. The added noise to the high-kVp projections improved the consistency between the edge maps in simulation with that in real kVp-switching scans.

D. Experimental datasets

We acquired full-view single-kVp helical CT scans from four subjects. Out of these, three scans were used to create training data, while the fourth scan was reserved for testing. Each full-view CT (FVCT) scan comprised 1,200 projections per rotation. The tube current of the FVCT scans ranged from 103 to 461 mAs and the tube voltage was 120 kVp. The high-kVp SVCT subsets were simulated by setting the pixel values of the projections at the transition and opposite kVp positions to 0. The image size for both reconstructed FVCT and SVCT images was 512×512. It should be noted that in this study, no low-kVp SVCT was included for training.

Furthermore, we obtained a pair of full-view scans at 135 kVp and 80 kVp on one subject. The tube currents were 65 mAs for the 135 kVp scan and 330 mAs for the 80kVp scan. The data were used to simulate a real kVp-switching scan by mixing the 135 kVp and 80 kVp projections following the 3-1-3-1 pattern with data in the transition views set to zero. Other acquisition parameters were similar to those described above. SVCT images were obtained using the projections corresponding to either the high- or low-kVp and the edge map was obtained using the full-view mixed-kVp projections.

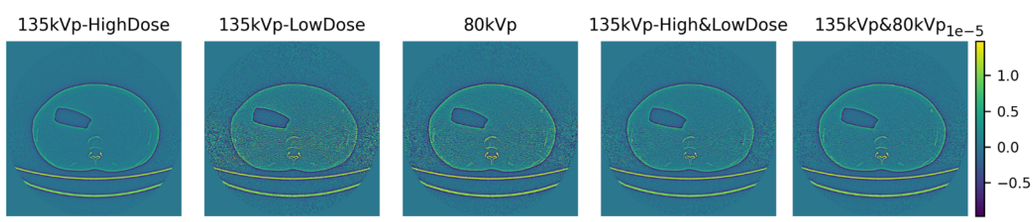


Fig. 2 The edge map comparison. From left to right, the edge map extracted from: (1) high-dose 135 kVp, (2) low-dose 135kVp, (3) 80 kVp, (4) high- & low-dose 135 kVp, and (5) high dose 135 kVp & 80 kVp scans.

E. Metrics for evaluation

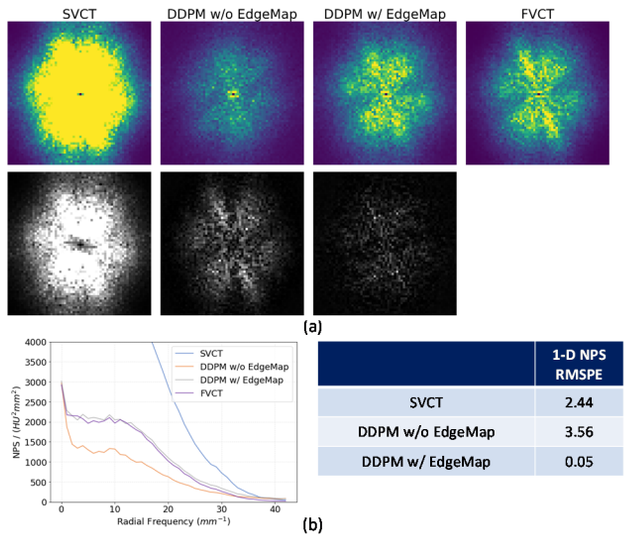
In this study, mean absolute error (MAE), signal-to-noise ratio (SNR), and structural similarity index (SSIM) [6], which are the commonly used metrics, were evaluated to compare different trained networks. In addition, three other metrics were used to further evaluate the effect of texture restoration: (1) histogram correlation (HC), (2) local-binary-pattern-based texture error (LBP-Err) [7], [8], and (3) noise power spectrum (NPS). The definitions of the two texture metrics can be found in our previous study [4].

III. Results

A. Simulated testing dataset

Fig. 3 compares two representative image slices from the simulated testing dataset. The images, from the left to right, are SVCT images, network predictions from DDPM without edge map, network predictions from DDPM with edge map (proposed method), and FVCT references, respectively. The display windows are $W = 1500$ HU/$L = -600$ HU for the lung slice and $W = 400$ HU/$L = 40$ HU for the abdominal slice, respectively. The MAE, SNR, SSIM, HC and LBP-Err of each slice were calculated and listed at the bottom of each image.

It can be observed that both network predictions offer significant improvements over the SVCT images in terms of MAE, SNR, and SSIM, by reducing noise and suppressing streaky artifacts. As shown in the enlarged regions, both predictions sharpened anatomical boundaries in lung regions and preserved the textures of the CT images in abdominal regions. Furthermore, when comparing the two network predictions, the results trained using the edge map restored more texture and preserved fine structures in the regions indicated by the red arrows.



	1-D NPS RMSPE
SVCT	2.44
DDPM w/o EdgeMap	3.56
DDPM w/ EdgeMap	0.05

Fig. 4 NPS comparison of the simulated testing data. (a) 2D NPS and corresponding difference images, (b) 1D NPS and the root mean squared percentage error (RMSPE) of 1D NPS with respect to that of the FVCT.

Fig. 4 shows the NPS computed from a unfirm liver region. Both 2D and 1D NPS further confirmed that the edge maps can guide the network to restore accurate textures in the liver region, matching that of the full view CT reference image.

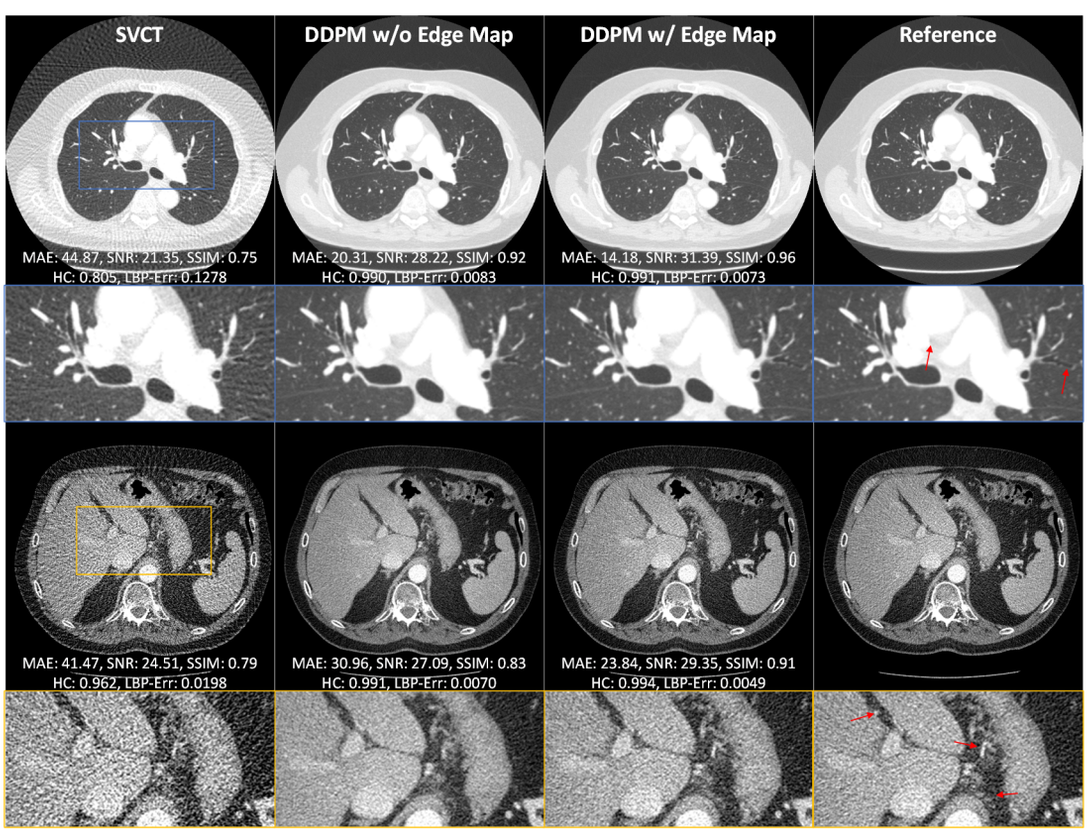


Fig. 3 The image comparison of the simulated data. From left to right, (1) SVCT (input), (2) the prediction from DDPM trained w/o edge map, (3) the prediction from DDPM trained w/ edge map, and (3) FVCT (reference).

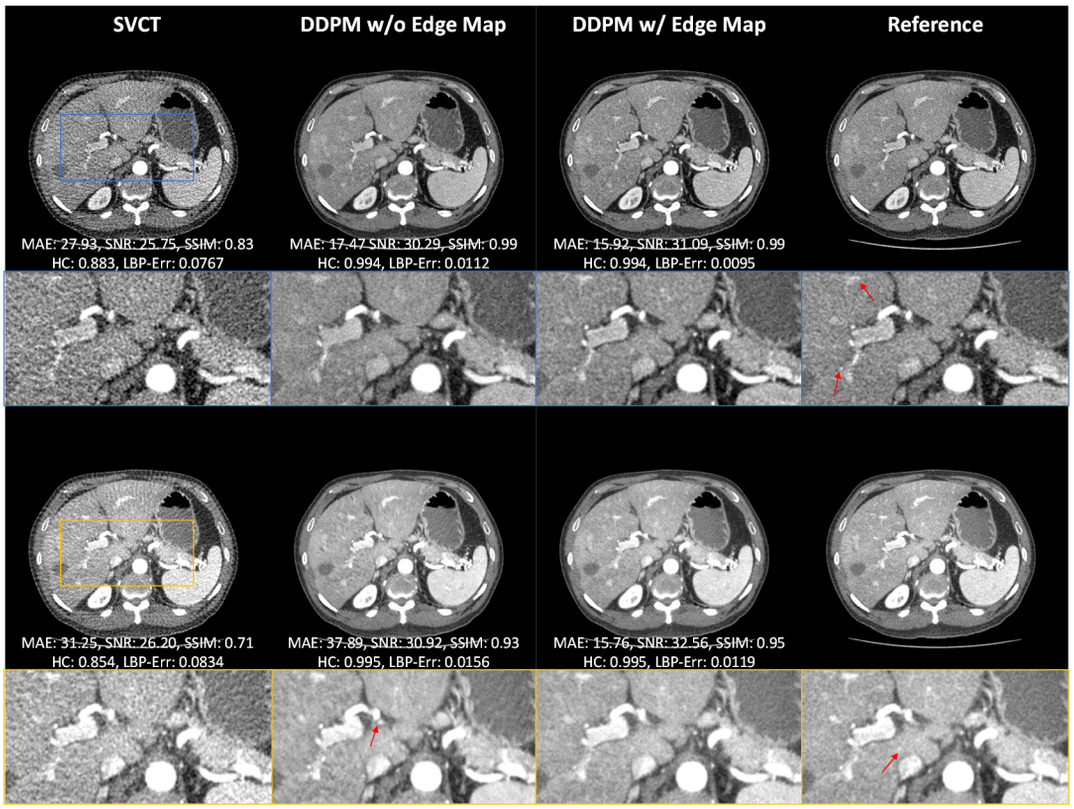


Fig. 5 The image comparison of the real 135 kVp (top row) and 80 kVp (bottom row) data. From left to right, (1) SVCT (input), (2) the prediction from DDPM trained w/o edge map, (3) the prediction from DDPM trained w/ edge map, and (3) FVCT (reference).

B. Real testing dataset

We applied the networks trained on simulation data to the mixed 135 kVp and 80 kVp scan. Fig. 5 shows the network outputs at both energies with comparison to the full-view references. The display windows are W=400 HU/L=40 HU for the 135 kVp slice (top row) and W=600 HU/L=160 HU for the 80 kVp slice (bottom row), respectively. Although the network was trained using 120 kVp data, it still had the capability to restore the SVCT at 135 kVp and 80 kVp energy. These results further confirmed the effectiveness of the proposed method.

IV. Discussion

In this study, we developed an edge prior conditioned DDPM for enhancing kVp-swtching DECT images. The most challenging aspect of this study is how to complement the lacking projections of the SVCT while taking advantage of existing SECT images for network training. The proposed edge map provides a solution to accomplish this goal. We found that we can generate edge maps using single-kVp CT scans that are relatively consistent with those from real mixed-kVp scans. As shown in Fig. 4, although the network without the edge maps was able to enhance the image quality to some extent, it was unable to restore fine structures when they were overshadowed by noise and streaky artifacts. In comparison, the proposed edge map was able to provide the missing structure information that comes from the missing projections while not requiring another kVp scans. The results of this study suggest that edge maps can guide the network to reduce noise and artifacts while restoring fine structures.

V. Conclusion

Our study showed that the proposed edge prior conditioned DDPM is effective in enhancing single-kVp SVCT images from kVp-switching DECT systems. The improvements in fine structure restoration and detail preservation are promising, indicating the potential use of this technology in practice. In the future work, we will include single-kVp CT scans at multiple kVp levels and also quantitatively analyze and evaluate the proposed method on a larger dataset.

References

[1] J. Hsieh, “TU-E-210A-01: Dual-energy CT with fast-kVp switch,” *Med. Phys.*, vol. 36, no. 6Part24, p. 2749, 2009.

[2] I. Goodfellow *et al.*, “Generative adversarial nets,” in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[3] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 6840–6851, 2020.

[4] N. Yuan, J. Zhou, and J. Qi, “Sparse-view CT Spatial Resolution Enhancement via Denoising Diffusion Probabilistic Models,” in *Proceedings of the 17th International Meeting on Fully 3D*, 2023, pp. 470–473.

[5] N. Yuan, J. Zhou, and J. Qi, “Half2Half: deep neural network based CT image denoising without independent reference data,” *Phys. Med. & Biol.*, vol. 65, no. 21, p. 215020, 2020.

[6] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, and others, “Image quality assessment: from error visibility to structural similarity,” *IEEE Trans. image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

[7] M. Sultana, N. Bhatti, S. Javed, and S. K. Jung, “Local binary pattern variants-based adaptive texture features analysis for posed and nonposed facial expression recognition,” *J. Electron. Imaging*, vol. 26, no. 5, p. 53017, 2017.

[8] T. Ojala, M. Pietikainen, and D. Harwood, “Performance evaluation of texture measures with classification based on Kullback discrimination of distributions,” in *Proceedings of 12th international conference on pattern recognition*, 1994, vol. 1, pp. 582–585.

DiffusionX2CT: Zero-Shot 3D CBCT Reconstruction from Sparse-view X-Rays with Warping-based Diffusion Model

Xiaoyue Li, Mark D. Butala, *Member, IEEE* and Gaoang Wang, *Member, IEEE*

***Abstract*—This paper introduces an innovative zero-shot solution, DiffusionX2CT, specifically devised for sparse-view 3D Cone Beam Computed Tomography (CBCT) reconstruction only from sparse X-ray measurements without external training on an extensive set of 3D volumes. The proposed approach harnesses on the advantageous generative capabilities acquired through the pre-training of diffusion models on 2D CT slices, coupled with the incorporation of physical information derived from clinical measurements. Essentially, we propose the method by enhancing the diffusion prior through model-based proximal optimization during the inference process. To exploit semantic similarity effectively among adjacent captured X-rays, our method relies on a novel regularization that integrates information from warping-based views into the conditional denoising process, thereby ensuring that the method can achieve reconstructions of high fidelity and accuracy even in the most challenging sparse-view scenarios. Our proposed method demonstrates remarkable generalization capabilities, capable of constructing 3D CBCT volumes that deviate from the training dataset. Moreover, it exhibits adaptability to diverse measurement scenarios without necessitating retraining.**

***Index Terms*—3D CBCT, sparse View, score-based diffusion model, proximal Optimization, warping-based regularization.**

I. Introduction

CONE beam computed tomography (CBCT) has emerged as a vital imaging technique for the noninvasive examination of internal structures in clinical settings [1]. However, the imperative of capturing a large number of measurements to ensure imaging quality raises concerns about potential excessive radiation exposure. Consequently, the longstanding issue of radiation dose associated with X-rays remains a focal point in CBCT imaging. Prior research endeavors have aimed at mitigating radiation dose in CBCT by reducing the number of measurements while upholding reconstruction quality through analytical methods such as FDK [2], model-based iterative algorithms [3], [4], and deep learning approaches [5], [6]. The former two struggle with ill-posed problems, while the latter often demands extensive datasets for network training, posing challenges due to the scarcity of clinical data. Moreover, many of these approaches require retraining when alterations occur in the measurement process or data distribution, leading to a lack of generalizability and flexibility. Recent attempts at neural rendering techniques for CBCT [7], [8] relying solely on projections of the target object, present object-specific limitations and challenges in extending one neural radiation field to previously unseen volumes.

More recently, diffusion-based models [9]–[12], known for their robustness compared to other generative models, have gained prominence in the biomedical imaging field. However, the majority of diffusion-based methods explored to date have primarily focused on 2D medical imaging scenarios, owing to the high-dimensional nature of generative constraints associated with 3D volumes in training datasets. A recent work, named DiffusionMBIR [12], attempts to address 3D medical imaging using a 2D diffusion-based prior and total variation (TV) constraint. However, this method is exclusively applied to conventional CT imaging and does not explore the potential application to more complex 3D CBCT imaging.

To tackle the challenges mentioned above, we propose DiffusionX2CT, a novel diffusion model for sparse-view 3D CBCT reconstruction that incorporates proximal optimization into diffusion sampling steps. To fully explore the inherent semantic information from adjacent measurements, DiffusionX2CT introduces a novel warping-based regularization by generating synthetic views between the limited set of captured views, which can help enhance the conditional information available during the sampling process. All these views guide the pre-trained 2D diffusion model to independently reconstruct batches of 2D slices in the XY-plane, with the batch dimension corresponding to the Z-plane, and these 2D slices are then concatenated to form the target 3D volume. Geometric consistency based on TV is also employed to further maintain coherent relationships in the Z-plane of the reconstructed volumes. The reconstructed results from different sparse cases demonstrate the generality of DIffusionX2CT and its ability to reconstruct high-fidelity 3D CBCT imaging using prior knowledge from pre-trained 2D diffusion models.

II. Theory and Method

A. Forward Model for CBCT Imaging

In the context of a CBCT imaging system, we consider a standard forward model where $\boldsymbol{x} \in \mathbb{R}^n$ represents the unknown volumes to be reconstructed. The volume is uniformly

This study was supported in part by National Key R&D Program of China (No. 2022ZD0162000) and the National Natural Science Foundation of China Research Fund for International Young Scientists (No. 41850410490).

Xiaoyue Li is with the College of Information Science and Electronic Engineering, Zhejiang University (e-mail: xiaoyue98@zju.edu.cn).

Mark D. Butala is with the Zhejiang University-University of Illinois Urbana-Champaign Institute, and also with the College of Information Science and Electronic Engineering, Zhejiang University (e-mail: markbutala@intl.zju.edu.cn).

Gaoang Wang is with the Zhejiang University-University of Illinois Urbana-Champaign Institute, also with the College of Computer Science and Technology, Zhejiang University (e-mail: gaoangwang@intl.zju.edu.cn).

scanned by N different cameras over a 180° angular range around the object. The measurement y_i ($1 \leq i \leq N$), captured from the angle $\theta_i = (i-1)/N \times 180^\circ$, is given by the equation $y_i = a_i \boldsymbol{x} + \boldsymbol{n}i$, where ai signifies the linear operator related to the ith projection, and n_i accounts for corresponding acquisition noise. The system can be succinctly expressed in matrix form as $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n}$.

To mitigate radiation exposure, only a sparse set of measurements $y_1, \ldots, y_M$ is available, where $M < N$. We define $\mathcal{M}$ as the set corresponding to the captured projections, where y_i ($\theta_i \in \mathcal{M}$) are the known projection images, and y_j ($\theta_j \in \mathcal{N} - \mathcal{M}$) are the unknown ones. A standard approach for addressing this ill-posed inverse problem involves the following regularized reconstruction:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{x} \right\|_2^2 + \boldsymbol{L}(\boldsymbol{x}), \tag{1}$$

where the data fidelity term $\frac{1}{2} \left\| \boldsymbol{y} - \boldsymbol{A}_{\mathcal{M}} \boldsymbol{x} \right\|_2^2$ enforces data consistency, and the prior regularization $\boldsymbol{L}(\boldsymbol{x})$ guides the reconstruction to align with the prior constraint.

From a Bayesian perspective, the prior constraint is conceptualized by more accurately estimating the complex prior data distribution. This allows for achieving higher quality samples, and the samples from the prior distribution can be obtained through the reverse stochastic differential equation (SDE) of a score-based diffusion model [11].

B. Conditional Score-based Diffusion Models

The score-based diffusion model constitutes a generative model that formulates the generative process as the reverse of the data noising process. The diffusion process $\boldsymbol{x}(t)_{t=0}^{T}$ is constructed with $\boldsymbol{x}(t) \in \mathbb{R}^n$, where $t \in [0, T]$ represents the timestep, and n is the image dimension. Here we focus on developing our method based on VE-SDE.

Following the definition from [11], the reverse SDE of the score-based model can be constructed as:

$$d\boldsymbol{x} = -\frac{d\left[\sigma^2(t)\right]}{dt} \underbrace{\nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x})}_{\text{score function}} + \sqrt{\frac{d\left[\sigma^2(t)\right]}{dt}} d\bar{\boldsymbol{w}}, \tag{2}$$

where dt is the infinitesimal negative time step, and $\bar{\boldsymbol{w}}$ is the standard n-dimensional Brownian motion running backwards.

Solving the reverse SDE with the approximated score function (2) involves sampling from the prior distribution $p(\boldsymbol{x})$. The score function can be estimated using a time-conditional neural network $s_{\boldsymbol{\theta}}(\boldsymbol{x}(t), t) \simeq \nabla_{\boldsymbol{x}} \log p_t(\boldsymbol{x}(t))$ [10]. Substituting this approximation into (2) enables solving the reverse SDE. In the context of solving inverse problem, sampling from the posterior distribution $p(\boldsymbol{x} \mid \boldsymbol{y})$ is paramount.The relationship between the prior and posterior is governed by Bayes' rule $p(\boldsymbol{x} \mid \boldsymbol{y}) = p(\boldsymbol{x}) p(\boldsymbol{y} \mid \boldsymbol{x}) / p(\boldsymbol{y})$, leading to

$$\nabla_{\boldsymbol{x}} \log p(\boldsymbol{x} \mid \boldsymbol{y}) = \nabla_{\boldsymbol{x}} \log p(\boldsymbol{x}) + \nabla_{\boldsymbol{x}} \log p(\boldsymbol{y} \mid \boldsymbol{x}). \tag{3}$$

Hence the likelihood term serves to enforce both the prior data distribution and the data consistency that satisfies $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$.

1) Flow-based Novel View Augmentation: In contrast to the previous diffusion-based methods for CT imaging [12], [13] which are only guided by the known measurements and ignore the implicit relationship between the known measurement sets and the adjacent unknown ones, DiffusionX2CT introduces a new regularization to incorporate the contribution of the $N - M$ non-captured views. To estimate the novel views $\boldsymbol{y}_j$ from the known view geometry, we interpolate them from its two neighboring captured views $\boldsymbol{y}_i$ and $\boldsymbol{y}_k$, given $\theta_i < \theta_j < \theta_k$. It worth noting that $\theta_j \in \mathcal{N} - \mathcal{M}$ and $\theta_i, \phi_k \in \mathcal{M}$.

We initiate the process by computing the optical flow $\boldsymbol{v}_{i \to k}$ from the acquired image pair $\boldsymbol{y}_i$ and $\boldsymbol{y}_k$. Utilizing the angle indexing ratio $s = (j-i)/(k-i)$, we obtain two intermediate images $\widetilde{\boldsymbol{y}}_i$ and $\widetilde{\boldsymbol{y}}_k$ by warping $\boldsymbol{y}_i$ and $\boldsymbol{y}_k$, respectively:

$$\widetilde{\boldsymbol{y}}_i = \boldsymbol{wa}\left(\boldsymbol{y}_i, s\boldsymbol{v}_{i \to k}\right), \tag{4}$$

$$\widetilde{\boldsymbol{y}}_k = \boldsymbol{wa}\left(\boldsymbol{y}_k, -(1-s)\boldsymbol{v}_{i \to k}\right). \tag{5}$$

The unknown view $\boldsymbol{y}_j$ is then interpolated using the warping-based images $\widetilde{\boldsymbol{y}}_i$ and $\widetilde{\boldsymbol{y}}_k$, as follows:

$$\boldsymbol{y}_j = (1-s)\widetilde{\boldsymbol{y}}_i + s\widetilde{\boldsymbol{y}}_k, \tag{6}$$

and all $N - M$ missing projections are interpolated from the captured views.

We term the new regularization related to the novel projections as $\mathcal{L}_{\text{novel}}$. As the accuracy of the projection estimation module is highly dependent on the angular distance to the captured views, we enhance the optimization framework by introducing a confidence estimation weight ($0 < \omega_j < 1$), as follows:

$$\mathcal{L}_{\text{novel}} = \sum_{j \mid \phi_j \in \mathcal{N} - \mathcal{M}} \omega_j \left\| \boldsymbol{y}_j - \boldsymbol{a}_j \boldsymbol{x} \right\|_2^2, \tag{7}$$

$$\to \frac{1}{2} \left\| \boldsymbol{y}_{\text{novel}} - \boldsymbol{A}'_{\mathcal{N} - \mathcal{M}} \boldsymbol{x}) \right\|_2^2. \tag{8}$$

The confidence weight ω_j ensures that the closer the angles are to the captured views, the higher the confidence in the estimation. The maximum weight is 1.

2) Geometric Consistency Constraint: To optimize the utilization of the diffusion models for 3D reconstruction, we implement 2D diffusion models slice by slice basis in the XY-plane. As each denoising step disregards correlations between slices, the reconstructed slices, especially in cases of extreme sparsity, lack coherence with each other. Consequently, images acquired from coronal or sagittal directions exhibit noticeable artifacts.

In this framework, we perform the estimation of prior distribution in the first term of (3) in 2D slice-by-slice, Subsequently, the 2D prior samples from the target volume are concatenated in sequence and guided by the captured sets and novel sets of projections, along with corresponding linear operators using Alternating Direction Method of Multipliers (ADMM) and conjugate gradient (CG) in the 3D domain. To further mitigate artifacts, we follow [12] to set $\boldsymbol{L}(\boldsymbol{x})$ term

as the ℓ_1 norm of the difference along Z-plane, denoted as $\|\boldsymbol{L}_{\boldsymbol{z}}\boldsymbol{x}\|_1$. The total optimization is replaced as:

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \frac{1}{2}\left\|\boldsymbol{y}_{\text{captured}} - \boldsymbol{A}_{\mathcal{M}}\boldsymbol{x})\right\|_2^2 + \arg\min_{\boldsymbol{x}} \frac{1}{2}\left\|\boldsymbol{y}_{\text{novel}} - \boldsymbol{A}'_{\mathcal{N}-\mathcal{M}}\boldsymbol{x})\right\|_2^2 + \|\boldsymbol{L}_{\boldsymbol{z}}\boldsymbol{x}\|_1 . \tag{9}$$

Here we apply the predictor-corrector (PC) sampling algorithm [11] and extend the sampling scheme to a proximal optimization, such that to replace the reverse process as a conditional sampling algorithm guided by X-rays and novel views by adjacent measurements. The detailed algorithm for the inference process of DiffusionX2CT can be found in Algorithm 1. Here we denote ρ and α as the hyper-paremeters, and $\mathcal{D}$ represents the first term of (3), and $\mathcal{S}$ is the soft thresholding operator. $\mathcal{CG}\ (\boldsymbol{x}, \boldsymbol{r}, \rho, \boldsymbol{A}_{\mathcal{M}}, \boldsymbol{A}_{\mathcal{N}-\mathcal{M}}, O)$ means running O iterations of CG with initial point $\boldsymbol{x}$.

Algorithm 1 The inference process of DiffusionX2CT

Require: $s_\theta, N, \rho, \alpha, \{\sigma_i\}$
$\boldsymbol{x}_N \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_T^2\boldsymbol{I}\right)$
$\boldsymbol{z} \leftarrow \boldsymbol{0}$
$\boldsymbol{w}_1 \leftarrow \boldsymbol{0}$
$\boldsymbol{w}_2 \leftarrow \boldsymbol{0}$
for iteration $= 1$ to max_steps **do**
  $\boldsymbol{x} \leftarrow \mathcal{D}(\boldsymbol{x}, s_\theta)$
  $\boldsymbol{r} \leftarrow \boldsymbol{A}_{\mathcal{M}}^T(\boldsymbol{A}_{\mathcal{M}}\boldsymbol{x} - \boldsymbol{y}_{\text{captured}}) + \boldsymbol{A}_{\mathcal{N}-\mathcal{M}}^T(\boldsymbol{A}_{\mathcal{N}-\mathcal{M}}\boldsymbol{x} - \boldsymbol{y}_{\text{novel}}) + \rho L^T(L\boldsymbol{x} - \boldsymbol{z} + \frac{\boldsymbol{w}_1}{\rho} + \frac{\boldsymbol{w}_2}{\rho})$
  $\boldsymbol{x} \leftarrow \mathcal{CG}(\boldsymbol{x}, \boldsymbol{r}, \rho, \boldsymbol{A}_{\mathcal{M}}, \boldsymbol{A}_{\mathcal{N}-\mathcal{M}}, \boldsymbol{L}, 1)$
  $\boldsymbol{z} \leftarrow \mathcal{S}(\boldsymbol{L}\boldsymbol{x} + \frac{\boldsymbol{w}_2}{\rho}, \frac{\alpha}{\rho})$
  $\boldsymbol{w}_1 \leftarrow \boldsymbol{w}_1 + \rho(\boldsymbol{A}_{\mathcal{M}}\boldsymbol{x} - \boldsymbol{y}_{\text{captured}} + \boldsymbol{x} - \boldsymbol{z})$
  $\boldsymbol{w}_2 \leftarrow \boldsymbol{w}_2 + \rho(\boldsymbol{A}'_{\mathcal{N}-\mathcal{M}}\boldsymbol{x} - \boldsymbol{y}_{\text{novel}} + \boldsymbol{x})$
end for
function $\mathcal{S}(\boldsymbol{v}, \boldsymbol{h})$
  return $\max(0, 1 - \frac{\boldsymbol{h}}{\max(|\boldsymbol{v}|, 1e-10)}) \cdot \boldsymbol{v}$
end function

III. Experiments

A. Implementation Details

We hold out three volumes from the publicly available LIDC-IDRI dataset [14] for testing. Note that we use a pre-trained score function from [13] that was trained on AAPM 2016 CT images only, and hence we need not split the train/test data here. The whole steps of our PC sampler is 2000.

B. Baselines

To assess the effectiveness of our method, we conduct a comparative analysis with five baseline techniques. FDK [2] is chosen as a representative of analytical methods. OS-SART [3], is a modified iterative reconstruction algorithm. ASD-POCS [4] represents another iterative method with a TV regularization. We also compare with concurrent neural volume rendering methods, Neural Attenuation Field (NAF) [8], specifically designed for CBCT. We additionally compare our method with Chung et al. [13]. In response to memory space constraints, we implement a modification inspired by [12]. This involves replacing the manifold constrained gradient (MCG) with the use of the algebraic reconstruction technique (ART) in each data-consistency step.

For ablation purposes, we present two variants of DiffusionX2CT by removing the regularization for view augmentation and geometric consistency, denoted as DiffusionX2CT (w/o novel view aug.) and DiffusionX2CT (w/o geometric consist.), respectively. The experiments are conducted with different numbers of input views (i.e., 10, 20, and 30 views). Notably, the 10-view scenario is extremely sparse, representing approximately 50 times fewer projections than typical clinical settings.

IV. Numerical Results

The qualitative results are presented in Fig. 1, demonstrating the superior performance of our proposed method compared to alternative approaches. To facilitate a comprehensive comparison among different methods, ROIs have been magnified at the bottom of each image. Specifically, FDK, known for its efficacy with an ample number of views, encounters challenges in delivering satisfactory results under the extremely sparse conditions presented here. OS-SART shows improvement over analytical methods in sparse view cases, but it compromises on reconstruction details, exhibiting an over-smoothing issue, particularly affecting tracheal structures. ASD-POCS exhibits enhanced edge preservation but still falls short in depicting intricate structures with clarity. Regarding the latest neural rendering methods, NAF demonstrates reduced artifacts compared to other counterparts but struggles with the accurate reconstruction of local intensity contrast. MCG performs well in reconstructing the axial slices but lacks spatial integrity across the Z-plane, leading to sharp artifacts that are visible clearly in both coronal and sagittal slices. In contrast, our method consistently achieves the best results with more details and less noise for all configurations.

We quantize the quality of the reconstructed 3D CBCTs from different methods using PSNR and SSIM metrics with different view configurations. By comparison, DiffusionX2CT (w/o novel view aug.) and DiffusionX2CT (w/o geometric consist.) already surpass all other methods, providing high-quality visual results, while DiffusionX2CT (ours) with novel view augmentation and geometric consistency constraint further outscores the baselines by large margins in most of the settings. Especially, our method with only 10 views outperforms all the others with 30 views.

V. Conclusions

In this paper, a diffusion model for 3D CBCT from insufficient 2D X-rays is devised, effectively addressing the ill-posed CBCT imaging by incorporating the proximal optimization into each reverse step of the 2D pre-trained score-based model. We introduce a new measurement augmentation warping-based regularization, which jointly optimizes with TV prior to deal with the great challenges introduced by extremely sparse views. This novel method offered more generality and flexibility in different measurement cases compared to other comparing techniques respectively. For further work, we aim

TABLE I
PERFORMANCE EVALUATION OF THE CBCT RECONSTRUCTION TASK IN DIFFERENT PLANES. **BOLDFACE** INDICATES THE BEST RESULTS.

	10-view						20-view						30-view					
	Axial		Coronal		Sagittal		Axial		Coronal		Sagittal		Axial		Coronal		Sagittal	
Method	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FDK	10.02	0.21	12.54	0.36	10.01	0.22	12.12	0.27	14.22	0.43	12.07	0.42	14.23	0.46	15.79	0.50	14.20	0.46
OS-SART	10.23	0.24	10.12	0.22	10.10	0.21	12.70	0.33	12.03	0.46	12.10	0.47	14.67	0.49	14.54	0.47	14.46	0.45
ASD-POCS	13.45	0.54	12.34	0.49	12.96	0.53	15.52	0.66	13.63	0.57	14.53	0.65	17.03	0.72	15.71	0.68	15.97	0.69
NAF	15.21	0.62	14.97	0.56	15.05	0.60	16.93	0.64	15.01	0.61	15.26	0.63	19.27	0.79	18.43	0.74	18.62	0.74
MCG	28.92	0.83	25.46	0.72	25.11	0.62	29.79	0.87	26.35	0.69	26.56	0.70	30.52	0.89	27.45	0.82	27.30	0.81
DiffusionX2CT (w/o novel view aug.)	28.96	0.84	26.37	0.78	26.32	0.76	29.82	0.88	28.35	0.84	28.69	0.86	30.66	0.91	29.07	0.86	29.04	0.86
DiffusionX2CT (w/o geometric consist.)	33.65	0.94	26.26	0.72	25.32	0.74	34.54	0.94	27.85	0.81	27.68	0.80	34.70	0.95	28.59	0.84	28.51	0.83
DiffusionX2CT (Ours)	**33.69**	**0.94**	**34.75**	**0.95**	**31.58**	**0.92**	**34.76**	**0.95**	**35.28**	**0.96**	**33.19**	**0.93**	**35.59**	**0.96**	**34.87**	**0.94**	**34.58**	**0.94**

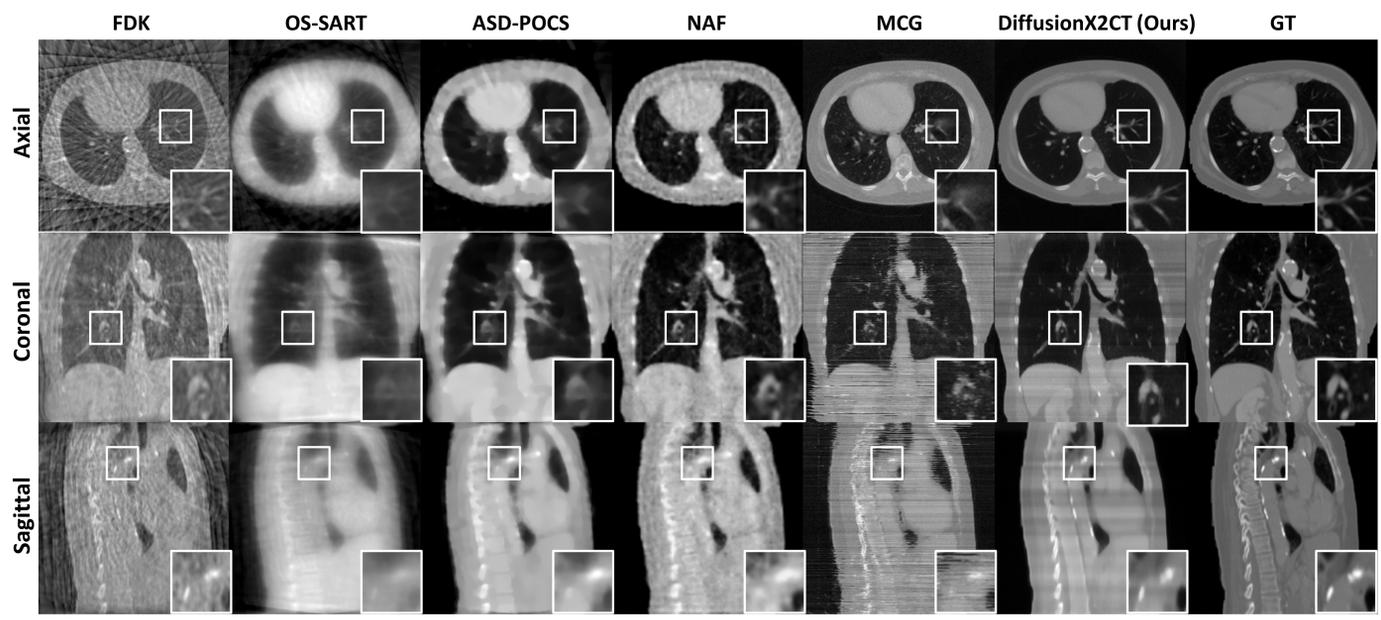


Fig. 1. Illustration of the 10-view reconstructed results of our proposed framework. (First row: axial slice, second row: sagittal slice, third row: coronal slice) We also visualize the results from FDK, OS-SART, ASD-POCS, NAF, and MCG. The last column shows the ground truth of the target volume.

to extend the proposed method to 4D (time-involved) cases and explore physically motivated view synthesis to achieve even better reconstruction results for sparse-view tomography problems.

REFERENCES

[1] W. C. Scarfe, A. G. Farman, P. Sukovic *et al.*, "Clinical applications of cone-beam computed tomography in dental practice," *Journal-Canadian Dental Association*, vol. 72, no. 1, p. 75, 2006.

[2] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Josa a*, vol. 1, no. 6, pp. 612–619, 1984.

[3] G. Wang and M. Jiang, "Ordered-subset simultaneous algebraic reconstruction techniques (os-sart)," *Journal of X-ray Science and Technology*, vol. 12, no. 3, pp. 169–177, 2004.

[4] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, p. 4777, 2008.

[5] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng, "X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10 619–10 628.

[6] G. Wang, J. C. Ye, and B. De Man, "Deep learning for tomographic image reconstruction," *Nature Machine Intelligence*, vol. 2, no. 12, pp. 737–748, 2020.

[7] 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.

[8] R. Zha, Y. Zhang, and H. 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.

[9] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in neural information processing systems*, vol. 32, 2019.

[10] Y. Song, S. Garg, J. Shi, and S. Ermon, "Sliced score matching: A scalable approach to density and score estimation," in *Uncertainty in Artificial Intelligence*. PMLR, 2020, pp. 574–584.

[11] 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. [Online]. Available: https://openreview.net/forum?id=PxTIG12RRHS

[12] H. Chung, D. Ryu, M. T. McCann, M. L. Klasky, and J. C. Ye, "Solving 3d inverse problems using pre-trained 2d diffusion models," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 22 542–22 551.

[13] H. Chung, B. Sim, D. Ryu, and J. C. Ye, "Improving diffusion models for inverse problems using manifold constraints," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25 683–25 696, 2022.

[14] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman *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*, vol. 38, no. 2, pp. 915–931, 2011.

Prior Frequency Guided Diffusion Model for Limited Angle (LA)-CBCT Reconstruction

Jiacheng Xie, Hua-Chieh Shao, Yunxiang Li, You Zhang

***Abstract*—Cone-beam computed tomography (CBCT) is widely used in image-guided radiotherapy. Reconstructing CBCTs from limited-angle acquisitions (LA-CBCT) is highly desired for improved imaging efficiency, dose reduction, and better mechanical clearance. LA-CBCT reconstruction, however, suffers from severe under-sampling artifacts, making it a highly ill-posed inverse problem. Diffusion models are known for generating high-quality, diverse, and realistic images as a new class of generative models. We proposed a diffusion model-based framework, prior frequency guided diffusion model (PFGDM), to solve the LA-CBCT reconstruction problem. PFGDM uses high-frequency information from patient's prior computed tomography (CT) scans as anatomical detail preservation conditions for a diffusion model-based denoiser, which was incorporated into an iterative reconstruction framework based on the alternating direction method of multipliers (ADMM). We compared PFGDM with current solutions and showed that it achieves the best performance under different limited angle scenarios, especially for very small angles that are challenging for other methods (e.g., 30°).**

***Index Terms*—CBCT, Diffusion model, Deep learning, Limited angle, Reconstruction.**

I. Introduction

In recent years, cone-beam computed tomography (CBCT) has been widely applied to fields like radiotherapy [1] for its efficient 3D imaging capability. Limited-angle CBCT (LA-CBCT), which allows faster imaging speed, lower X-ray dose, and better mechanical clearance, has been investigated extensively. Conventional reconstruction methods like the Feldkamp-Davis-Kress (FDK) algorithm produce significant artifacts and distortions in the reconstructed images given the limited-angle sampling, which prohibits the wide clinical adoption of LA-CBCT.

With the recent success of deep learning (DL), several works have been proposed for limited-angle/view image reconstruction. Preprocessing methods used frameworks like generative adversarial networks (GANs) to predict full-angle/view projection data from limited-angle/view acquisitions for fully sampled reconstruction [2]. Postprocessing methods employed convolutional neural networks (CNNs) to improve image quality by directly mapping the degraded reconstructions to corresponding high-quality versions [3]. Compared with these methods, methods that combine neural networks with iterative physics-informed reconstruction models can often achieve better results and are more robust to imaging condition variations [4, 5]. These hybrid methods effectively leverage the power of neural networks in feeding prior information and serving as denoisers, and the proven records of physics-driven iterative models in solving under-sampling inverse problems. Recently, diffusion models were introduced into the inverse problems of image generation/reconstruction, and shown to outperform popular generative models such as GAN [6]. Diffusion models operate through forward and backward diffusion steps [7]. In the forward process, a diffusion model gradually adds noise to an image, transforming the data into a Gaussian noise distribution. Starting from the noise, the model then iteratively reduces the noise, through a reverse diffusion process which is trained to reconstruct the original image or generate new samples from the learned data distribution. In terms of limited-angle CT (LA-CT) reconstruction, DOLCE [8] and DiffusionMBIR [9] are two latest diffusion model-based methods that showed superior performance over other DL-based methods. By embedding the diffusion model as a denoiser in an ADMM iterative reconstruction framework, DiffusionMBIR uses the prior distribution learned in the diffusion model to improve the LA-CT image quality. However, both the DOLCE and the DiffusionMBIR were developed for CT reconstruction using parallel beam geometries. A more clinically realistic cone-beam imaging geometry can be more challenging for both methods. In addition, although both methods showed improved results in limited-angle reconstruction, it is still challenging to reconstruct high-quality images using very small gantry angles (*i.e.,* $\theta \leq 90°$).

To address the above challenges, we propose a prior frequency-guided diffusion model (PFGDM) for LA-CBCT reconstruction. PFGDM is built upon the findings of our recent publication on zero-shot image translation [10], which observed that a diffusion model conditioned on image-domain high-frequency information can better retain the anatomical details within the images. For source and target domains with the same or similar underlying anatomy (*i.e., CBCT and CT images of the same patient*), the high-frequency information from the source image domain can be applied as the *prior* to condition the diffusion steps for structural preservation during source-to-target domain translation. For radiotherapy patients, there will always be high-quality CT images of the same patients acquired for treatment simulation/planning before the daily CBCT acquisition. Such CT information can serve as a strong anatomical *prior* to condition the reconstruction of daily LA-CBCTs. Inspired by these observations, we developed PFGDM for LA-CBCT reconstruction, which is conditioned on high-frequency information from patient-specific prior CT scans. However, although the anatomical structures are expected to be similar between CT/CBCT, they may not be the same due to daily anatomical variations. To allow the reconstruction of CBCT anatomical structures that are different from those of CT,

The study was supported by the US National Institutes of Health (R01 CA240808, R01 CA258987, and R01 CA280135).
J. Xie, H.-C. Shao, Y. Li, and Y. Zhang are with the Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX 75390, USA.

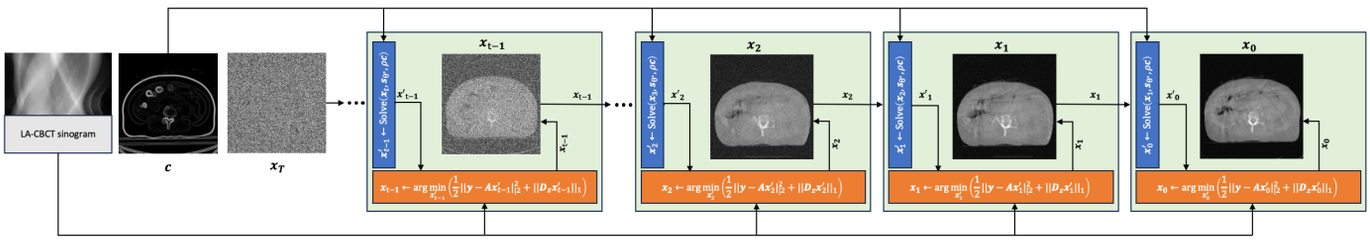


Fig. 1. Overview of PFGDM. Starting from the Gaussian noise $\boldsymbol{x}_T$, in each SDE step, we sample an image $\boldsymbol{x}_t$ from the posterior by solving the conditional diffusion and updating it via ADMM based on the given LA-CBCT sinogram, until we reach a high-quality reconstruction.

PFGDM also introduced a condition-dropping mechanism that allows the reconstruction algorithm to drop the prior CT condition after preset steps, and then use the available projections to reconstruct the anatomical differences.

We evaluated PFGDM and compared it with the other methods for LA-CBCT reconstructions across a range of limited angles ($\theta \in 30°, 90°, 120°$) and three anatomical sites (lung, pelvis, and head and neck). Trained as a single model, PFGDM achieves the best performance across all scenarios, making it a highly promising tool for clinical applications.

II. Materials And Methods

II.A. Background

Inverse problems. Reconstructing LA-CBCT can be formulated as a linear inverse problem:

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{n} \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the image to be reconstructed, $\boldsymbol{y} \in \mathbb{R}^m$ is the measurement (i.e. sinogram), $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is the measurement system matrix, and $\boldsymbol{n}$ is the measurement noise in the system. Reconstructing $\boldsymbol{x}$ from $\boldsymbol{y}$ is highly ill-posed in LA-CBCT, thus additional assumptions on $\boldsymbol{x}$ are often required (adding a denoiser, regularizer, etc).

Score-based diffusion models. Score-based diffusion models [11] are a type of generative models that simulate the generation of data/image by reversing a data-noising process, which can be incorporated as a denoiser in inverse problems. Consider a stochastic process with the time variable $t \in [0,1]$, the noise-corrupted data/image $\boldsymbol{x}$ can be represented as $\boldsymbol{x}(t) = \boldsymbol{x}_t$. Suppose the data is sampled from a distribution $p(\boldsymbol{x})$ for $\boldsymbol{x}(0)$, as time progresses ($t = 1$), the data distribution $\boldsymbol{x}(1)$ approaches a Gaussian distribution. The evolution of the data over time is described by a Stochastic Differential Equation (SDE):

$$d\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{x}, t)dt + g(t)d\boldsymbol{w} \tag{2}$$

where $\boldsymbol{f}(\boldsymbol{x}, t)\colon \mathbb{R}^{n\times 1} \longmapsto \mathbb{R}^n$ is the drift function, $g(t)\colon \mathbb{R} \longmapsto \mathbb{R}$ is the scalar diffusion function, and $\boldsymbol{w}$ is the n- dimensional standard Brownian motion. To solve the reverse-time SDE that defines the generative process of the diffusion model, we train the neural network through Denoising Score Matching (DSM) [11, 12]:

$$\min_{\theta} E_{t,\boldsymbol{x}(t)}\left[\lambda(t)||\boldsymbol{s}_\theta(\boldsymbol{x}(t), t) - \nabla_{\boldsymbol{x}_t}\log p\left(\boldsymbol{x}(t)\middle|\boldsymbol{x}(0)\right)||_2^2\right] \tag{3}$$

where $\nabla_{\boldsymbol{x}_t}\log p\left(\boldsymbol{x}_t\right)$ is the score function, representing the gradient of the log probability with respect to the data; $\boldsymbol{s}_\theta(\boldsymbol{x}(t), t)\colon \mathbb{R}^{n\times 1} \longmapsto \mathbb{R}^n$ is a time-dependent neural network and $\lambda(t)$ is the weighting function. Ultimately, given this robust training schema [9], the reverse diffusion process is approximated by:

$$d\boldsymbol{x} \simeq -\frac{d[\sigma^2(t)]}{dt}\boldsymbol{s}_{\theta^*}(\boldsymbol{x}(t), t)dt + \sqrt{\frac{d[\sigma^2(t)]}{dt}}d\bar{\boldsymbol{w}} \tag{4}$$

To solve Eq. 4, one can employ a method such as the predictor-corrector (PC) sampler, which involves discretizing the time interval from 0 to 1 into N equal parts. This allows for the numerical integration of the reverse diffusion process to generate samples from the target distribution.

DiffusionMBIR. DiffusionMBIR allows 3-D reconstruction by applying 2-D diffusion models on a slice-by-slice basis along the z-axis, and uses additional regularization along the z-axis to leverage the inter-dependency among slices [9]. DiffusionMBIR incorporates the diffusion models into an iterative 3-D volume update step using the Alternating Direction Method of Multipliers (ADMM). The specific objective function is formulated as:

$$\min_{\boldsymbol{x}} \frac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2 + ||D_z\boldsymbol{x}||_1 \tag{5}$$

Where $||D_z\boldsymbol{x}||_1$ represents the L1 norm of the finite differences along the z-axis only. Given Eq. 4, DiffusionMBIR is solved as [9]:

$$\boldsymbol{x}'_{i-1} \leftarrow \text{Solve}(\boldsymbol{x}_i, \boldsymbol{s}_{\theta^*}) \tag{6}$$

$$\boldsymbol{x}_{i-1} \leftarrow \arg\min_{\boldsymbol{x}'_{i-1}}\left(\frac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}'_{i-1}|_{2}^2 + ||\boldsymbol{D}_z\boldsymbol{x}'_{i-1}||_1\right) \tag{7}$$

where *Solve* denotes the reverse-SDE solver of Eq. 4.

High-frequency information extraction. To extract the high-frequency information from prior CT images, we used the Sobel operator [13], a discrete differential operator. Specifically, we convolved the CT image CT_{prior} with horizontal and vertical Sobel filter kernels, denoted as K_x and K_y, respectively, resulting in two filtered images, G_x and G_y. The high-frequency information i_{mag} is then extracted as:

$$G_x = K_x * CT_{prior}, \quad G_y = K_y * CT_{prior} \tag{8}$$

$$i_{\text{mag}} = \sqrt{G_x^2 + G_y^2} \tag{9}$$

II.B. The proposed method: PFGDM

An overview of the PFGDM method is provided in **Fig. 1**. Our goal is to reconstruct CBCTs from LA-CBCT projections with the guidance of a conditional diffusion model, where the condition $\boldsymbol{c}$ is obtained by applying the Sobel operator on the prior CT image (Eqs. 8 and 9). Solving the inverse problem with diffusion involves solving the reverse SDE with the approximated score function defined in Eq. 4, which can be viewed as sampling from the conditional prior distribution $p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{c})$. Thus, the solution of PFGDM can be formulated as:

$$\boldsymbol{x}'_{i-1} \leftarrow \text{Solve}(\boldsymbol{x}_i, \boldsymbol{s}_{\theta^*}, \rho\boldsymbol{c}) \tag{10}$$

$$\boldsymbol{x}_{i-1} \leftarrow \arg\min_{\boldsymbol{x}'_{i-1}}\left(\frac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}'_{i-1}|_{2}^2 + ||\boldsymbol{D}_z\boldsymbol{x}'_{i-1}||_1\right) \tag{11}$$

Specifically, Eq. 10 equals to denoising each slice in parallel with the trained score-based diffusion model under the conditional input, whereas Eq. 11 applies the z-directional TV prior and enforces data fidelity as in DiffusionMBIR. Notably, ρ is a hyper-parameter that determines whether the conditional input $\boldsymbol{c}$ exists or not, which controls the condition-dropping mechanism.

II.C. Experimental setup

We trained the conditional score-based model on an in-house dataset which consists of ~3000 CBCT slices from 30 patients of three anatomical sites: head and neck, lung, and pelvis. In the training process, the Sobel-filtered slices are fed in as optional conditional input to guide the diffusion model. Specifically, during the diffusion model training, we used a random number generator to guide random introduction of this extracted high-frequency information with a probability of 0.5. In this way, the model is trained to fit scenarios with and without the conditional input, allowing us to drop the condition when appropriate to reconstruct CT/CBCT differences. For testing, we used CT and CBCT scans from 6 different patients of the three anatomical sites (2 patients for each site). The CT scans were rigidly registered to the corresponding CBCT images using the open-source software package Elastix. Cone-beam projection sets of different limited-angle acquisition scenarios were simulated for reconstruction and comparison between five different methods (FDK, ADMM-TV, DOLCE, DiffusionMBIR, and PFGDM). We adopted DOLCE[8] and DiffusionMBIR[9] in this study by modifying the projection geometry from parallel beam to cone beam. These two are representative, recently-developed methods based on diffusion models, and are considered state-of-the-art in our study. We evaluated limited-angle acquisitions from a single direction (single-view) or orthogonal directions (orthogonal-view), for the latter has shown better sampling efficiency under the same total scan angle as in previous works[14].

III. Results

III.A. Hyper-parameter Analysis

As aforementioned, in Eq. 10, hyper-parameter ρ controls the presence of the conditional input during reconstruction. With ρ, we can choose to stop introducing the high-frequency information after a preset number of steps during reconstruction, to allow differing CT/CBCT anatomies to be better reconstructed. **Fig. 2** shows reconstructed LA-CBCTs (second row), with the condition extracted from the prior CT dropped at various iteration steps. It can be seen that dropping the condition at an intermediate step (1400) allows the LA-CBCT to inherit similar anatomical features from the prior CT (1400 vs. 500), while better correcting the anatomical mismatches (1400 vs. 2500).

To quantitatively determine the optimal step of dropping the high-frequency information, we performed a parameterized study using steps ranging from 500 to 2500. As shown in **Fig. 3**, the optimal condition-dropping step is around 1400, which gives the best SSIM and PSNR scores. All reconstructions performed in the following study for PFGDM are based on this step number.

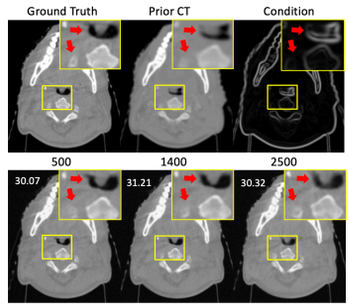


Fig. 2. Comparison between reconstructed LA-CBCTs, with the high-frequency prior CT condition dropped at different iteration steps (500, 1400, 2500). The reconstructions are based on an orthogonal-view 90º scan angle (45º for each view), using a total of 3000 iteration steps. PSNR scores measured against the 'ground-truth' fully-sampled CBCT are indicated at the top-left corner.

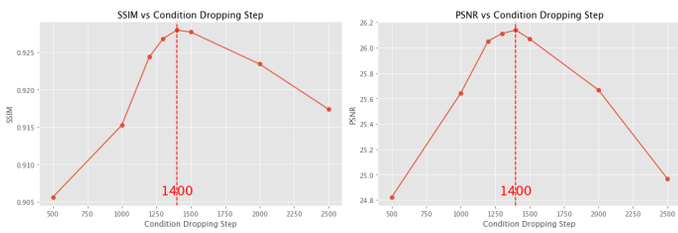


Fig. 3. Variations of the reconstructed LA-CBCT quality given different condition-dropping steps, with a total reconstruction iteration number of 3000.

III.B. Qualitative Evaluation

We visually compared the reconstructed LA-CBCT images between FBP, ADMM-TV, DOLCE, DiffusionMBIR, and PFGDM in **Fig. 4.** Different rows show the results of different anatomical sites and scan angles. The LA-CBCTs reconstructed by the conventional FDK algorithm were compromised by strong under-sampling artifacts, while the iterative ADMM algorithm with total variation regularization (ADMM-TV) manages to substantially suppress these artifacts but unable to reconstruct the anatomical details, especially for very small scan angles (30º). With diffusion models introduced as denoisers, DiffusionMBIR and DOLCE outperformed ADMM-TV, while the reconstructed anatomical details show clear mismatches with the 'ground-truth', especially for the 30º scan angle. In contrast, with the high-frequency prior CT condition and the condition-dropping mechanism, PFGDM reconstructs the anatomical structures to match well with the 'ground-truth', even for the 30º scan angle scenario. Its reconstruction results using single-view projections are generally better than DiffusionMBIR and DOLCE, even if DiffusionMBIR and DOLCE used the more efficient orthogonal-view sampling.

III.C. Quantitative Evaluation

TABLE I. Average (±s.d.) PSNR and SSIM results in axial slices on the test set by different methods. The arrows are pointing to the direction of improved accuracy.

Metric	**PSNR↑**			**SSIM↑**		
/Angle	**30°**	**90°**	**120°**	**30°**	**90°**	**120°**
FDK(Ortho)	12.66±1.12	13.96±1.06	14.83±1.74	0.304±0.115	0.431±0.201	0.511±0.201
ADMM-TV(Ortho)	12.86±1.74	14.18±2.99	15.59±3.87	0.533±0.072	0.688±0.044	0.773±0.068
DOLCE(Ortho)	18.57±2.53	22.74±3.82	25.51±3.03	0.776±0.068	0.868±0.096	0.896±0.106
DiffusionMBIR(Ortho)	19.61±2.47	24.58±4.18	26.18±4.89	0.807±0.048	0.912±0.068	0.926±0.067
PFGDM(Single)	22.21±1.16	25.80±1.28	27.77±1.42	0.855±0.055	0.929±0.021	0.949±0.013
PFGDM (Ortho)	**23.81±2.25**	**26.63±2.79**	**27.97±3.10**	**0.896±0.036**	**0.937±0.029**	**0.949±0.027**

Table I quantitatively compares different methods on their reconstruction accuracy. Most methods used the orthogonal-view scheme (ortho), while PFGDM used both single-view (single) and orthogonal-view (ortho) sampling scenarios. As shown, PFGDM consistently outperforms other methods under

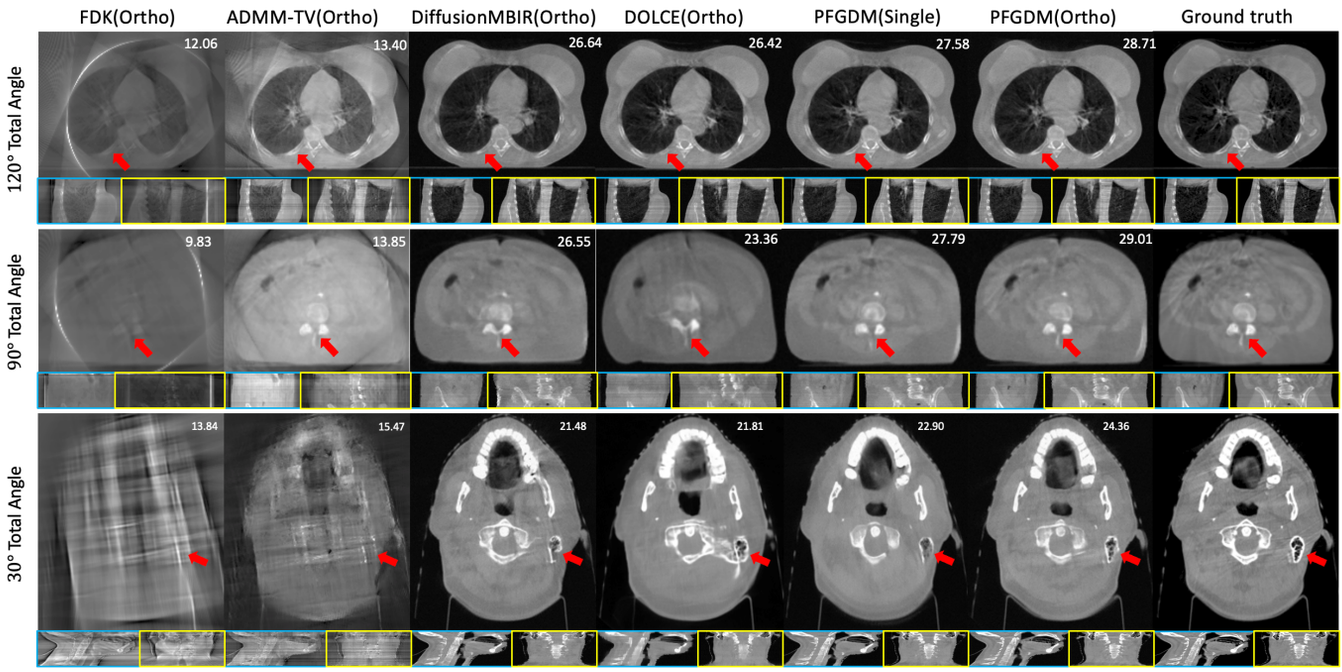


Fig. 4. Visual comparison between the reconstructed LA-CBCTs of different methods and the 'ground truth'. The limited-angle projections were simulated with either orthogonal-view (ortho) acquisitions for a total scan angle of 30°([0°, 15°], [90°, 105°]), 90°([0°, 45°], [90°, 135°]), or 120°([0°, 60°], [90°, 150°]); or single-view (single) acquisitions for a total scan angle of 30°([0°, 30°]), 90°([0°, 90°]), or 120°([0°, 120°]). PFGDM results with single-view acquisitions are shown in the fifth column. Coronal and sagittal views are shown in blue and yellow boxes, respectively. PSNR scores of each reconstruction against the 'ground truth' in axial slice are indicated at the top-right corner of each figure. See Table I for quantitative results.

all limited-angle scenarios for axial, coronal, and sagittal views. The advantage of PFGDM becomes especially evident when the scan angle decreases, showing the efficacy of introducing prior CT information when the on-board projection information is limited. Notably, PFGDM based on single-view projections already yields better or comparable results when compared with other methods that are based on orthogonal-view projections.

IV. Conclusion

In this work, we introduced PFGDM, a diffusion model-based method for LA-CBCT reconstruction. PFGDM incorporates the high-frequency information from the patient's existing CT scans as additional prior to compensate the missing angle information in LA-CBCT reconstruction. We demonstrated the robustness of this method under different limited-angle scenarios, using scan angles as small as 30°. PFGDM outperformed the other LA-CBCT reconstruction techniques by a clear margin, especially for small scan angles. PFGDM also makes a general method that does not need re-training for different imaging geometries or limited-angle scenarios, providing it a substantial advantage over methods like DOLCE for potential clinical translation.

References

1. C. Walter, J. C. Schmidt, C. A. Rinne, S. Mendes, K. Dula, and A. Sculean, "Cone beam computed tomography (CBCT) for diagnosis and treatment planning in periodontology: systematic review update," *Clinical Oral Investigations*, vol. 24, no. 9, pp. 2943–2958, Jul. 2020, doi: https://doi.org/10.1007/s00784-020-03326-0.
2. Z. Li et al., "A sinogram inpainting method based on generative adversarial network for limited-angle computed tomography," *arXiv (Cornell University)*, May 2019, doi: https://doi.org/10.1117/12.2533757.
3. J. Gu and J. Chul Ye, "Multi-scale wavelet domain residual learning for limited-angle CT reconstruction," 2017, arXiv:1703.01382.
4. 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," in *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2912-2924, Oct. 2022, doi: 10.1109/TMI.2022.3175529.
5. 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. Imag.*, vol. 41, no. 7, pp. 1778–1790, Jul. 2022.
6. P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis," *arXiv:2105.05233 [cs, stat],* Jun. 2021, Available: https://arxiv.org/abs/2105.05233
7. J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," *arXiv:2006.11239 [cs, stat],* Dec. 2020, Available: https://arxiv.org/abs/2006.11239
8. Liu, Jiaming et al. "DOLCE: A Model-Based Probabilistic Diffusion Framework for Limited-Angle CT Reconstruction." *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* (2022): 10464-10474.
9. H. Chung, D. Ryu, M. Mccann, M. Klasky and J. Ye, "Solving 3D Inverse Problems Using Pre-Trained 2D Diffusion Models," in *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 2023 pp. 22542-22551.doi: 10.1109/CVPR52729.2023.02159
10. Y. Li et al., "Zero-shot Medical Image Translation via Frequency-Guided Diffusion Models," in *IEEE Transactions on Medical Imaging*, doi: 10.1109/TMI.2023.3325703.
11. Song, Y., et al., Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. 1, 2, 3, 5
12. P. Vincent, "A Connection Between Score Matching and Denoising Autoencoders," *Neural Computation*, vol. 23, no. 7, pp. 1661–1674, Jul. 2011, doi: https://doi.org/10.1162/neco_a_00142.
13. N. Kanopoulos, N. Vasanthavada, and R. L. Baker, "Design of an image edge detection filter using the sobel operator," *IEEE Journal of solid- state circuits*, vol. 23, no. 2, pp. 358–367, 1988.
14. Y. Zhang, F.-F. Yin, W. P. Segars, and L. Ren, "A technique for estimating 4D-CBCT using prior knowledge and limited-angle projections," *Medical Physics*, vol. 40, no. 12, p. 121701, Nov. 2013, doi: https://doi.org/10.1118/1.4825097.

Diffusion Model Based Limited-angle CT Reconstruction for Flat Objects

Zhizhong Ran, Dimeng Xia, Shusen Zhao, Weibin Zhang, and Xiangyuan Liu

***Abstract*—Inspecting chips, circuit boards, and other flat electronic package components is essential for limited-angle X-ray CT applications. To solve the limited-angle artifacts in flat objects CT reconstruction and improve the quality of CT images, regularization-based methods are relatively mainstream research directions. In practice, it is not easy to meet the requirements of limited-angle CT for both reconstruction quality and reduced iteration. In this paper, we proposed an edge-preserving reconstruction method based on the diffusion model. The method takes full advantage of the property that the gradient and gray values at the visible boundary in the limited-angle CT reconstruction image are both approximately accurate and construct an optimization model. Moreover, the proposed method combines the time-stepping technique and the edge information to obtain the heat diffusion equation, which is solved to get high-quality images. Numerical experiments show that the proposed method requires only a small number of iterations to obtain the noticeable effect of limited-angle artifact removal.**

***Index Terms*—Limited-angle CT, diffusion model, flat objects.**

I. Introduction

X-RAY computed tomography (CT) is currently an essential approach for non-destructive testing of electronic packaging components (e.g., chips, circuit boards) and other flat objects. In practice, such flat objects are often placed close to the X-ray tube to improve the image resolution, so in this paper, we consider a limited-angle CT scanning system for flat objects shown in Figure 1. Figure 1(a) shows a schematic diagram of the limited-angle X-ray scanning trajectory and the relative position of the object. The object image is shown in Figure 1(b), and the reconstruction results of the projection data acquired with the scanning mode in 1(a) are shown in Figure 1(c). Figure 1(c) has more obvious limited-angle artifacts covering the long-edge structure of the object. In the past decades, many researchers have studied various algorithms for limited-angle CT reconstruction of flat objects, among which regularization-based methods are relatively mainstream research directions. Since the CT images of chips, circuit boards, and other flat objects are sparse in some transform domains, TV regularization-based methods are usually used to solve these ill-posed inverse problems of limited-angle CT reconstruction, such as the regularization model based on isotropic TV[1–3], on anisotropic TV_{l0}-regularization[4, 5], *et al*. These methods have improved the quality of limited-angle CT reconstruction images. However, they do not adequately consider the direction of limited-angle artifacts, which cannot achieve the ideal artifacts removal effect. In 2019, Xu *et al*. proposed the AEDS method[6], which constructs a single-direction optimization model by taking advantage of the property that the gray values at visible boundaries are approximately accurate, which can achieve high-quality reconstructed images with a sufficient number of iterations. Then, Pan *et al*. proposed the DTV model[7], which can reconstruct ideal CT images with projection data in a small amount of angular range and can be regarded as a constrained representation of the AEDS model in the case of multi-subdirectional TV regularization.

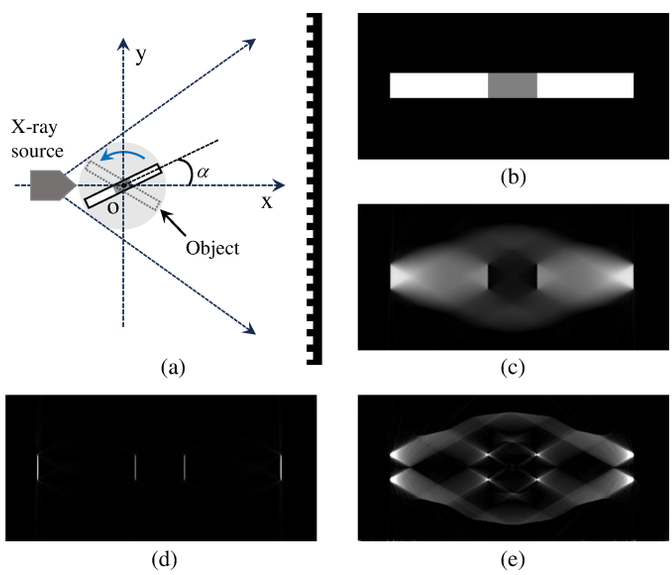


Fig. 1. Schematic of a limited-angle scanning and reconstructed images. (a) Illustration of circular and limited-angle fan-beam CT scanning configuration. (b) The object image. (c) Reconstructed image with scanning angular range $\alpha \in [\pi/6, 5\pi/6]$. (d) The gradient magnitude image along the x direction of the reconstruction image. (e) The gradient magnitude image along the y direction of the reconstruction image. The gray scale window of (b)(c) are set to [0, 1], the gray scale window of (d) is set to [0, 0.5] and the gray scale window of (e) is set to [0, 0.05].

In 2021, our team proposed a limited-angle CT reconstruction algorithm based on the differential property of Fourier transform (DPFT)[8], which constructed an optimization model by using the property that the gradient values at the visible boundaries are accurately approximated. The differential property of the Fourier transform is utilized in the implementation process. The method effectively improves the convergence speed of the limited-angle CT reconstruction of the flat

This work was supported by the Shenzhen Science and Technology Program (Grant No. RCBS20221008093316046 and JSGGZD20220822095600001)

Z. Ran, Dr. D. Xia, Dr. S. Zhao, Dr. W. Zhang and Dr. X. Liu are with the National Center for Applied Mathematics Shenzhen, Southern University of Science and Technology, Shenzhen 518055, China.

Corresponding author: Dr. Shusen Zhao(zhaoss@sustech.edu.cn).

objects but makes an approximation when processing the 0 frequency with the Fourier transform differentiation property, which introduces sawtooth artifacts in the reconstructed image. Therefore, we propose a new diffusion model that makes full use of the property that the gradient and gray values at the visible boundary in the limited-angle CT reconstructed image are both accurately approximated, and a high-quality reconstructed image can be achieved by solving this model.

II. Method

A. Inspiration of the Proposed Method

In 2017, Quinto *et al*. proposed the concept of "visible and invisible boundaries" in CT reconstruction based on the theory of microlocalization analysis[9]. According to the scanning trajectory shown in Figure 1(a), there are more rays in the vertical direction but no rays in the horizontal direction, so the vertical boundary of the reconstructed image is visible, and the horizontal boundary is invisible. Figures 1(d) and 1(e) show the gradient images of 1(c) along the horizontal and vertical directions, respectively. The vertical edges in 1(d) are sharp and less different from phantom Figure 1(b), while Figure 1(e) shows that the horizontal edges of Figure 1(c) are unsatisfactory. When a visible boundary similar to Figure 1(d) is obtained, the most direct treatment is to use some interpolation methods to recover the structure inside the boundary. However, the simple interpolation methods usually fail to achieve a satisfactory recovery effect for flat objects with complex structures. Moreover, it's not easy to obtain a high-quality CT image with only one interpolation calculation.

B. The Proposed Method

Based on the above theory and observation, in this paper, we propose a new diffusion model and combine it with an iterative reconstruction method to recover limited-angle CT images iteratively. In the iterative framework, first, the CT image with limited-angle artifacts is obtained by statistical reconstruction method[8]; second, the proposed diffusion model is solved to recover the structural information of the limited-angle CT image while realizing the edge-preserving effect; then, the solution result is used as the initial value of the following statistical reconstruction to start the next round of iteration.

Here, we proposed a diffusion model as below

$$\begin{cases} u_t(t;\mathbf{x}) = u_{xx}(t;\mathbf{x}) + \mu(u_{xx}(t;\mathbf{x}) - \phi_{xx}(\mathbf{x})), & \mathbf{x} \in \Omega' \\ u(t;\mathbf{x}) = \phi(\mathbf{x}), & \mathbf{x} \in \partial\Omega \\ u(0;\mathbf{x}) = \phi(\mathbf{x}), & \mathbf{x} \in \Omega \end{cases} . \tag{1}$$

where the function u denotes the CT image, t and $\mathbf{x}$ are two variables that represent two-dimensional coordinates in time and space, respectively. $\partial\Omega$ denotes the set of edge points, Ω' the set of non-edge points. ϕ denotes the reconstruction image with limited-angle artifacts. u_t denotes the partial derivative of u with respect to time, u_{xx} the second-order partial derivative of u with respect to the horizontal direction x. In this work, image information of the horizontal direction x is missing, which means it is the direction of invisible boundaries. Therefore, only the propagation along the long edge of the flat object is discussed in the proposed diffusion model, and taking the second partial derivative of x with respect to u will suffice. μ is an adjustable coefficient. Let $\mathrm{RECON}(u^{(j)}, q)$ denote the procedure of the statistical reconstruction with projection data q and initial image $u^{(j)}$. How the proposed method can be solved iteratively is summarized shown in **Algorithm1**.

TABLE I
NUMERICAL INDEXES OF RECONSTRUCTION RESULTS BY THE METHODS IN FIG.3

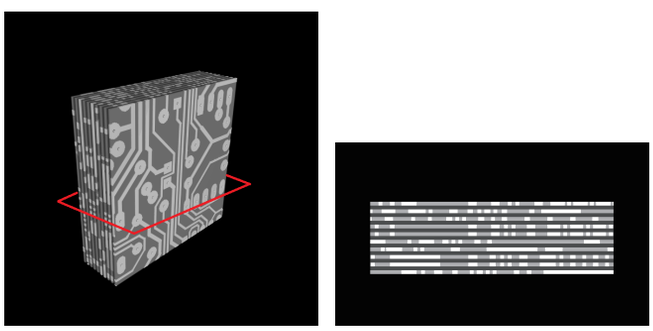
	SART	TV_{l0}	AEDS	DPFT	the proposed method
PSNR	85.5032	83.9537	86.9038	90.0829	**90.4969**
NMAD	0.7866	0.6362	0.4050	0.2977	**0.2896**
NRMSE	0.3960	0.3570	0.1810	0.1449	**0.0791**

Algorithm 1 the proposed algorithm

Given the projection data q.
Initialization: image $u^{(0)} = 0$
$j := 0$
while (stop criterion is not met) **do**
  Update the image by
  $u^{(j+1/2)} = \mathrm{RECON}(u^{(j)}, q)$
  Update the image by solving Eq.(1)
  $u^{(j+1)} = \mathrm{SOLVE}(\mathrm{Eq.}(1))$
  $j := j + 1$
end while

III. Results

In this paper, we validated the effectiveness of the proposed method for flat CT reconstruction using noisy simulated PCB data with the scanning angle set at $\alpha \in \left[\frac{\pi}{6}, \frac{5\pi}{6}\right]$. The experimental scanning parameters are set as follows: the distance from the source to the center of the rotation center (SOD) is 600 mm, and the distance from the source to the detector (SDD) is 1,000 mm; the line detector consists of 512 detector units, and the size of each detector unit is 0.25 mm; the noisy projection data are generated by adding Poisson and Gaussian noise to the noisy-free data; the initial number of photons of Poisson noise is 10^5; the mean of Gaussian noise is 0, and the standard deviation is 0.025.



(a) (b)

Fig. 2. The simulated PCB phantom. (a) Simulated 3D PCB phantom. (b) Simulated 2D PCB phantom used for experiments.

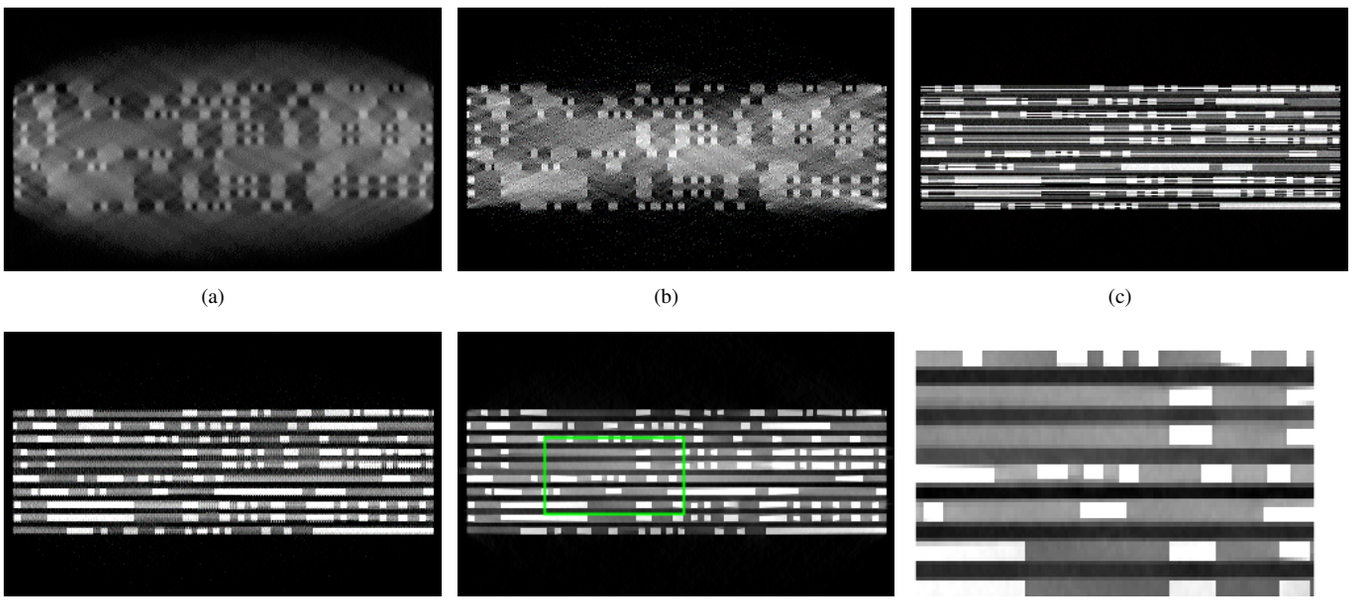


Fig. 3. Comparison of reconstructed images with different methods. (a) SART reconstruction (1 iteration). (b) TV_{l0} method (100 iterations). (c) AEDS method (100 iterations). (d) DPFT method (5 iterations). (e) The proposed method (5 iterations). (f) An enlarged view of the green box area in (e). All the display windows are set [0,0.15].

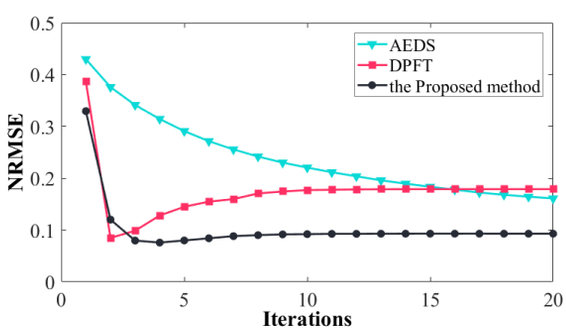


Fig. 4. NRMSE curves of AEDS method, DPFT method and the proposed method in 20 iterations.

Figure 3 shows the experimental results of the proposed method and the other four results for comparison. As shown in Figure 3, the image of the SART method[10] 3(a), in the presence of noise, leads to noticeable limited-angle artifacts because no prior information about the reconstructed image is used. For the image of the TV_{l0} method 3(b), the artifacts are removed to some extent, but for the horizontal boundaries of the long metal lines, the TV_{l0} method does not recover effectively and introduces pseudo-structures in the image. This is because the TV_{l0} method is isotropic, and unclear edges along the vertical direction reduce the ability of TV_{l0} to preserve boundaries along the horizontal direction. Obvious horizontal line artifacts appear in Figure 3(c), which is reconstructed by the AEDS method. Rough structures of the flat object are recovered by the AEDS method after 100 iterations, but due to the edge-preserving performance of the one-dimensional l_0 norm minimization algorithm used in the AEDS method is too strong, the errors in the noisy data gradually increase in the iterations. As for the DPFT method in Figure 3(d), it has removed most of the boundary artifacts after five iterations, but sawtooth artifacts are introduced in the reconstructed image due to the use of approximations near the 0 frequency in the computation. For the reconstruction image of the proposed method in Figure 3(e), the image quality is relatively higher, and the structure of the PCB phantom is more clearly recovered compared to the other four experimental comparison methods. Moreover, it does not appear similar sawtooth artifacts at the same number of iterations as the DPFT method shown in Figure 3(f). Some numerical metrics are in Table I, and the NRMSE curves are shown in Figure 4. It can also be seen that the result of the proposed method has relatively higher quality than those of the above four methods.

IV. Discussion and Conclusion

In this paper, we proposed a diffusion model-based limited-angle CT reconstruction method for flat objects. The method makes full use of the property that the gradient and gray values at the visible boundaries in the limited-angle CT reconstructed image are both approximately accurate, proposes an optimization diffusion model. And we solved the model by an iterative method with a small number of iterations to achieve high-quality limited-angle artifacts removal. Note that the total number of iterations of the image reconstruction process used in this method is small, yet in solving the diffusion model, the number of computation times required for convergence is related to the size of the image to be processed. Since the results of the previous computation must be used for each step, the parallelism of this method currently needs to be higher, and the processing time is still long.

APPENDIX

The Dirichlet energy functional is based on the gradient of image processing, which can smooth the image and remove noise and artifacts[11, 12]. At the same time, due to the significant difference between the gradient at the edge and the gradient at the non-edge, the gradient-based processing can divide the image from the perspective of the edge, which makes good use of the property that the edge information of the limited-angle CT reconstruction image is approximately accurate.

The Dirichlet energy functional minimization is as below

$$E[u] = \frac{1}{2}\int_{\Omega} \|\nabla u\|^2. \tag{2}$$

In this paper, we reconstruct the image using approximately accurate edge information, which is guaranteed to be invariant at the edges when iterating. We consider minimizing the Dirichlet energy generalization at all non-edges $v : \Omega' \to \mathbf{R}$,

$$E[v] = \frac{1}{2}\int_{\Omega'} \|\nabla v\|^2, \tag{3}$$

To avoid overfitting, we add a term $\mu E[v - \phi]$, μ could be adjusted by experimental data. Since the Dirichlet energy functional is based on the gradient, this term can be seen as a data fidelity term for the gradient, which ensures the consistency of the gradient change during the iteration process when subsequently solving the numerical PDE. The following energy functional is then obtained:

$$F[v] = \frac{1}{2}\int_{\Omega'} \|\nabla v\|^2 + \frac{\mu}{2}\int_{\Omega'} \|\nabla(v-\phi)\|^2, \tag{4}$$

Combined with the characteristics of the flat objects CT limited-angle scanning system, this method only needs to consider the propagation of x along a single direction. The l_2 norm is chosen for its high calculation efficiency, and the energy functional above is transformed into:

$$F[v] = \frac{1}{2}\int_{\Omega'} \|v_x\|_2^2 + \frac{\mu}{2}\int_{\Omega'} \|v_x - \phi_x\|_2^2, \tag{5}$$

Minimize the v value of this energy functional by solving it with the calculus of variations, considering its Euler-Lagrange equation

$$f_v - \frac{d}{dx} f_{v_x} = 0, \tag{6}$$

where,

$$f[v] = \frac{1}{2}v_x^2 + \frac{\mu}{2}(v_x - \phi_x)^2. \tag{7}$$

Solving the Euler-Lagrange equation above yields the solution in steady state, and its unsteady state is obtained using time stepping technique as follows:

$$v_t = v_{xx} + \mu(v_{xx} - \phi_{xx}), \tag{8}$$

Finally, in the non-edge area we have $u = v$, in the edge area and when time is 0 we have $u = \phi$, then the PDE model is

$$\begin{cases} u_t(t;\mathbf{x}) = u_{xx}(t;\mathbf{x}) + \mu(u_{xx}(t;\mathbf{x}) - \phi_{xx}(\mathbf{x})), & \mathbf{x} \in \Omega' \\ u(t;\mathbf{x}) = \phi(\mathbf{x}), & \mathbf{x} \in \partial\Omega \\ u(0;\mathbf{x}) = \phi(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{9}$$

By solving equation 9, the limited-angle CT reconstruction is realized.

REFERENCES

[1] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, "Improved total variation-based ct image reconstruction applied to clinical data," *Physics in Medicine & Biology*, vol. 56, no. 6, p. 1545, 2011.

[2] H. Yu and G. Wang, "A soft-threshold filtering approach for reconstruction from a limited number of projections," *Physics in Medicine & Biology*, vol. 55, no. 13, p. 3905, 2010.

[3] S. J. LaRoque, E. Y. Sidky, and X. Pan, "Accurate image reconstruction from few-view and limited-angle data in diffraction tomography," *JOSA A*, vol. 25, no. 7, pp. 1772–1782, 2008.

[4] Z. Chen, X. Jin, L. Li, and G. Wang, "A limited-angle ct reconstruction method based on anisotropic tv minimization," *Physics in Medicine & Biology*, vol. 58, no. 7, p. 2119, 2013.

[5] X. Jin, L. Li, Z. Chen, L. Zhang, and Y. Xing, "Anisotropic total variation for limited-angle ct reconstruction," in *Nuclear Science Symposium Conference Record (NSS/MIC), 2010 IEEE.* IEEE, 2010, pp. 2232–2238.

[6] J. Xu, Y. Zhao, H. Li, and P. Zhang, "An image reconstruction model regularized by edge-preserving diffusion and smoothing for limited-angle computed tomography," *Inverse Problems*, vol. 35, no. 8, p. 085004, 2019.

[7] 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*, vol. 70, p. 102030, 2021.

[8] S. Zhao, D. Xia, and X. Zhao, "A fast image reconstruction method for planar objects CT inspired by differentiation property of fourier transform (DPFT)," *Inverse Problems*, vol. 37, no. 7, 2021.

[9] E. T. Quinto, "Artifacts and visible singularities in limited data x-ray tomography," *Sensing and Imaging*, vol. 18, no. 1, p. 9, 2017.

[10] 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.

[11] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[12] Y. Zhao, J. Xu, H. Li, and P. Zhang, "Edge information diffusion-based reconstruction for cone beam computed laminography," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4663–4675, 2018.

Latent Space Reconstruction and its Application to CT Detruncation: Latent Detruncation

Anton Kabelac, Elias Eulig, Joscha Maier, Michael Knaup, and Marc Kachelrieß

***Abstract*—Truncation in CT occurs when parts of the patient laterally exceed the field of measurement (FOM). This is typically the case for obese patients in clinical CT and for most patients in CBCT unless CBCT uses a shifted detector or a very large detector. Conventional reconstruction algorithms (analytic, iterative, deep learning) will also suffer from severe truncation artifacts. Correcting for these artifacts within the FOM is easy, but providing a good reconstruction quality of the parts that lie outside the FOM is hard and has never been achieved, yet.**

In this work, we propose a novel deep learning-based approach called the latent space reconstruction (LSR). We apply LSR to the detruncation problem. It can not only correct for the inner FOM artifacts but is also able to extend the images far beyond the FOM and thereby achieves a yet unseen high image quality. Our LSR approach uses a variational autoencoder (VAE) that has been trained for full CT data representation to find the closest non-truncated latent space representation to the truncated rawdata.

Results: Images corrected by LSR show effective artifact suppression within the FOM, alongside a substantial high-quality extension of the FOM.

***Index Terms*—CT Image Reconstruction, Latent Space Reconstruction, Autoencoder, Detruncation**

I. Introduction

CT images are typically reconstructed using algorithms of filtered backprojection (FBP) type, of iterative reconstruction type, and, just recently, also with the help of deep learning as a post processing step. A high quality reconstruction requires the rawdata to be complete in a mathematical sense. Missing data can, e.g., be caused by lateral truncation, where parts of the patient laterally exceed the field of measurement (FOM). The lateral FOM is the region within the x-y-plane where each voxel is viewed under 180° or more. In clinical practice, truncation most often occurs with patients that are not centered properly on the table, obese patients, or when using C-arm systems, which generally are equipped with small, say, 30 to 40 cm, flat detectors and thus feature small FOMs of about 15 to 20 cm in diameter.

Truncation causes cupping-like artifacts inside the FOM which become apparent especially at the FOM edges. Furthermore, regions outside the FOM cannot be reconstructed at all since are seen by less than 180° and thus suffer from severe limited angle artifacts.

Over the last decades, a number of methods for detruncation, i.e. truncation artifact correction and field of view (FOV) extension, have been proposed. Early techniques for the reconstruction of incomplete projection data primarily fall roughly into two groups. The first group is made up of data completion methods, like more or less simple extrapolation algorithms. Reference [1] used the size and slope of a water cylinder fitted into each projection to estimate suitable projection extension. Other studies used cosine fitting and elliptical extrapolation, sometimes combined with consistency conditions, to estimate a convex hull of the patient in order to extrapolate the truncated projections [2]. A more sophisticated extrapolation uses information from a patient atlas, where the atlas patients are registered to the current patient and then forward projected [3]. This prior-based method, that was also used to correct other types of missing data artifacts, is quite inconvenient to implement and use since it requires large amounts of atlas data and a deformable registration. This work, as we will see below, can be thought of replacing the atlas and the registration by a neural network.

The second group consists of methods based on iterative reconstruction. Iterative reconstruction often involves creating an objective function that incorporates observed data and prior understanding of CT images, and then deriving the solution by optimizing this function [4, 5].

Recently, deep learning (DL)-based methods have proven to be powerful tools throughout many medical imaging applications. DL-based detruncation methods proposed in the literature can be categorized based on their data utilization: some methods work directly with measured projection data (sinogram-based methods), while others operate on the reconstructed CT images.

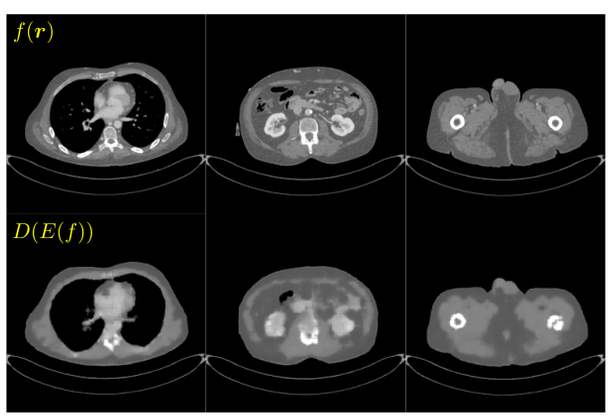


Fig. 1. Reconstructed images from the VAE. The first row shows ground truth images f from the test data, the second row shows the VAE output $D(E(f))$. $C = 50$ HU, $W = 1200$ HU

Anton Kabelac, Elias Eulig, and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls-University, Heidelberg, Germany. Dr. Joscha Maier and Dr. Michael Knaup are with the German Cancer Research Center (DKFZ), Heidelberg, Germany.

Corresponding author: anton.kabelac@dkfz.de

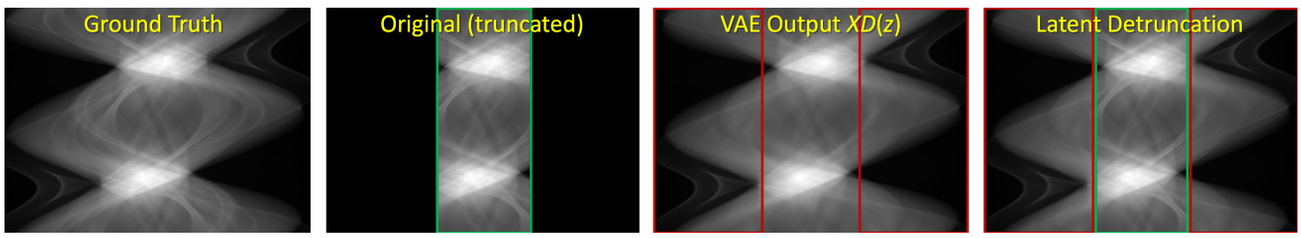


Fig. 2. Overview of our processing pipeline. From left to right: the ground truth sinogram, clipped sinogram to simulate truncation, forward projection of $D(z)$, and latent detruncated sinogram.

Rawdata-based networks act directly on the incomplete (truncated) measured projection data. These are typically extended with a U-net convolutional neural network, and the extended sinogram is reconstructed with FBP [6].

Image-based methods, such as [7] can be compared to semantic image inpainting problems like, where the mask is just the missing outer area of the image. The truncated area is usually defined via a mask and a network is trained to inpaint the missing regions.

Dual-domain based methods combine the previous two methods and use two or more deep neural networks, one in the projection data domain and one in the image domain. In [8] two GANs where used, trained end-to-end in two stages: first, for sinogram extension, and second, for post-processing the initial reconstruction, with each stage involving its own generator and discriminator trained.

In this paper, we propose LSR which is a novel method based on latent space optimization of a generative model, which here is chosen to be a variational autoencoder (VAE), trained on complete CT images. We apply LSR to the truncation problem and call this method the latent detruncation. Note that a similar approach may be applied to other reconstruction problems in CT. Exploring these is left for future work.

II. Method and Data

Here, we restrict ourselves to 2D, for convenience. CT images are denoted by $f(x, y)$, non truncated raw projection data by $p(\vartheta, \xi)$, and truncated raw projection data by $q(\vartheta, \xi)$. X is representing a forward projection. The angle ϑ and the distance parameter ξ parameterize the ray $x \cos \vartheta + y \sin \vartheta = \xi$.

A. Data Sets

The CT data were acquired with a Siemens Somatom Force CT system (Siemens Healthineers, Forchheim, Germany). Our data set consists of 85 adult patients (2/3 male, 1/3 female), with a total of 94117 slices. Images where acquired at 70 kV and reconstructed into a 512×512 matrix with a slice thickness of 0.6 mm and an axial voxel spacing between 0.69 and 0.98 mm. The patients were randomly split into 59 patients for training, 13 for validation and 13 for testing (70:15:15). To decrease training time, all images where downsampled to 256×256 pixels.

B. Truncation Simulation

We generate full $p(\vartheta, \xi)$ and truncated sinograms $q(\vartheta, \xi)$ by forward-projecting the ground truth CT images. The simulations are performed in parallel beam geometry with $N = 360$ projections covering an angular range of 180°. The full sinogram has $M = 500$ detector pixels with a pixel size of $\triangle\xi = 1$ mm and corresponds to a 500 mm FOM (large FOM). We investigate two levels of truncation, one with a 350 mm FOM (medium FOM), corresponding to the B source of today's dual source CT systems and one with a 150 mm FOM (small FOM) to mimic typical C arm FOMs. Truncation simulation is achieved by zeroing the outermost left and right 75 detector channels for the medium FOM and the outermost left and right 175 detector channels for the small FOM.

C. Proposed Approach: The Latent Detruncation

Initially, a variational autoencoder $D(E(f))$, consisting of an encoder E and a decoder D, is trained to learn a full non-truncated CT image representation. Given the VAE we are looking for a latent space vector z such that $D(z)$ generates a viable CT image while simultaneously ensuring that $\mathsf{X}D(z)$ closely matches the measured truncated rawdata, $q(\vartheta, \xi)$, within the measurement boundaries. We find z by solving the following optimization problem

$$z = \arg\min_z \|\mathsf{X}D(z) - p\|, \tag{1}$$

which we call the latent space reconstruction (LSR). The measured truncated rawdata q are then detruncated by extrapolating them in the outer regions with the values from $\mathsf{X}D(z)$.

D. Network Architecture

Here, we opted for a VAE architecture, primarily because LSR prefers a generative model, i.e. a well-behaving generator function (our decoder) that, given a low-dimensional parameter vector z, is able to generate a realistic representation of the object and close neighbors of z generate neighboring objects. The VAE has these properties. Its latent space corresponds to the parameters of a contiguous variational distribution. This characteristic facilitates effective sampling within the latent space. The input and output dimensions of the VAE are 256×256 pixels. Our encoder is based on the ResNet50 architecture. The ResNet-50 architecture is a variant of the ResNet architecture [9] and comprises 50 layers. It features a bottleneck design with 1×1 convolutions, comprising multiple layers including a 7×7 kernel convolution, max pooling, and various iterative layers with different kernel sizes, culminating in average pooling and a softmax-activated fully connected layer. The dimension of the latent space z is 256.

Our decoder reconstructs images from latent representations using a fully connected layer for initial expansion, followed by

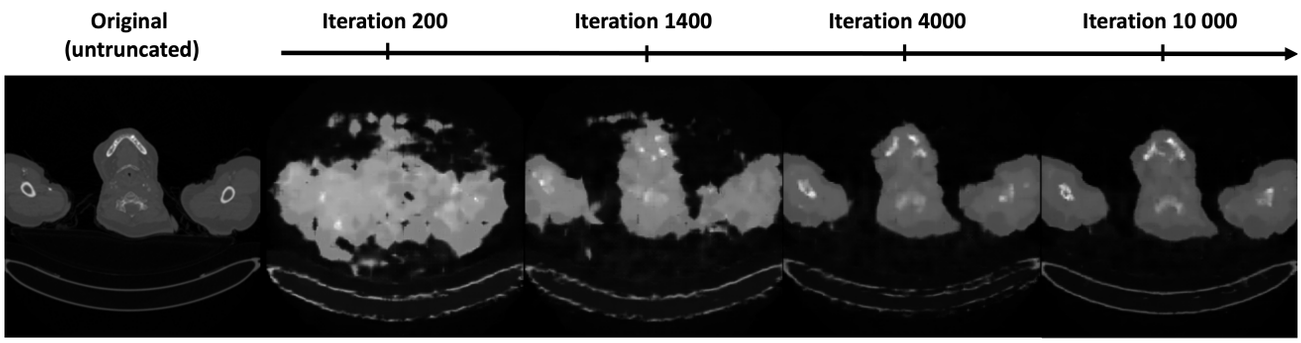


Fig. 3. Original image and optimization steps $D(z)$ of the latent space vector, starting from a randomly initialized latent space vector $z \sim \mathcal{N}(0, 1)$.

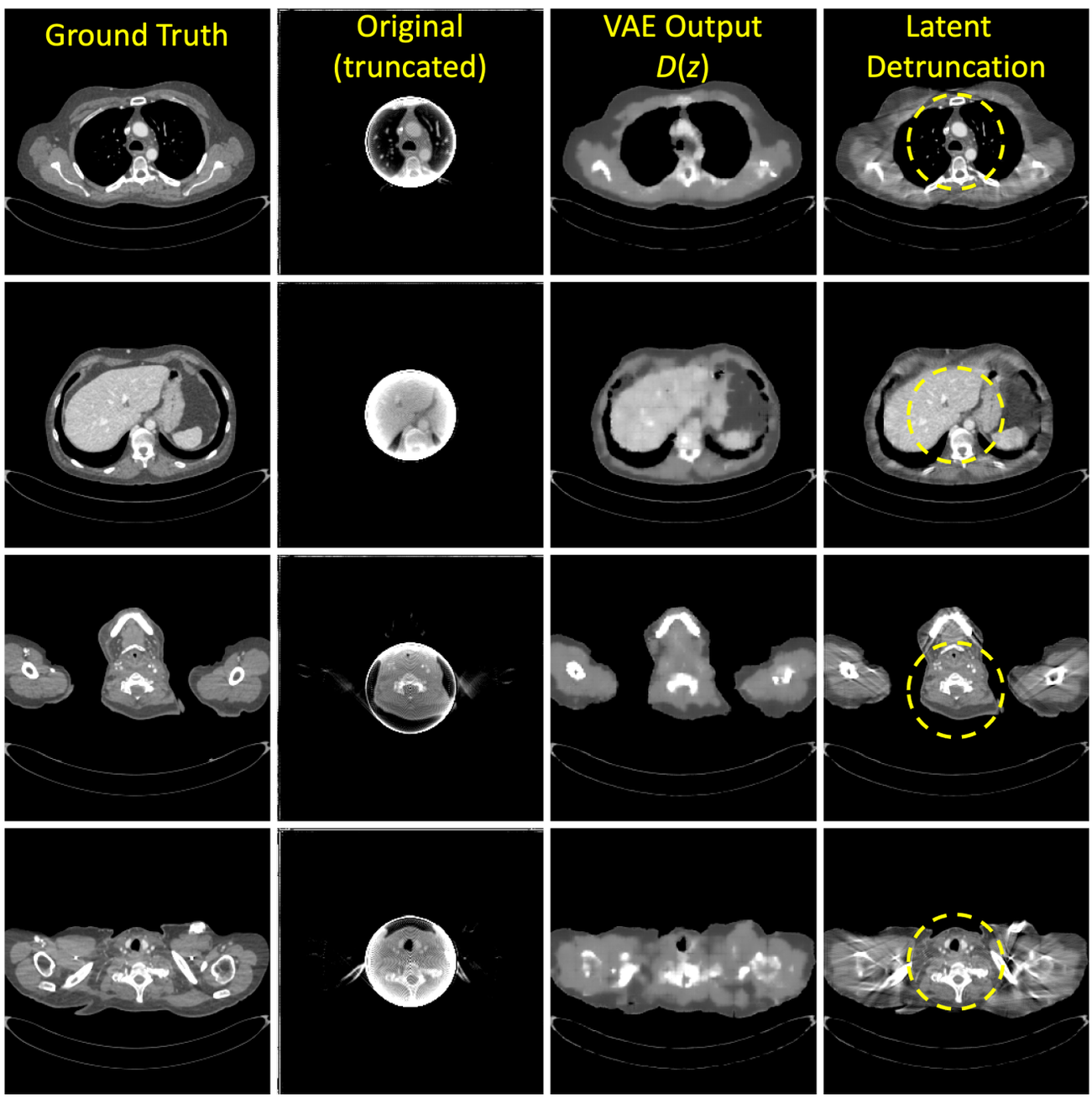


Fig. 4. Detruncation results for four different patients. From left to right: ground truth image, truncated image (reconstructed from truncated sinogram), $D(z)$, and reconstruction from latent detruncated sinogram. $C = 50$ HU, $W = 1200$ HU

a series of transposed convolutional layers with ReLU activations. The VAE employs a hybrid loss function, combining L1 loss and Kullback-Leibler (KL) divergence loss. To optimize this architecture, we set the learning rate to 0.001, the batch size to 16, and trained for 100 epochs.

III. Results

In Fig. 1, we present an illustration of the proficiency of our VAE in replicating a randomly selected CT image from our test dataset. The results, as shown in the second row, bear a close resemblance to the original CT images featured in the first row. However, it is noteworthy that while the VAE effectively captures the overall structure and composition of the images, the reproduced images exhibit a noticeable degree of blurriness.

Fig. 2 shows the sinograms of our latent detruncation pipeline: The forward-projected, untruncated ground truth sinogram $p(\vartheta, \xi)$, followed by the truncated sinogram $q(\vartheta, \xi)$, the forward projection of $D(z)$ and the latent detruncated sinogram. Measured data are framed with a green frame while the latent detruncated outer sinogram regions are indicated with a red frame.

The LSR optimization process is demonstrated in Fig. 3. LSR's aim in this paper here is to generate an image whose forward projection approximates the truncated rawdata of the patient. The original patient image is displayed in the top-left corner. Truncation to 15 cm FOM occured in the sinogram domain.

Figure 4 shows our approach applied to four randomly sampled images from the test dataset. For each sample we show the original CT image, the FBP reconstruction of the truncated sinogram $q(\vartheta, \xi)$, the decoded image $D(z)$ for the optimized latent vector z, and the FBP reconstruction of the latent detruncated data. Here we find that while the VAE is able to adequately reconstruct the overall shape of the patient. However, $D(z)$ is blurry and misses high-frequency information. The last column is the latent detruncation result, i.e. the reconstruction of the patient both outside and inside the FOM.

Similar results were achieved for the medium FOM, but are omitted here due to space constraints.

IV. Discussion and Conclusion

In our study, we presented a novel deep learning method termed latent space reconstruction (LSR), applied to the problem of detruncation. Our findings show that LSR, even when implemented with a basic variational autoencoder (VAE) architecture, proves to be a highly efficient solution for correcting truncation artifacts. Several improvements of the utilized VAE, such as the use of a perceptual loss to reduce the blurriness, are planned for the future aiming at enhancing the quality and generalizability of LSR for the truncation problem and for various other types of missing data artifacts.

V. Acknowledgements

The projection simulation software RayConStruct-PS and the image reconstruction software RayConStruct-IR were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] 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". In: *Medical Physics* 31.9 (Sept. 2004), pp. 2385–2391.

[2] K. Sourbelle, M. Kachelriess, and W. A. Kalender. "Reconstruction from truncated projections in CT using adaptive detruncation". In: *Eur Radiol* 15.5 (May 2005), pp. 1008–1014.

[3] T. Heußer, M. Brehm, L. Ritschl, S. Sawall, and M. Kachelrieß. "Prior-based artifact correction (PBAC) in computed tomography". In: *Medical Physics* 41.2 (Feb. 2014), p. 021906.

[4] A. Chambolle and P. Lions. "Image recovery via total variation minimization and related problems". In: *Numerische Mathematik* 76.2 (Apr. 1997), pp. 167–188.

[5] Ludwig Ritschl, Frank Bergner, Christof Fleischmann, and Marc Kachelrieß. "Improved Total Variation–Based CT Image Reconstruction Applied to Clinical Data". In: *Phys. Med. Biol.* 56.6 (Feb. 2011), pp. 1545–1561.

[6] J. H. Ketola, H. Heino, M. A. K. Juntunen, M. T. Nieminen, and S. I. Inkinen. "Deep learning-based sinogram extension method for interior computed tomography". In: *Medical Imaging 2021: Physics of Medical Imaging*. Feb. 2021, p. 123.

[7] G. P. Fonseca, M. Baer-Beck, E. Fournie, C. Hofmann, I. Rinaldi, M. C. Ollers, W. J.C. Van Elmpt, and F. Verhaegen. "Evaluation of novel AI-based extended field-of-view CT reconstructions". In: *Medical Physics* 48.7 (July 2021), pp. 3583–3594.

[8] J. H. Ketola, H. Heino, M. A. K. Juntunen, M. T. Nieminen, S. Siltanen, and S. I. Inkinen. "Generative adversarial networks improve interior computed tomography angiography reconstruction". In: *Biomed. Phys. Eng. Express* 7 (Nov. 2021), p. 065041.

[9] H. Kaiming, Z. Xiangyu, R. Shaoqing, and S. Jian. "Deep Residual Learning for Image Recognition". In: *arXiv* (Dec. 2015). eprint: 1512.03385.

Morphing from the TV-Norm to the l_0-Norm

Gengsheng L. Zeng and Ya Li

***Abstract*—For piecewise constant objects, the images can be reconstructed with under-sampled measurements. The gradient image of a piecewise image is sparse. If a sparse solution is a desired solution, an l_0-norm minimization method is effective to solve an under-determined system. However, the l_0-norm is not differentiable, and it is not straightforward to minimize an l_0-norm. This paper suggests a function that is similar to the l_0-norm function, and we refer to this function as meta l_0-norm. The subdifferential of the meta l_0-norm has a simple explicit expression. Thus, it is straightforward to derive a gradient descent algorithm to enforce the sparseness in the solution. In fact, the proposed meta norm is a transition that varies between the TV-norm and the l_0-norm. As an application, this paper uses the proposed meta l_0-norm for few-view tomography. Computer simulation results indicate that the proposed meta l_0-norm effectively guides the image reconstruction algorithm to a piecewise constant solution. It is not clear whether the TV-norm or the l_0-norm is more effective in producing a sparse solution.**

***Index Terms*—Inverse problem, optimization, total-variation minimization, l_0-norm minimization, few-view tomography, iterative algorithm, image reconstruction**

I. Introduction

COMPRESSED sensing is a technology to use under-sampled measurements for solving an inverse problem [1-4]. When the measurements are insufficient, the imaging system is underdetermined, and the object to be imaged is not well defined. Extra information is required for a useful reconstructed image. For certain objects, extra information is available. The sparseness of an object is an important piece of information. An image array is said to be sparse if most of the array elements are zero. For example, the edge image of a piecewise constant image is sparse.

Let us consider the following situation. The desired solution, X, of an inverse problem is sparse after a certain sparsifying transformation Ψ, that is, $\Psi(X)$ is sparse. The sparseness is measured by the l_0 norm, $\|\cdot\|_0$. If X is a vector,

$$\|X\|_0 \triangleq \text{the total number of nonzero elements in } X. \quad (1)$$

If X is a scalar, $\|X\|_0$ is a binary number as

$$\|X\|_0 \triangleq \begin{cases} 0, & if\ \mathrm{X} = 0 \\ 1, & if\ \mathrm{X} \neq 0. \end{cases} \quad (2)$$

An under-determined inverse problem has infinitely many solutions. A compressed sensing solution is a solution with a minimum $\|\Psi(X)\|_0$ value.

Mathematically speaking, the l_0 norm is not really a norm nor even a pseudo-norm, because, in general,

$$\|cX\|_0 \neq |c| \cdot \|X\|_0. \quad (3)$$

In this paper, we use the term 'norm' loosely.

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).

Y. Li is with the Department of Mathematics, Utah Valley University, Orem, UT 84058 USA.

The difficulty of using the l_0-norm is that the l_0-norm is non-differentiable. Many researchers have attempted to tackle the l_0-norm minimization problem. The work in [5] converted the l_0-norm minimization problem into an unconstrained augment Lagrange problem. The work in [6] solved the l_0-norm minimization problem by introducing auxiliary variables. The l_0-norm minimization solution used in [7] involved a hard thresholding, linearization, and proximal points techniques. The l_0-norm minimization is usually achieved by solving some sub problems [8].

Due to the difficulty in minimizing the l_0-norm, the total variation (TV) norm minimization, which is the l_1-norm of the gradient, has been given more attention [9-11]. This is because the implementation of the l_1-norm optimization is much easier than the l_0-norm optimization. The main purpose of this current paper is to present a simple way to implement the l_0-norm optimization.

II. Methods

A. The l_0-norm and proposed meta l_0-norm

In our two-dimensional (2D) image reconstruction applications, we choose the finite differences in the row and column directions as the sparsifying transformation Ψ. Let the 2D image be X and its pixels be denoted as $x_{i,j}$, where i is the row index and j is the column index. At each pixel, $\Psi(x_{i,j})$ is a finite difference gradient, which is a 2D vector

$$\Psi(x_{i,j}) \triangleq \begin{bmatrix} x_{i+1,j} - x_{i,j} \\ x_{i,j+1} - x_{i,j} \end{bmatrix}. \quad (4)$$

The optimization metric (i.e., the objective function) can be set up as the sum of the l_0-norms of all pixels in the image as

$$M(X) = \sum_i \sum_j \left\| \begin{bmatrix} x_{i+1,j} - x_{i,j} \\ x_{i,j+1} - x_{i,j} \end{bmatrix} \right\|_0. \quad (5)$$

One difficulty of minimizing (5) is that the l_0-norm is not differentiable and is not easy to optimize. One common method to mitigate the difficulty is to use the l_1-norm approximation [9-13]. The famous total variation (TV) method is to use the l_1-norm of the finite difference gradient to replace the l_0-norm of the finite difference gradient. For a 2D image X, the isotropic TV norm is defined as

$$TV_{iso}(X) = \sum_i \sum_j \left\| \begin{bmatrix} x_{i+1,j} - x_{i,j} \\ x_{i,j+1} - x_{i,j} \end{bmatrix} \right\|_1 = \sum_i \sum_j \sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2}, \quad (6)$$

and the anisotropic TV norm is defined as

$$TV_{aniso}(X) = \sum_i \sum_j \left(|x_{i+1,j} - x_{i,j}| + |x_{i,j+1} - x_{i,j}| \right). \quad (7)$$

This paper proposes an alternative 'norm' to replace the l_0-norm. We refer to the proposed 'norm' as the meta l_0-norm, which is defined for a scalar X as

$$\|X\|_{meta0} \triangleq 1 - e^{-a|X|}, \quad (8)$$

where $0 < a < \infty$ is user-defined parameter. When $X = 0$, $\|X\|_{meta0} = \|X\|_0 = 0$. When $X \neq 0$, $\|X\|_{meta0} \approx \|X\|_0 = 1$ if a is chosen large enough. Figure (1) illustrates the approximation of $\|X\|_0$ by proposed $\|X\|_{meta0}$.

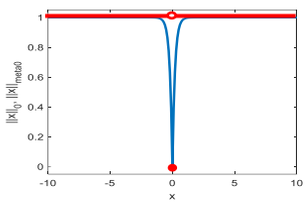


Figure 1. (Blue) $\|X\|_{meta0}$ with $a = 5$; (Red) $\|X\|_0$.

For a 2D vector $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, we define the *anisotropic* meta l_0-norm as

$$\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_{anisoMeta} \triangleq \left(1 - e^{-a|x_1|}\right) + \left(1 - e^{-a|x_2|}\right), \quad (9)$$

and define the *isotropic* meta l_0-norm as

$$\left\| \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \right\|_{isoMeta} \triangleq 1 - e^{-a\sqrt{x_1^2 + x_2^2}}. \quad (10)$$

For a 2D image X, discarding the constant term, we replace (9) by the anisotropic metric:

$$M_{aniso}(X) = -\sum_i \sum_j \left(e^{-a|x_{i+1,j} - x_{i,j}|} + e^{-a|x_{i,j+1} - x_{i,j}|}\right), \quad (11)$$

or the isotropic metric with the constant term discarded:

$$M_{iso}(X) = -\sum_i \sum_j e^{-a\sqrt{(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2}}. \quad (12)$$

The subdifferential exists for the proposed metric (11) as

$$\begin{aligned} \frac{\partial M_{aniso}(X)}{\partial x_{i,j}} = & -a \times e^{-a|x_{i+1,j} - x_{i,j}|} sgn\left(x_{i+1,j} - x_{i,j}\right) \\ & -a \times e^{-a|x_{i,j+1} - x_{i,j}|} \times sgn\left(x_{i,j+1} - x_{i,j}\right) \\ & +a \times e^{-a|x_{i,j} - x_{i-1,j}|} \times sgn\left(x_{i,j} - x_{i-1,j}\right) \\ & +a \times e^{-a|x_{i,j} - x_{i,j-1}|} \times sgn\left(x_{i,j} - x_{i,j-1}\right), \end{aligned} \quad (13)$$

where the sign function in (13) is defined as

$$sgn(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0, & \text{if } x = 0 \\ -1, & \text{if } x < 0. \end{cases} \quad (14)$$

The subdifferential given in (13) readily leads to a gradient descent algorithm to minimize the metric given in (11). The subdifferential for (12) can be similarly obtained.

B. Application: Few-view tomography

Here we consider a 2D parallel-beam imaging system, where the image array was 256 × 256, the detector contained 256 bins, and the number of views over 180° was 30.

The noiseless projection line integrals were generated analytically (without using pixels). A method based on 'projections onto convex sets' (POCS) was selected to reconstruct the image. This method alternated between the maximum-likelihood expectation-maximization (MLEM) image reconstruction algorithm and a gradient descent algorithm that minimized the metric $M_{anisoMeta}(X)$ or $M_{isoMeta}(X)$. There were 200 iterations used in the POCS algorithm. At each POCS iteration, there were 5000 iterations for the gradient descent algorithm. A flow chart of the algorithm is illustrated in Fig. 2.

In the computer simulations, the Shepp-Logan phantom was used [14]. In addition to the noiseless data set, a noisy data set was also generated with the zero-mean Gaussian noise added.

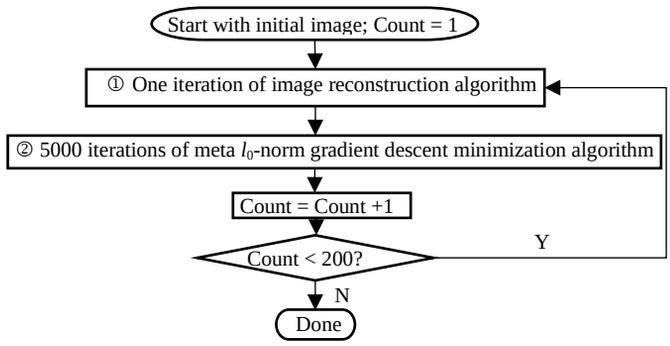


Figure 2. A flowchart of the algorithm used in computer simulations.

In Fig. 2, the image reconstruction algorithm ① is the MLEM algorithm:

$$x_{i,j}^{(k+1)} = \frac{x_{i,j}^{(k)}}{\sum_m a_{i,j,m}} \sum_m a_{i,j,m} \frac{p_m}{\sum_{\hat{i},\hat{j}} a_{\hat{i},\hat{j},m} x_{\hat{i},\hat{j}}^{(k)}}, \quad (15)$$

where p_m is the mth projection, $a_{i,j,m}$ is the projection contribution from the pixel (i,j) to the projection bin m, and k is the iteration index. In fact, the user can choose any iterative image reconstruction algorithm for algorithm ①. The gradient descent algorithm ② in Fig. 2 is given as

$$x_{i,j}^{(k+1)} = x_{i,j}^{(k)} - \eta \frac{\partial M_{(an)iso}(X)}{\partial x_{i,j}}, \quad (16)$$

where $\eta = 2 \times 10^{-7}$ in our computer simulations, and $\frac{\partial M_{aniso}(X)}{\partial x_{i,j}}$ is defined by (13) or $\frac{\partial M_{iso}(X)}{\partial x_{i,j}}$. The number of iterations shown in Fig. 2 is served as an example only.

A TV constrained image reconstruction algorithm was also implemented for the comparison purposes. The TV implementation was in the same format as depicted in Fig. (2), except that the gradient $\frac{\partial M_{(an)iso}(X)}{\partial x_{i,j}}$ was calculated with the TV norm (6) or (7).

III. Results

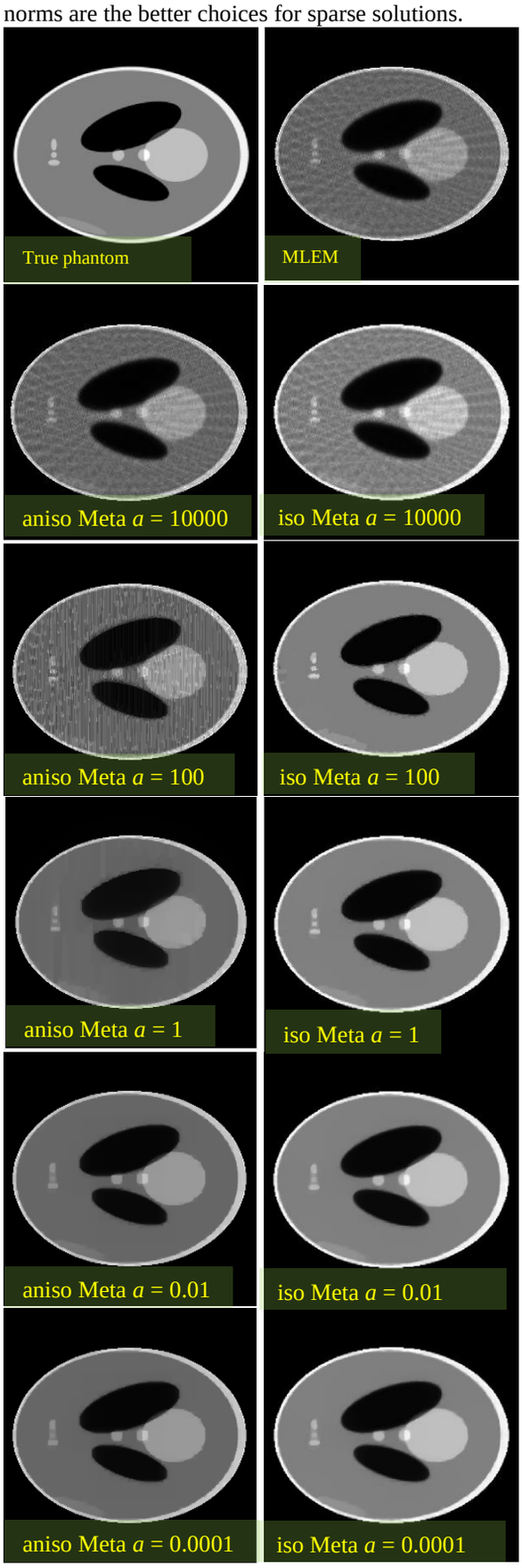
The reconstructions via the MLEM, anisotropic/isotropic TV, and the proposed anisotropic/isotropic meta l_0 method are shown in Figs. 3 and 4, respectively, for the noiseless and noisy data. Tables 1-4 show the quantitative comparison studies with structure similarity (SSIM), peak signal-to-noise ratio (PSNR), and signal-to-noise ratio (SNR).

When the parameter '*a*' is small, the proposed meta l_0-norms behave like TV norms. On the other hand, when the parameter '*a*' is large, the meta l_0-norms behave like the l_0-norms. Thus, the proposed meta l_0-norms are snapshots of the morphing from the TV norm to the l_0-norm for different '*a*' values.

When a = 10000, the numerical values of the exponential function become underflow. The constraints are not effective and are ignored.

It is observed that the l_0-norms give better SSIM results. However, the TV norms give better PSNR and SNR results. Therefore, we cannot conclude whether the TV norms or the l_0-

norms are the better choices for sparse solutions.



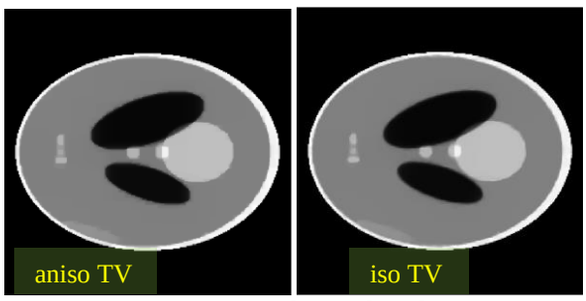
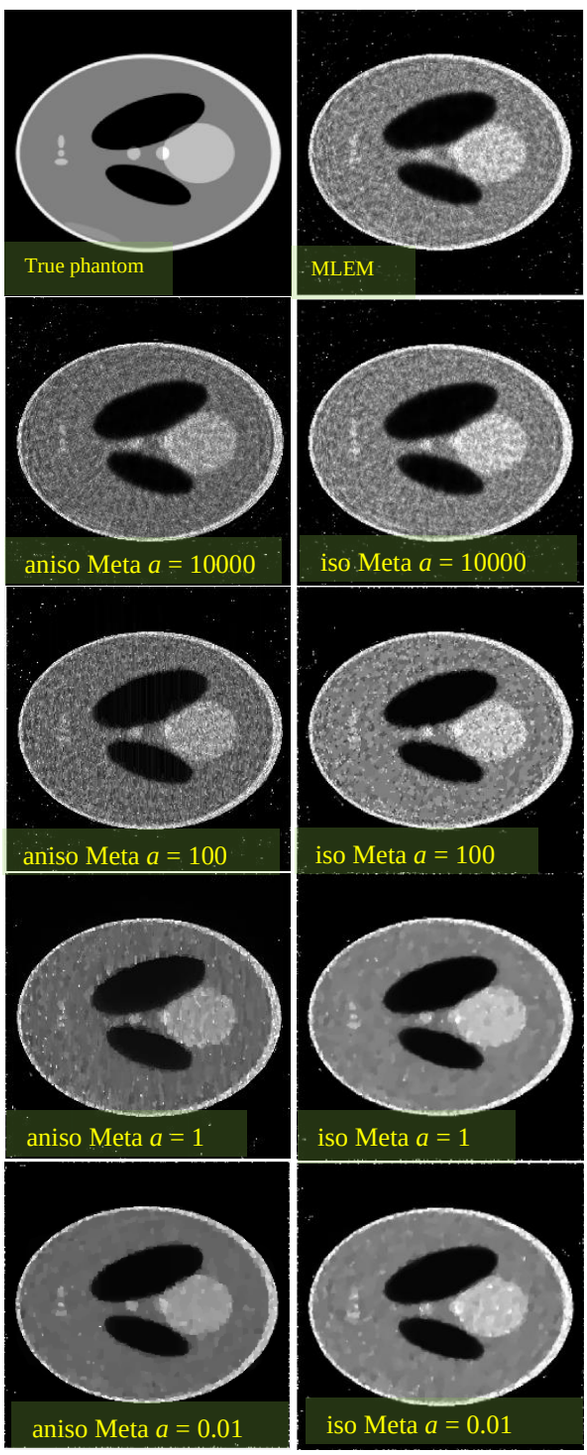


Fig. 3. Reconstructions with noiseless data.



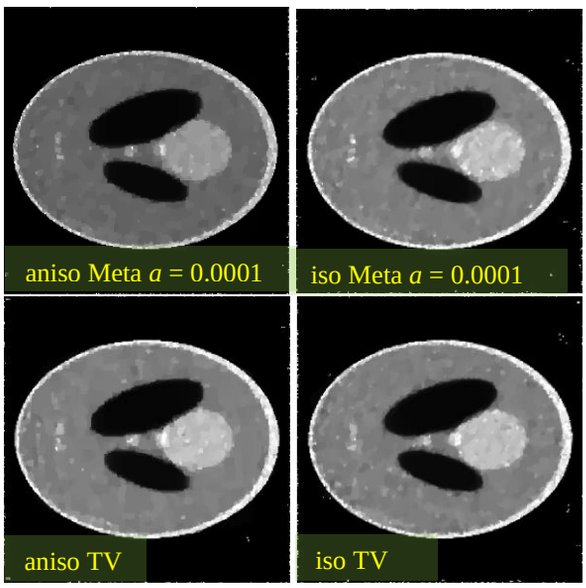


Fig. 4. Reconstructions with noisy data.

Table 1. Comparison studies with noiseless data (**anisotropic**)

Method	SSIM	PSNR	SNR
Ideal case	1.00	∞	∞
MLEM	0.8771	32.8887	16.5828
anisoMeta a=10000	0.8745	32.8744	16.5684
anisoMeta a=100	0.8059	31.6825	15.3766
anisoMeta a=1	0.9120	34.1341	17.8281
anisoMeta a=0.01	0.9474	36.9150	20.6090
anisoMeta a=0.0001	**0.9475**	36.9199	20.6199
anisoTV	**0.9475**	**36.9199**	**20.6139**

Table 2. Comparison studies with noiseless data (**isotropic**)

Method	SSIM	PSNR	SNR
Ideal case	1.00	∞	∞
MLEM	0.8771	32.8887	16.5828
isoMeta a=10000	0.8775	32.8967	16.5908
isoMeta a=100	**0.9557**	34.9854	18.6795
isoMeta a=1	0.9509	36.7247	20.4187
isoMeta a=0.01	0.9514	37.4011	21.0951
isoMeta a=0.0001	0.9514	37.4094	**21.1035**
isoTV	0.9514	**37.4095**	**21.1035**

Table 3. Comparison studies with noisy data (**anisotropic**)

Method	SSIM	PSNR	SNR
Ideal case	1.00	∞	∞
MLEM w/ noise	0.6522	28.1125	11.8066
anisoMeta w/ noise, a=10000	0.6507	27.9083	11.6023
anisoMeta w/ noise, a=100	0.6039	26.9093	10.6033
anisoMeta w/ noise, a=1	0.7585	28.6529	12.3470
anisoMeta w/ noise, a=0.01	0.8667	30.5347	14.2287
anisoMeta w/ noise, a=0.0001	**0.8709**	30.4943	14.1883
anisoTV w/ noise	0.8663	**30.7298**	**14.4238**

Table 4. Comparison studies with noisy data (**isotropic**)

Method	SSIM	PSNR	SNR
Ideal case	1.00	∞	∞
MLEM w/ noise	0.6522	28.1125	11.8066
isoMeta w/ noise, a=10000	0.6467	27.9845	11.6785
isoMeta w/ noise, a=100	0.6860	26.9188	10.6128
isoMeta w/ noise, a=1	0.8511	29.8379	13.5319
isoMeta w/ noise, a=0.01	0.8605	30.1981	13.8921
isoMeta w/ noise, a=0.0001	**0.8638**	**30.3116**	**14.0056**
isoTV w/ noise	0.8625	30.3089	14.0029

IV. Discussion and Conclusions

Simple meta l_0-norms are suggested in this paper to replace the l_0-norm in searching a sparse solution for an inverse problem with under-sampled data. These meta norms have a user defined parameter 'a'. The meta l_0-norms behave like the l_0-norm when 'a' is large. The most significant advantage of the meta l_0-norms is that it is subdifferentiable. Therefore, it is straightforward to derive a gradient descent algorithm to minimize the meta l_0-norms. Computer simulations in this paper show that the meta l_0-norms are effective in producing a piecewise solution. The meta norms are snapshots between the TV norms and the l_0-norms. The choice of a better norm is task dependent. The l_0-norms are not always preferred.

We must point out that the l_0-norm minimization problem is far from being solved. The l_0-norm minimization problem is NP-hard. The gradient descent algorithm to minimize the meta l_0-norms is merely a greedy approach. The l_0-norm minimization problem has multiple minima. Any gradient-based methods can only find a local minimum.

References

[1] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal 1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006. doi:10.1002/cpa.20132

[2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006. doi:10.1109/TIT.2006.871582

[3] E. J. Candès, J. K. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information," *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 489–509, 2006. *doi:10.1109/tit.2005.862083*. *S2CID 7033413*

[4] L. I. Rudin, S. Osher and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–68, 1992.

[5] H. Chen, J. Tao, Y. Sun, Z. Ye and B. Qiu, "Magnetic resonance image reconstruction via L_0-norm minimization," *2015 IET International Conference on Biomedical Image and Signal Processing* (ICBISP 2015), Beijing, 2015, pp. 1-6, doi: 10.1049/cp.2015.0780.

[6] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia, "Image smoothing via L_0 gradient minimization," *ACM Transactions on Graphics*, vol. 30, no. 6, Article 174, Publication date: December 2011.

[7] Y. Sun and J. Tao. "Few views image reconstruction using alternating direction method via L_0-minimization," *International Journal of Imaging Systems and Technology*, 24, pp. 214-223, (2014).

[8] Wei Yu and Li Zeng, "ℓ_0 gradient minimization based image reconstruction for limited-angle computed tomography," *PLOS ONE*, vol. 10, no. 7, Article e0130793, 2015. https://doi.org/10.1371/journal.pone.0130793

[9] Xuan Fei, Zhihui Wei, and Liang Xiao, "Iterative directional total variation refinement for compressive sensing image reconstruction," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1070–1073, 2013. *doi:10.1109/LSP.2013.2280571*

[10] E. Y. Sidky and X. Pan. "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization." *Physics in Medicine & Biology*, vol. 53, pp. 4777 – 4807, 2008.

[11] V. P. Panin, G. L. Zeng, and G. T. Gullberg, "Total variation regulated EM algorithm," *IEEE Trans. Nucl. Sci.*, vol. 46, no. 6, pp. 2202-2210, 1999.

[12] G. L. Zeng, Y. Li and A. Zamyatin, "Iterative total-variation reconstruction vs. weighted filtered-backprojection reconstruction with edge-preserving filtering," *Phys. Med. Biol.*, vol. 58, pp. 3413-3431, 2013.

[13] G. L. Zeng, "A projection-domain iterative algorithm for metal artifact reduction by minimizing the total-variation norm and the negative-pixel energy," *Visual Computing for Industry, Biomedicine, and Art*, vol. 5, 1, 2022.

[14] L. A. Shepp and B F Logan, "The Fourier reconstruction of a head section," *IEEE Transactions on Nuclear Science*, vol. NS-21, no. 3, pp. 21–43, 1974. doi:10.1109/TNS.1974.6499235

CT Brain Perfusion Assessment using X-ray Angiography Images

Shalini Subramanian[1], W. Paul Segars[2], Ting-Yim Lee[3,4], Andreia. V. Faria[1], and Katsuyuki Taguchi[1]

***Abstract*— The ability to assess brain perfusion in real time could significantly improve the application of interventional techniques, particularly endovascular thrombectomy (EVT), in treating ischemic strokes. This potential is not fully realized due to the lack of intraoperative real-time perfusion imaging. Our novel approach, the intra-interventional perfusion with no C-arm rotation (IPEN) method, aims to utilize x-ray angiography images to perform in-depth brain perfusion analysis in the interventional suite. IPEN is constructed on tomographic principles. It enables estimating the values of *N* homogenous 3D regions-of-interests (ROI) of known shapes using *N* x-ray beams from a single projection. Initially, to develop and evaluate IPEN-based brain perfusion assessment, a perfusion XCAT phantom was used. IPEN was initially evaluated for computer-simulated data on eight ischemic stroke cases. Following this, as a part of a pilot clinical study, IPEN was evaluated on seven clinical datasets. Dynamic enhanced bi-plane x-ray angiography images with intra-arterial contrast injection were used for each scenario. IPEN was employed to estimate time-enhancement curve (TEC) and key perfusion indices (PIs) for the predefined ROIs in the brain. The results obtained were rigorously validated against known standards for both studies. The normalized root mean square error (NRMSE) and linear correlation (R value) for TEC and PI were computed for the simulation study. In both studies' the estimated TEC and PI showed good agreement with ground truth values. For the computer simulation study, the NRMSE values were 13.8±1.9% for TEC, 15.1±3.1% for time-to-maximum (Tmax) and 16.0±4.0% for relative cerebral bold flow (rCBF). The R values were 0.92, 0.80, and 0.84 for TEC, Tmax and rCBF respectively. In the clinical study, we saw good agreement between the estimated and actual lesion size and location. The time-to-peak values also significantly differed between ischemic and non-ischemic regions, emphasizing the study's promising outcomes.**

***Index Terms*—brain perfusion imaging, IPEN, endovascular thrombectomy, perfusion index, time enhancement curve, x-ray angiography**

I. Introduction

STROKE is the second leading cause of death and the leading cause of severe disability in the United States. [1]

Interventional treatment techniques such as endovascular thrombectomy (EVT) have shown promising outcomes in treating ischemic stroke. Real-time brain tissue perfusion imaging can be highly beneficial in assessing the risk–benefit of interventional treatments and the viability of downstream vascular beds. In addition, analyzing contrast enhancement before EVT also helps provide critical information about the size of the infarct core. It can also help evaluate the success during and after the procedure.

The current standard of care employs X-ray computed tomography (CT) or magnetic resonance perfusion imaging only at initial patient admission due to their inaccessibility in the interventional suite. In its place, in the interventional suite, x-ray angiography (XA) images are used to assist interventional procedures. XA images provide high temporal resolution and assist in delineating arterial blockage and recanalization. Nonetheless, assessing perfusion volumetrically in smaller brain regions is challenging due to overlapping regions, mainly when an ischemic area is obscured by surrounding healthy, enhanced brain tissue. Alternative CT-based approaches have been investigated to overcome this limitation; nevertheless, these techniques cannot achieve high temporal resolution as XA images. The higher radiation exposure, equipment cost, and less-than-ideal workflow render these alternatives less favorable.

To address this, we propose an intra-interventional perfusion with no C-arm motion (IPEN)-based method to perform quantitative volumetric brain tissue perfusion analysis using XA images in the interventional suite. This approach relies solely on the standard XA images obtained in the interventional suite, thus preserving the current interventional workflow without disruptions.

II. Methods

We outline the proposed IPEN algorithm for brain perfusion analysis in section A, the computer simulation-based evaluation study in section B, and a pilot clinical study in section C.

A. IPEN algorithm and workflow

Tomography is the backbone of IPEN workflow. The concept is that when an object is divided into two homogenous regions-of-interest (ROIs) of known thickness, two x-ray beams from a single view are sufficient to estimate their values. Expanding on this concept, *N* x-ray beams at a single view angle are sufficient to estimate values of *N* homogeneous 3D ROIs. This is possible even if ROIs overlap along the x-ray paths if co-linearity is avoided. [2]

The IPEN algorithm for assessing brain perfusion consists of four major steps (Figure 1): (1) obtain patient-specific ROIs using arterial territory atlas; (2) align CT head voxels with XA image; (3) estimate dynamic time-enhancement curves (TEC) for each ROI and the arterial input function (AIF) and (4) estimate perfusion indices (PIs). This study focuses on steps (1), (3), and (4), improving ROIs in arterial territory atlas, estimating dynamic TEC and AIF, followed by PIs such as time-to-maximum (Tmax) and relative cerebral blood flow

[1] S. Subramanian, A. V. Faria, and K. Taguchi are affiliated with Johns Hopkins University

[2] W. P. Segars is affiliated with Duke University

[3,4] Ting-Yim Lee is affiliated with University of Western Ontario, Lawson Health Research Institute

(rCBF). A brief overview of steps 1, 3, and 4 is described in the following paragraphs.

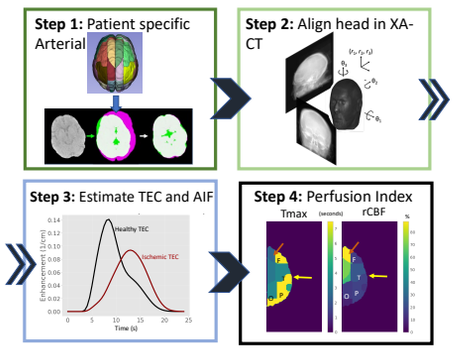


Figure 1: IPEN algorithm workflow

In step (1), the arterial atlas from [3] was modified to increase the ROI count while decreasing each ROI volume. Patient-specific brain ROIs were generated by performing 3D affine registration of the arterial atlas with the pre-operative CT volume.

In step (3), dynamic TEC was estimated by solving an optimization problem [2]. The object modeling revolved around the basic principle of IPEN, i.e., homogeneous enhancement in each ROI. Let $\boldsymbol{x_t}$ be the vector for the object's linear attenuation coefficient at time $\boldsymbol{t}$. Then, $\boldsymbol{x_t}$ can be expressed as the summation of pre-contrast value $\boldsymbol{x_0}$ and ROI enhancement value $\boldsymbol{z_t}$, $\boldsymbol{x_t} = \boldsymbol{x_0} + \boldsymbol{A_{z2x}z_t}$, where $\boldsymbol{A_{z2x}}$ operator converts ROI values to voxel values in the object. The projection $\boldsymbol{y_t}$ of object $\boldsymbol{x_t}$ can be expressed as, $\boldsymbol{y_t} = \boldsymbol{A_x}(\boldsymbol{x_0} + \boldsymbol{A_{z2x}z_t}) = \boldsymbol{A_x x_0} + \boldsymbol{A_z z_t} = \boldsymbol{y_0} + \boldsymbol{A_z z_t}$, where $\boldsymbol{A_x}$ is the forward projector for a pre-contrast object, $\boldsymbol{A_z}$ is a forward and $\boldsymbol{y_0}$ projector is the pre-contrast projection. Using measured projections at $t = 0$ and t, $\widetilde{\boldsymbol{y_0}}$ and $\widetilde{\boldsymbol{y_t}}$, respectively, the ROI enhancement vector at t, $\boldsymbol{z_t}$, was estimated by minimizing,

$$\widehat{\boldsymbol{z_t}} = argmin\, ||\boldsymbol{A_z z_t} + \widetilde{\boldsymbol{y_0}} - \widetilde{\boldsymbol{y_t}}\,||^2_W,$$

where W is a statistical weighting by the inverse variance approximated by the measurement. The separable paraboloidal surrogates (SPS) optimizer minimizes the cost function. Next, the AIF, $\boldsymbol{C_a}$, was computed using identified artery pixels in XA image.

Subsequently, in step (4), the perfusion index estimation comprised two steps: estimating the residue function $\boldsymbol{R_f}$ and computing two PIs, Tmax and rCBF, from the residue function. The residue function for each ROI can be defined as the tissue response function at t in response to a unit of contrast bolus injected into an artery supplying the ROI as a delta function at $t = 0$ [4]. $\boldsymbol{R_f}$ was estimated by solving an optimization problem to deconvolve AIF from TEC estimated in step (3), as discussed below.

The object modeling was adapted from the theoretical physiological model for tissue perfusion in intravascular tracer systems [5] for ROI-based perfusion analysis. This approach assumes that each ROI is homogeneous and the contrast transport in the tissue is stationary. Then, the contrast concentration in the tissue can be expressed as,

$$\boldsymbol{z_t} = \boldsymbol{C_a} \otimes \boldsymbol{R_f},$$

where $\otimes$ is a convolution operator.

By analyzing the shape of TEC obtained in Step (4), we have determined $\boldsymbol{R_f}$ to be modeled as an asymmetric Gaussian function. The parameters estimated were the mean μ, the standard deviations σ_1 and σ_2, the scaling factor a, and the bias b:

$$\boldsymbol{R_f} = a * \boldsymbol{G} + b,$$

$$\text{where } \boldsymbol{G} = \begin{cases} \exp\left(\frac{-(t-\mu)^2}{2\sigma_1^2}\right) & t - \mu \le 0 \\ \exp\left(\frac{-(t-\mu)^2}{2\sigma_2^2}\right) & t - \mu > 0 \end{cases}.$$

The residue function parameters for each ROI j were estimated by minimizing the difference between the estimated ROI TEC vector $\widehat{\boldsymbol{z}}_j$ and the convolution of residue function with estimated AIF, $\boldsymbol{C_a}$ using the conjugate gradient method. Residue function vector was obtained by minimizing:

$$\widehat{\boldsymbol{R}}_{f,j} = argmin_{\mu,\sigma_1,\sigma_2,a,b} \left\| \widehat{\boldsymbol{z}}_j - \left(\boldsymbol{C_a} \otimes \boldsymbol{R_f}\right) \right\|^2.$$

Using this estimated $\widehat{\boldsymbol{R}}_{f,j}$, the PIs, Tmax, and rCBF were obtained. Tmax is the time taken from the bolus injection to its maximum value in the tissue, i.e., the time corresponding to the maximum value of $\widehat{\boldsymbol{R}}_f$. The CBF is computed as the maximum slope of the function $\widehat{\boldsymbol{R}}_f$ and rCBF is obtained relative to the maximum CBF value amongst all ROIs in the brain.

B. Computer simulation study

This section outlines the evaluation study using computer simulation and the data analysis performed.

We used the perfusion XCAT phantom to simulate dynamic contrast-enhanced head CT volumes [6]. Eight acute ischemic stroke cases with thrombus occlusion in the left and right brain were simulated. In each patient, a thrombus occluded one of the four arteries—the middle cerebral (MCA), anterior cerebral, anterior temporal, and posterior parietal —and dynamic intra-arterial contrast-enhanced volumes were simulated. A bi-plane angiography system was used with two flat panel detectors of 512×512 pixels with a pixel size of 0.8 mm (projected at the iso-center). A controlled intra-arterial injection of contrast agent containing 2,000 mg iodine was administered over 4–6 seconds. XA images were acquired over 25 seconds at two frames/second. Five noisy XA images were generated for each case by converting line integrals to photon counts and adding Poisson noise. Using these XA images, IPEN was performed to estimate the TEC and PI parameters.

In this study, we implemented all steps in IPEN except step (2), assuming that the XCAT phantom head was registered to XA images accurately.

The accuracy of the estimated TEC was quantitatively evaluated by comparing it to ground truth values using the normalized root-mean-square error adjusted for the volume proportion of each ROI (NRMSE) and a linear correlation coefficient (R value). Similarly, the accuracy of estimated PI (Tmax and rCBF) was assessed against ground truth using both NRMSE and R value.

C. Pilot Clinical Study

In this section, we discuss the specifications of the data used in the pilot clinical study, followed by the evaluation of IPEN on this dataset.

This study retrospectively accessed ten patient datasets who underwent admission CT, CT perfusion, EVT, and follow-up

CT. Amongst the 10 cases, two cases (20%) displayed large lesions, four cases (40%) had medium-sized lesions, and four cases (40 %) had small lesions. Three cases (30%) had strong artifacts and were excluded. The developed IPEN algorithm was implemented with minor adaptations to accommodate the challenges faced with clinical data. Any small head motion between frames was not compensated to simplify the translation from the computer simulation study. In step (4) of the IPEN algorithm, the only perfusion index estimated was time-to-peak of the estimated TEC (T2P).

Admission CT perfusion images were used to establish the ground truth as follows. Regions with Tmax values ≥ 6 seconds were identified as ischemic; regions with Tmax < 6 seconds non-ischemic. Using the identification, ROIs used in IPEN were defined as ischemic ROIs (or non-ischemic ROIs). Finally, an ischemic volume size was measured for each patient.

IPEN results were assessed as follows. An ROI was called ischemic if the T2P value was larger than 5 seconds. An ischemic volume size for IPEN was a summation of ischemic ROI volumes.

The accuracy of IPEN perfusion was assessed as follows. A Student's *t*-test was performed to assess if there was a statistically significant difference between T2P values of ischemic ROIs and those of non-ischemic ROIs. A linear correlation coefficient was calculated between the ischemic lesion sizes measured by CT perfusion and those by IPEN. A P value less than 0.05 indicated significance.

III. Results and Discussion

A. Computer simulation study

This section focuses on IPEN evaluation using perfusion XCAT phantom and the results of estimated TEC and PIs with respect to the ground truth values.

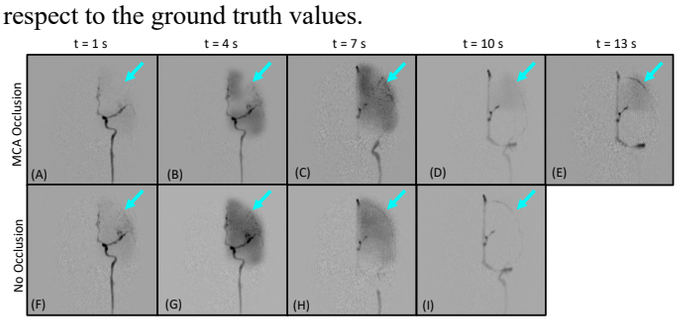


Figure 2: DSA images of MCA occlusion (A-E) and healthy (F-I) cases generated using perfusion XCAT phantom. The blue arrows show the affected ischemic region (A-E) and the same region with no lesion (F-I).

Figure 2 presents the digital subtraction angiography (DSA) images of perfusion XCAT phantom with MCA occlusion case (A–E) and without any occlusion (F–I) over 15 seconds at a time interval of 3 seconds. In Figure 2(A–E, arrows), a large ischemic region presents delayed and lower contrast enhancement, whereas in Figure 2(F–I, arrows), the healthy case had neither. The healthy DSA images show faster and uniform enhancement throughout the brain.

For the MCA occlusion case above, dynamic contrast-enhanced 3D CT maps were generated using each ROIs true and estimated TEC values. Figure 3(a) shows the axial slice of the contrast-enhanced CT maps at three time points across the TEC measured. First, at the initial contrast uptake phase; second, at the peak of a healthy ROI; and last, at the peak of an average ischemic ROI. Qualitatively, Figure 3(a, b), presents excellent agreement between the estimated and true contrast-enhanced CT maps TEC curves.

Figure 3(c) presents the scatter plot for all ROIs for all 8 cases across multiple scans. The fitted line and R value demonstrated that the performance of IPEN was consistent across different cases. Qualitatively, the estimated TEC showed good agreement with the true TEC values. The NRMSE and R values computed for TEC over eight cases and multiple scans are shown in the first row of Table 1. The standard deviation represents the range of all data analyzed. We see a low NRMSE value and a large R value, indicating IPEN's reliability. Considering perfusion imaging is currently not feasible in the interventional suite, the NRMSE and R values in Table 1 are good and show promising value.

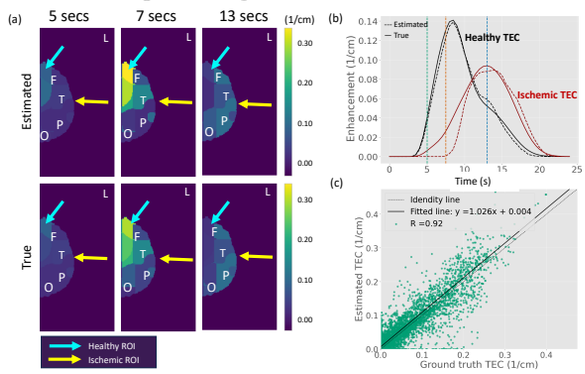


Figure 3: (a) Contrast-enhanced CT maps of the MCA occlusion case, synthesized from the estimated TECs of ROIs (top row) and the true ROI values calculated from (heterogenous) voxel values (bottom row). The frontal/temporal/occipital ROIs (F/T/O, yellow arrows) supplied by MCA are significantly less perfused compared to unaffected frontal ROIs (F, blue arrows) supplied by ACA, which were accurately delineated by the proposed method. (b) The estimated and true enhancement in healthy ("F" in (a)) and ischemic ("F/T/O" in (a)) brain ROIs of the MCA occlusion case. Vertical lines represent three time points presented in (a). (c) shows the scatter plot of all estimated enhancement values plotted over the true values of all brain ROIs for all eight cases.

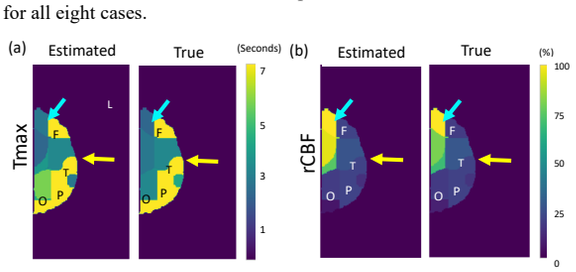


Figure 4 Perfusion index maps, Tmax (left, a), and rCBF (right, b), for the MCA occlusion case. (a) The frontal/temporal/occipital ROIs (F/T/O, yellow arrows) supplied by MCA show significantly delayed Tmax compared to unaffected frontal ROI (F blue arrows). (b) The ischemic ROIs ("F/T/O") show substantially lower rCBF values compared to the unaffected healthy ("F") brain ROIs.

In step (6), the estimated PIs were used to generate ROI-based 3D perfusion maps. Figure 4 shows the estimated and true Tmax and rCBF values for the MCA occlusion case. In Figure 4(a), Tmax is significantly larger for ischemic ROIs supplied by MCA (>6 secs, yellow arrow) compared to healthy ROIs supplied by ACA (2-5 secs, blue arrow). The quantitative analysis confirmed this with low NRMSE and high R values, as seen in Table 1.

Table 1: Quantitative analysis of estimated values

Metric	NRMSE (%)	R value
TEC	13.8 ± 1.9	0.92 ± 0.04
Tmax	15.1 ± 3.1	0.80 ± 0.17
rCBF	16.0 ± 4.2	0.84 ± 0.17

B. *Pilot Clinical study*

In this section, we will discuss the results obtained from the initial pilot clinical study performed. First, we use one case with a medium-sized lesion. Figure 5(A–E) shows the initial pre-EVT DSA images. Figure 5(F–I) shows post-EVT DSA images. The pre-EVT angiogram shows no clear view of any ischemic lesion due to overlapping regions in the angiograms. DSA images are effective in viewing vascular flow but not brain tissue perfusion with smaller ischemic lesions.

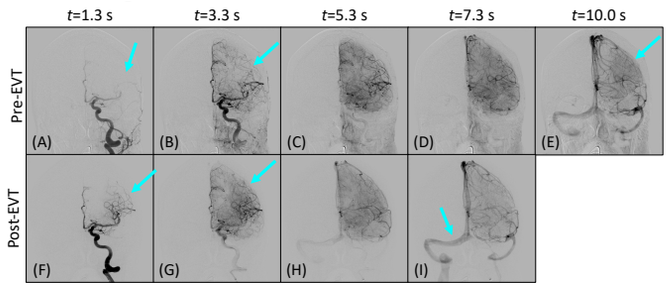


Figure 5: DSA images of a patient with a medium-sized ischemic lesion in the left MCA territory. The initial pre-EVT angiogram (top row) appears to have unobstructed blood flow, although the arterial flow seems slower (A, B). The ischemic lesion was indistinct in the images (C, D), with prolonged enhancement in the delayed venous phase (E). The post-EVT angiogram (bottom row) showed faster blood flow (F), enhancement (G), and venous drainage (I).

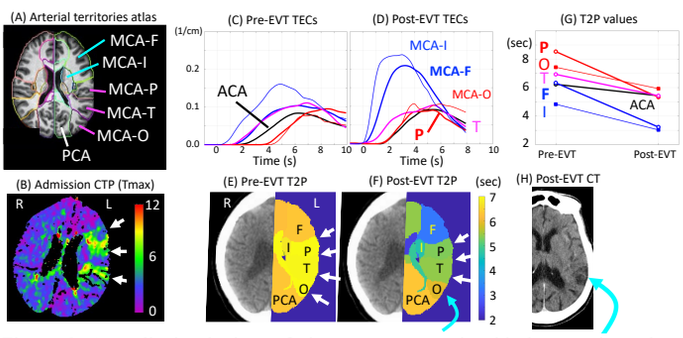


Figure 6: A qualitative brain perfusion assessment using bi-plane angiography images. (B) Admission CTP presents the medium-sized ischemic lesion over MCA-P/T. (C) The pre-EVT TECs showed a delayed enhancement in all the MCA territory ROIs, including MCA-T/P/O. (E) The pre-EVT T2P map showed an ischemic lesion. (D, F) The post-EVT assessment showed heterogeneous changes: MCA-F/I had sufficiently improved TECs and T2P, whereas MCA-P/T/O showed improved but still lower peaks and slower flow in TECs and high T2P values (>5.3 s, arrows). (G) The changes in T2P values between pre-EVT and post-EVT were noticeable but insufficient for MCA-P/T/O. (H) A 2-hour post-EVT NE-CT showed a hypodense lesion (blue arrow), which agreed with the persistently high T2P value at post-EVT (F, curved arrow).

Figure 6 shows the IPEN assessment of the medium-sized lesion case described above. In the pre-EVT T2P map (Figure 6(E)), ROIs MCA-P/T/O (white arrows) were identified as the affected ROIs. This ischemic region was larger than the region delineated by the admission CTP scan (Figure 6 (B)). A larger region could indicate a possible growth of the lesion in the 2-hour window before EVT. In addition, ROI MCA-F showed reduced perfusion, which may imply affected flow. The post-EVT assessment displayed a heterogeneous response to the thrombus retrieval. The T2P map in Figure 6(F) shows that ischemic ROIs MCA-P/T showed ample improvement compared to ROI regions MCA-O. In Figure 6(H), the 2-hour post-EVT non-contrast CT showed a hypo-dense region, corroborating our findings in the post-EVT T2P maps. ROIs MCA-F/I also show significant improvement, indicating restored flow.

For 34 ROIs in the MCA territory from seven patients, pre-EVT T2P values in ischemic regions were 6±2.5 seconds (N=18), compared to 3.8±1.4 seconds (N=16) in non-ischemic ROIs ($P < 0.01$). The linear correlation coefficient between ischemic lesion size in CT perfusion and IPEN was 0.69 ($P>0.05$). Considering a 2-hour difference between CTP and pre-EVT angiograms used in IPEN evaluation, this value is good.

One limitation of the IPEN algorithm is the possibility of heterogeneous enhancement and partial volume effect in pre-segmented ROIs. This heterogeneity may result in shorter delays or double Gaussian-shaped TEC curves in ischemic regions, especially if the ischemic region is a smaller portion of the ROI. Nonetheless, overall, the proposed IPEN method showed promising results, quantitatively and qualitatively, in both the simulation evaluation study and the pilot clinical study. The future work will focus on improving the algorithm and overall accuracy. In addition, we aim to perform IPEN-based assessments on more clinical datasets and compute PIs, Tmax, and rCBF for these cases.

IV. Conclusion

A four-step IPEN algorithm was developed and implemented to assess brain perfusion using XA images. The evaluation demonstrated its feasibility in the interventional suite. The performance of IPEN was evaluated for both simulated perfusion XCAT data and clinical data. The simulation study showed good accuracy with low NRMSE and high R values. Extending beyond simulations, in the pilot clinical study, IPEN continued to show promising results, achieving reasonable R values. This combined evaluation through simulation and clinical studies highlights IPEN's robustness and effectiveness in assessing brain perfusion. It also exhibited IPEN's potential applicability in real-world clinical settings.

V. Acknowledgments

This research was funded by R01 NS126256 and R21 EB029049. We thank Nafi Aygun for acquiring clinical data.

References

[1] Xu JQ, Murphy SL, Kochanek KD, Arias E. Mortality in the United States, 2018. NCHS Data Brief, no 355. Hyattsville, MD: National Center for Health Statistics. 2020.

[2] Taguchi K, Sauer TJ, Segars WP, et al. Three-dimensional regions-of-interest–based intra-operative four-dimensional soft tissue perfusion imaging using a standard x-ray system with no gantry rotation: A simulation study for a proof of concept. Med Phys. 2020;47(12):6087-6102.

[3] Liu C, Hsu J, Xu X, et al. Deep learning-based detection and segmentation of diffusion abnormalities in acute ischemic stroke. Communications Medicine. 2021;1(1):61

[4] Lee TY, Yang DM, Li F, Marants R, "CT perfusion techniques and applications in stroke and cancer," in Samei E and Pelc NJ (Ed.), Computed Tomography: Approaches, Applications, and Operations, Cham: Springer International Publishing, 2020: 347–365.

[5] Fieselmann A, Kowarschik M, Ganguly A, et al. Deconvolution-based CT, and MR brain perfusion measurement: Theoretical model revisited and practical implementation details. Journal of Biomedical Imaging. 2011; 2011:1-20.

[6] Divel SE, Christensen S, Segars WP, et al. A dynamic simulation framework for CT perfusion in stroke assessment built from first principles. Med Phys. 2021;48(7):3500-3510.

Optimizing Conditional DDPM for Head CT Motion Artifact Reduction: Brain vs. Skull and 3D vs. 2D

Zhennong Chen[1], Matthew Tivnan[1], Siyeop Yoon[1], Rui Hu[1], Quanzheng Li[1], and Dufan Wu[1]

Abstract— **In this study, we introduce a conditional Denoising Diffusion Probabilistic Model (DDPM) approach that employs motion-corrupted images generated by FBP as the condition to reduce motion artifacts in 3D head CT scans. We address two critical questions in this application. First, how can we overcome the disparate performance observed in the skull and brain regions, which is attributable to their distinct intensity ranges? Second, which is more effective for accommodating the 3D nature of head CT and head motion: a 3D or 2D DDPM backbone? The resolution of these questions guides us towards an optimized, image-domain-only DDPM method, demonstrating significant efficacy in reducing motion artifacts in head CT scans.**

Index Terms— **Head CT, Motion Correction, Diffusion Model**

I. Introduction

Head motion in paediatric patients or those with head trauma or stroke is a majoy source of image artifacts in head computed tomgoraphy (CT), degrading the image quality and impacting diagnosis. On the software side, solutions were developed to reduce motion artifacts either based on sinogram domain or based on image domain. The sinogram-domain-based solutions utilize projection data to estimate head motion during acquisition[1]–[4]. However, obtaining and processing projection data in daily clinical settings remains challenging.

In contrast, despite the lack of data fidelity, image-domain-based solutions don't require projection data and thus are more practical in daily use. These solutions typically employ a convolutional neural network (CNN) trained on pairs of motion-corrupted and motion-corrected images. For instance, Ko et al.[5] added attention modules to the U-Net to selectively amplify or attenuate residual features. Su et al.[6] integrated ResNet and dilation techniques into the UNet and show better performance over vanilla U-Net. Nevertheless, we observe empirically that their methods may over-smooth images in cases of moderate to severe 3D motion.

Recently, denoising diffusion probabilistic models (DDPM)[7] have shown remarkable success in various generative and inverse problems[8], [9], offering improved image quality and more stable training. Their ability to effectively incorporate conditioning information[10] motivates us to apply contidional DDPM to reduce motion artifacts in head CT, using motion-corrupted CT image generated by FBP as a condition. To find the most effective way to apply DDPM to this task, there are two key questions:

(1) **Managing distinct intensity ranges in skull and brain**: The skull and brain in head CT scans have markedly different intensity ranges, with the skull displaying higher values and a wider range (hundreds to >2000 HU) and the brain showing lower values and a narrower range (0 to 100 HU). As the reader will see in the **Figure 1 (row 3 and 5)**, direct implementation of DDPM to original head CT images cause much poorer performance in brain than skull. We hypothesize this disparate performance should be atributable to their distinct intensity ranges. How to solve this challenge? We propose to introduce an intensity cutoff in the image (i.e., [-100,100] HU, targeting only brain and background pixels), focusing both the training and the sampling process on brain. We then train two anatomy-specific DDPM models: θ_{skull} for the skull recovery and is trained using original CT images, and θ_{brain} for brain recovery and is trained using intensity-cutoff images. The final motion-corrected image combines predictions from θ_{skull} and θ_{brain}.

(2) **Accommodating the 3D nature of head CT and head motion**: The first solution is to use a 3D DDPM, converting all conbolution and attention layers to 3D and using the 3D CT volume as the model input. The second solution, inspired by Lee et al.[11] and Niu et al.[12], involves two 2D DDPM models, one trained on 2D axial planes (along z-axis) to learn intra-plane features (i.e., correct artifacts caused by the intra-plane motion) and the other trained on 2D perpendicular planes (can be either coronal or sagittal according to [11], [12]) to learn inter-plane features (i.e., inter-plane motion). The sampling is performed using both models in a switching mode. While the 3D model comprehensively processes intra- and inter-plane motion, it is computationally expensive and may lead to poor convergence. The 2D approach is less demanding in computation but its efficacy in reducing 3D head CT motion artifacts is yet to be established.

In summary, we aim to investigate two questions in this paper. Question 1, if introducing an intensity cutoff in the image preprocessing can resolve the disparate performance caused by the distinct intensity ranges of skull and brain. Question 2, whether a 3D or 2D DDPM backbone is more effective for 3D head CT motion reduction. The findings will contribute to developing a practical image-domain-only approach for leveraging conditional DDPM to significantly reduce motion artifacts in head CT scans.

[1]All Authors are with Center for Advanced Medical Computing and Analysis, Massachusetts General Hospital, Boston, USA. Corresponding author: Dufan Wu, dwu6@mgh.harvard.edu

II. METHODS

A. Conditional DDPM

The details of DDPM process and the integration of condition can be referred to reference [7] and [10] respectively. Briefly, we formulate the motion reduction as a conditional generation, approximating a conditional distribution $p(x_i|c_i)$ where $\{x_i, c_i\}_{i=1}^{N}$ is a pair of motion-free (x_i) and motion-corrupted (c_i, produced by FBP) CT images. *N* is the number of image pairs.

Regarding DDPM, it involves a forward and a sampling Markovian process. The forward process has a series of image $x_0 \to x_1 \to \cdots \to x_T$, where random Gaussian noise is gradually added in each step, turning a motion-free image (x_0) into a pure Gaussian noise image (x_T) after T iteration steps. In this Markov chain, we define the distribution of any intermediate step x_t given x_0 as a Gaussian governed by a step-dependent variance α_t:

$$q(x_t|x_0) \sim \mathcal{N}\left(\sqrt{\overline{\alpha_t}}x_0, (1-\overline{\alpha_t})\boldsymbol{I}\right) \quad (1)$$

where $\overline{\alpha_t} = \prod_{t=1}^{t} \alpha_t$. Thus, any x_t can be calculated as:

$$x_t = \sqrt{\overline{\alpha_t}}x_0 + \sqrt{1-\overline{\alpha_t}}\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0, \boldsymbol{I}) \quad (2)$$

where ε_t represents the noise added in each step.

The sampling process in DDPM is the generation process from a random noise to a motion-free image: $x_T \to x_{T-1} \to \cdots \to x_0$. In each step, we need to estimate the noise ε_t that was added in the forward step and remove it from x_t. Practically, we train a network θ (a U-Net with attention module used in [10]) to estimate ε_t given its resultant x_t, the time step t and the condition c with the objective function:

$$L(\theta) \coloneqq \mathbb{E}\|\varepsilon_t - \varepsilon_\theta(x_t, t, c)\|^2 \quad (3)$$

Simply, this function calculates the mean-squared-error (MSE) between ground truth noise ε_t and estimated noise ε_θ. Once trained, the sampling process turns x_t to x_{t-1} via:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha_t}}}\varepsilon_\theta(x_t, t, c)\right) + \sqrt{1-\alpha_t}z \quad (4)$$

$z \sim \mathcal{N}(0, \boldsymbol{I})$, a standard Gaussian distribution. In our motion reduction study, c is always set to be the motion-corrupted image in the pair of $\{x_i, c_i\}$. T, the total number of sampling steps, is 2000.

All diffusion models mentioned in this abstract were trained for over 1,000,000 iterations and showed loss convergence.

B. Solving Question 1: Distinct Intensity Ranges

Problem: We intuitively started with using 3D DDPM. As shown in **Figure 1 row 3**, the implementation of 3D DDPM on original images reveals visually poorer reconstruction in brain than skull, as evidenced by dramatic intensity shift and blurred anatomies. This discrepancy is also evident when we switched to the 2D DDPM approach (**Figure 1 row 5**).

Hypothesis: We have two hypotheses to explain the failure. First, as the first step of the sampling process, the random noise image x_T spans the full intensity range of head CT images, usually from -1000 (background) to >2000HU (skull). However, the brain pixel intensity usually ranges from 0 to 100HU, only a small fraction of the full CT range. Therefore, only few pixels in x_T initially align with intensity range of brain, making it challenging for the subsequent steps to recover brain. Second, in the training process, the training loss is predominatntly influenced by the skull since the skull pixel values are significantly higher and show a larger difference between motion-free and motion-corrupted images compared to the brain.

Solution: To address these issues, we proposed an intensity cutoff in image preprocessing, cutting imag pixel values into [-100, 100]HU to target the brain and background pixels. This led to the training of two anatomy-specific DDPM models: θ_{skull} for skull recovery and was trained using original CT images, and θ_{brain} for brain recovery and was trained using intensity-cutoff images. It's important to note that θ_{skull} and θ_{brain} were trained and sampled *independently*. The final motion-corrected image, I_{DDPM}, combines the brain from I_{brain} sampled by θ_{brain} and the skull from I_{skull} by θ_{skull}. Practically, our combination algorithm is as follows. For pixel *P*:

$$I_{DDPM}(P) = \begin{cases} I_{brain}(P) & X_{low} \le I_{skull}(P) \le X_{high} \\ I_{skull}(P) & \text{otherwise} \end{cases} \quad (5)$$

Here, X_{low} and X_{high} are cutoff thresholds close to the θ_{brain} training window, -100 and 100 HU, respectively. -95 and 95 HU were used in our experiments.

C. Solving Question 2: 3D or 2D

Problem: We would like to investigate whether (a) using a 3D DDPM model, taking the entire 3D volume as input, or (b) using a pair of 2D DDPM models would yield better motion artifact reduction.

Solution: For (b) 2D approach, we assumed 3D head motion can be decomposited into intra-plane and inter-plane 2D motions. Therefore we trained two 2D models, one on 2D axial planes (planes along z-axis, learning intra-plane motion artifacts) and the other on perpendicular planes (planes along x-axis, learning inter-plane artifacts). The sampling process then alternated between these two models in a switching mode according to Lee et al.[11] and Niu et al.[12].

D. Data and Simulation

We did a simulation study where we had pairs of motion-corrupted (used as condition, made by FBP) and motion-free (ground truth) images. We retrospectively collected 100 motion-free head CT scans from a single center. The pixel size was re-sampled into [1,1,2.5]mm, and the image dimension was set to 256×256×50 covering from the nose plane to the top of the head. To fit 3D DDPM in GPU (memory 40GB), we further down-sampled images by a factor of 2 in x and y, resulting in dimension of 128×128×50. The intensity range was always normalized to [-1,1].

In simulation, a head pose was described using 6 rigid motion variables including translations and rotations along three image axes. A *B*-spline model with 5 control points equally distributed across one gantry rotation used in Jang et al.[3] and Chen et al.[4] was utilized to model each motion variable. Each control point was randomly sampled from [-5,5]mm or degree for each translation or rotation variable. We split 100 scans into a training (80) and a testing (20) dataset and produced 50 simulations for each scan in training (resulting in 4000 training cases) and 5 simulations for each in testing (100 testing cases).

Table 1. Quantitative Results

Category	Image	ROI*	MAE (HU)	SSIM	DICE
Motion	I_m	Skull	330±97	/	0.80±0.07
		Brain	15.4±3.6	0.26±0.09	/
		Overall	119±40	0.77±0.11	/
3D DDPM	$I_{skull,3D}$	Skull	246±64	/	0.87±0.05
	$I_{brain,3D}$	Brain	30.6±3.5	0.34±0.11	/
	$I_{DDPM.3D}$	Brain	7.5±2.7	0.60±0.20	/
		Overall	75±24	0.89±0.06	/
2D DDPM	$I_{skull,2D}$	Skull	124±38	/	0.94±0.02
	$I_{brain,2D}$	Brain	8.4±0.9	0.56±0.08	/
	$I_{DDPM.2D}$	Brain	6.5±1.2	0.68±0.12	/
		Overall	34±11	0.97±0.02	
CNN	I_{CNN}	Skull	274±67	/	0.85±0.05
		Brain	11.9±2.3	0.33±0.08	/
		Overall	90±23	0.85±0.07	/

* **Skull**: the metrics were calculated for pixels > 100 HU on the ground truth image;
Brain: all images were clipped to [-100, 100] HU and metrics were calculated for brain pixels > 0HU;
Overall: the metrics were calculated all foreground pixels > -100HU.

E. Comparative Studies

We conducted four studies for both questions. The first two studies used 3D DDPM, one trained on original images ($\theta_{skull,3D}$) and the other trained on intensity-cutoff images ($\theta_{brain,3D}$). The next two studies employed 2D DDPM, similarly divided between training on original images ($\theta_{skull,2D}$) and intensity-cutoff images ($\theta_{brain,2D}$). Regarding question 1, we compared brain recovery in $I_{skull,3D}$ against $I_{brain,3D}$ and $I_{skull,2D}$ against $I_{brain,2D}$, highlighting the impact of intensity cutoff on improving brain performance. Regarding question 2, we compared the anatomy-specific and overall performance in the final motion-corrected image, $I_{DDPM,3D}$ and $I_{DDPM,2D}$ (combination of the skull in I_{skull} and brain in I_{brain}) and concluded the more effective approach. Additionally, we compared the overall performance of the best-performing DDPM method with I_{CNN} (reproduced according to Su et al.[6]), showcasing DDPM is superior over CNN. MAE and SSIM were used for quantitative evaluation of brain and the entire image, while MAE and DICE (binary cutoff = 100HU) were used for skull evaluation.

III. Results

A. Solving Question 1: Distinct Intensity Ranges

Figure 1 row 3 shows that 3D DDPM trained on original images, $\theta_{skull,3D}$, resulted in blurred brain tissues with dramatic intensity shift. Similarly, **Figure 1 row 5** shows that 2D DDPM trained on original images, $\theta_{skull,2D}$, also experienced severe blurring. However, trained on intensity-cutoff images, $\theta_{brain,3D}$ and $\theta_{brain,2D}$ both achieved a more precise depiction of brain tissue, including finer structures like the sulci (highlighted in **Figure 1 row 4 and 6**). We compare MAE and SSIM in **Table 1**. Notice that inside the brain window, as shown by the shaded rows in **Table 1**, $I_{brain,3D}$ has better performance than $I_{skull,3D}$, and the same applies to $I_{brain,2D}$ and $I_{skull,2D}$. It indicates the effectiveness of the intensity cutoff in enhancing brain tissue recovery.

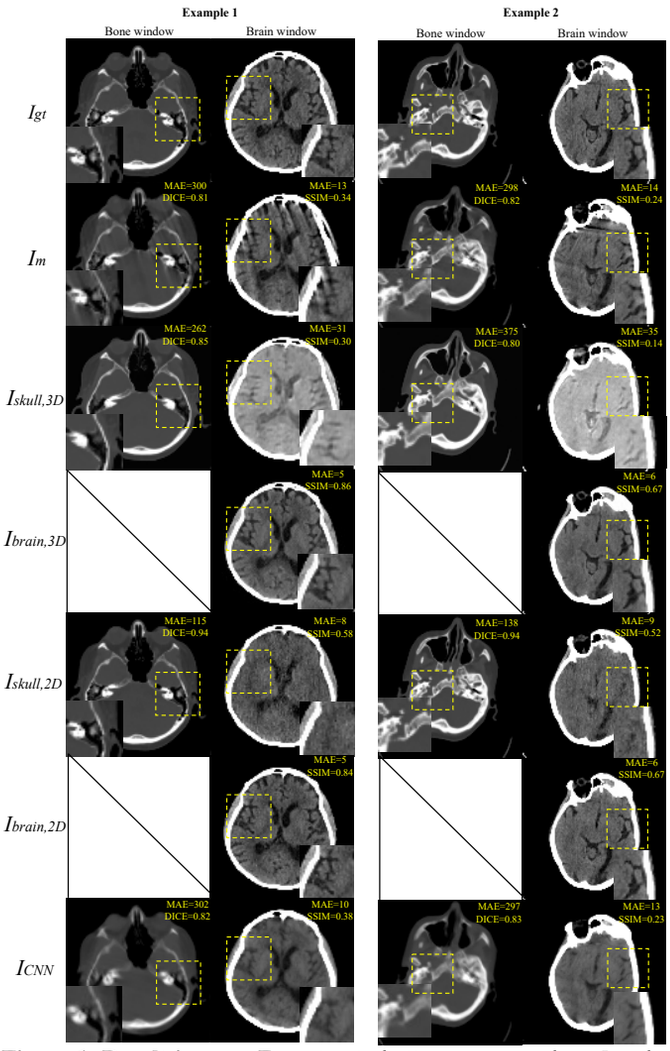


Figure 1. Result images. From top to bottom row, we show I_{gt}, the static ground truth; I_m, the motion-corrupted image; $I_{skull,3D}$ and $I_{brain,3D}$, the results by 3D DDPM; $I_{skull,2D}$ and $I_{brain,2D}$, the results by 2D DDPM; and I_{CNN} using image-domain CNN. MAE and DICE were added for skull pixels in the first column; MAE and SSIM were added for brain pixels in the second column.

B. Solving Question 2: 3D or 2D

Regarding skull recovery, a comparison between **Figure 1 row 3 and 5** reveals that although both 3D and 2D approaches reduced the motion blur, 2D was better at correcting anatomy shift caused by inter-plane motion (such as maxillofacial bones highlighted in yellow boxes). Furthermore, 3D struggled to accurately restore skull pixel values, leading to significantly higher MAE compared to 2D (246±64HU for 3D vs. 124±38 for 2D), as shown in the "skull" rows in **Table 1**.

For brain recovery, 3D and 2D might exhibit comparable visual quality in **Figure 1 row 4 and 6**. However, 2D achieved lower MAE and higher SSIM (MAE: 7.5±2.7HU for 3D vs. 6.5±1.2HU for 2D, SSIM: 0.60±0.20 for 3D vs. 0.68±0.12 for 2D), as shown in the "brain" rows in **Table 1**. The smaller standard deviation of 2D also indicates more stable performance.

Overall, $I_{DDPM,3D}$ yielded MAE = 75±24HU and SSIM = 0.89±0.06, while $I_{DDPM,2D}$ achieved significantly lower MAE = 34±11HU and higher SSIM = 0.97±0.02, indicating that the 2D approach has better recovery of the image, which is possibly due to the lack of model capacity and the hardness of training in the 3D model.

Furthermore, it's important to note the advantage of the 2D model in terms of sampling efficiency. The 2D process takes approximately 2 minutes to complete a case, significantly faster than the 10 minutes required by the 3D approach. This efficiency makes the 2D model not only more effective but also more practical for clinical applications.

However, it is worth noting that the 2D model does not come without limitations. Upon close inspection of the motion corrected images, the 3D results demonstrated brain structures slightly more faithful to the ground truth compared to the 2D results, despite the general blurring. It may be because that the 3D model has more accurate modeling of the 3D motion.

C. Overall Performance by I_{DDPM} and I_{CNN}

Figure 1 row 7 reveals that I_{CNN} tended to over-smooth images with overall higher MAE and lower SSIM than I_{DDPM} (90±23 vs. 34±11HU for MAE, 0.85±0.07HU vs. 0.97±0.02HU for SSIM).

IV. Conclusion

We conclude that employing a DDPM approach with a 2D backbone, combined with an intensity cutoff in image preprocessing is the most optimized DDPM approach to reduce motion artifacts in head CT scans. This method addresses the challenges posed by the distinct intensity ranges of skull and brain tissues and the 3D nature of head CT and motion.

V. Reference

[1] S. Ouadah et al., "Correction of patient motion in cone-beam CT using 3D-2D registration," *Phys. Med. Biol*.

[2] H. Bruder et al., "Compensation of skull motion and breathing motion in CT using data-based and image-based metrics, respectively," in *Medical Imaging 2016: Physics of Medical Imaging*, SPIE.

[3] S. Jang et al., "Head motion correction based on filtered backprojection for x-ray CT imaging," *Med. Phys.*

[4] Z. Chen et al., "A Deep Learning Approach to Estimate and Compensate Motion in Non-contrast Head CT scan," *17th Int. Meet. Fully 3D Image Reconstr. Radiol. Nucl. Med.*, Nov. 2023, doi: 10.48550/arXiv.2310.16846.

[5] Y. Ko et al., "Rigid and non-rigid motion artifact reduction in X-ray CT using attention module," *Med. Image Anal.*

[6] B. Su et al., "A deep learning method for eliminating head motion artifacts in computed tomography," *Med. Phys.*

[7] J. Ho et al., "Denoising Diffusion Probabilistic Models." arXiv.

[8] Y. Song et al., "Solving Inverse Problems in Medical Imaging with Score-Based Generative Models," arXiv.

[9] W. Xia et al., "Patch-Based Denoising Diffusion Probabilistic Model for Sparse-View CT Reconstruction," arXiv.

[10] C. Saharia et al., "Image Super-Resolution via Iterative Refinement," arXiv.

[11] S. Lee et al., "Improving 3D Imaging with Pre-Trained Perpendicular 2D Diffusion Models." arXiv.

[12] C. Niu et al., "3D Photon Couting CT Image Super-Resolution Using Conditional Diffusion Model," *17th Int. Meet. Fully 3D Image Reconstr. Radiol. Nucl. Med.*, Nov. 2023, doi: 10.48550/arXiv.2310.16846.

Ring Artifacts Removal from CT Images Using an Adaptive Patch Filter in the Frequency Domain

Xiangyuan Liu, Shusen Zhao*, Dimeng Xia, Weibin Zhang, and Fengwei An*, *Member, IEEE*

***Abstract*—Due to the inconsistent response of detectors, some undesired line -shaped components would appear in the sinogram data, so as to ring artifacts in the reconstructed computer tomography (CT) images and finally degrade image quality. Removing or minimizing such artifacts to the greatest extent is of great significance for accurate CT image analysis and processing. In this paper, we propose a novel and effective algorithm for correcting ring artifacts utilizing the remarkable properties of corrupted sinogram data in the frequency domain. According to the geometric and value attributes of undesired line-shaped elements in the frequency domain, we develop the patch-based rank-ordered distance (PROD) statistic to detect the abnormal patches, corresponding to ring artifacts in the CT image, in the frequency data of the sinogram image. Furthermore, an adaptive patch filter is developed to correct the undesired patches measured by the PROD detector and preserve the normal elements to a great extent. Therefore, this filter can effectively restore image structures while correcting the ring artifacts. The experimental results demonstrate that the proposed method outperforms state-of-the-art methods in terms of artifacts correction and structure preservation, so as to offer enhanced image.**

***Index Terms*—Ring artifact, computer tomography images, sinogram data.**

I. Introduction

RING artifact [1] is a common artifact in computed tomography (CT) images, and mainly caused by the inconsistent response of detectors. The defective detectors would cause line-shaped intensities in the sinogram data, which finally give raise to ring artifacts, undesired components in circle-shapes in the reconstructed CT images. Ring artifacts will seriously degrade image quality and greatly affect subsequent image processing and analysis. Therefore, removing or suppressing such artifacts to the greatest extent possible has great significance.

Correcting the detector plane can to some extent reduce artifacts during data acquisition, but the ring artifacts in the reconstructed usually can not be effectively avoided. Therefore, multiple post-processing techniques have been developed for reducing ring artifacts in CT images. Various techniques of ring artifacts removal are comprehensively compared in [2], and mainly be classified into two groups: sinogram processing and reconstructed images processing.

The sinogram domain methods attempt to correct the spurious strips in the sinogram data before reconstructing the CT images, thus, rings can be reduced if the method succeeds. The measured mean values of detector pixels are used to correct the undesired strips in the sinogram data in [3][4]. A measurement-driven correction method is developed in [5] by subtracting the mean of the dark current acquired projection images. A wavelet and Fourier filtering combined technique is provided in [6], where the discrete wavelet transform is applied to extract the image with significant feature and a Gaussian filter then works on the frequency data obtained by fast Fourier transform (FFT). The iterative filtering technique is used for ring artifacts removal in [7].

The reconstructed CT processing techniques aim at extract the undesired circles and correct them. It is a popular method of transforming ring artifacts in a Cartesian coordinate into linear artifacts in a polar coordinate and then processing them [8][9]. An optimization model that using a novel ring total variation (RTV) regularization is developed in [10] to compress ring artifacts by simultaneously updating the sinogram and the reconstructed images. A deep-learning network with spectral loss is proposed in [11] to correct ring artifacts in photon-counting spectral CT images.

In this paper, we propose a novel and effective method for removing ring artifacts by correcting the undesired line-shaped components in the frequency domain. The details of the proposed method are provided below.

II. Methodology

A. Ring artifacts in multi-domains

Computed tomography attempts to reconstruct the CT image of an object from the forward projections (also known as sinogram data). The ring artifacts degraded CT image, which we mainly focused on in this work, can be modeled as

$$Y + \epsilon = P(X + \varepsilon), \tag{1}$$

where X and Y respectively refer to the ideal sinogram data and the ideal reconstructed CT image, which is obtained by performing the back projection operation P. $X + \varepsilon$ represents the obtained sinogram data that be corrupted by the undesired data ε, which causes ring artifacts ϵ in the degraded CT image $Y + \epsilon$. Therefore, effectively correcting the data in the sinogram or in the reconstructed domain is the main technique for reducing ring artifacts in CT images.

This work was supported by the Shenzhen Science and Technology Program under Grant RCBS20221008093316046 and JSGGZD20220822095600001. (Corresponding author: Shusen Zhao and Fengwei An.)

Xiangyuan Liu, Shusen Zhao, Dimeng Xia and Weibin Zhang are with the National Center for Applied Mathematics Shenzhen, Southern University of Science and Technology, Shenzhen 518055, China (e-mail: liuxy7@mail.sustech.edu.cn, zhaoss@sustech.edu.cn).

Fengwei An is with the School of Microelectronics, Southern University of Science and Technology and Pengcheng Laboratory, Shenzhen, China (e-mail: anfw@sustech.edu.cn).

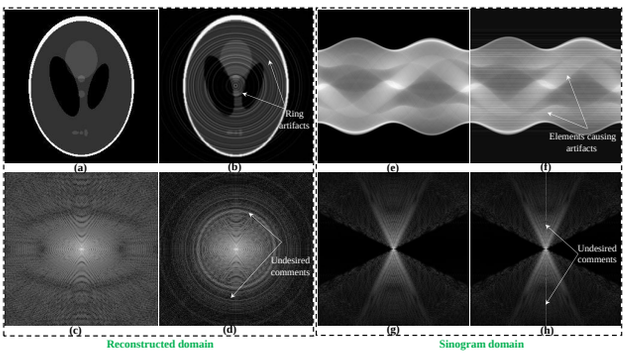


Fig. 1. Ring artifacts appear in various domains. The second row displays the frequency data obtained by performing FFT on the images in the first row. (a) and (b) represent the clear CT image and the degraded image with ring artifacts, respectively reconstructed using the sinogram data (e) and (f).

(***Properties***) The characteristics of ring artifacts in the sinogram and reconstructed data are examined in different domains, primarily in the spatial domain and the frequency domain, as depicted in Fig.1. It can be observed that the ring artifacts (indicated by the white arrows in Fig.1(b)) in the reconstructed CT image and the corresponding undesired components (indicated by the white arrows in Fig.1(d)) caused by these artifacts in the frequency data are all circular in shape. However, in the sinogram domain, the elements that cause ring artifacts in CT images are shaped like lines (indicated by the white arrows in Fig. 1(f)). In the frequency space, the corresponding undesired components are much simpler, appearing as approximately straight lines perpendicular to the horizontal direction (indicated by the white arrows in Fig.1(h)).

(***Motivation and framework***) The property analysis above shows that ring artifacts have more prominent and simpler features in the frequency space of the sinogram data (Fig. 1(h)). Meanwhile, the ring artifacts in the CT images still appear in circular shapes and disrupt some important structures of the frequency data, as can be seen by comparing Fig.1(c) with Fig.1(d). Therefore, it may be more effective to reduce the ring artifacts by correcting the streaks in the sinogram data in the frequency domain. The proposed method mainly includes two steps: *detecting* and *correcting*. **First**, the PROD statistic is defined based on the properties of ring artifacts in the sinogram domain and in the frequency space, including their geometric and value characteristics. It measures the signed difference between patches and is used to detect abnormal patches. The details are provided in *Section II-B*. **Next**, an adaptive patch filter is developed to correct the undesired patches that cause the line-shaped components detected by the PROD statistic, while preserving the normal patches to the greatest extent. The details are provided in *Section II-C*.

B. The PROD *detector*

FFig.1 shows that the undesired strips in the sinogram data, which cause the ring artifacts in CT images, are significantly different from their surrounding components in the frequency data. Therefore, to correct the line-shaped elements in the sinogram data (Fig.1(f)), so as to reduce the ring artifacts in CT images, we first attempt to define the patch-based detector PROD measure how a patch differs from its surrounding elements. The processed elements at the base are patches rather than pixels because the undesired components may be clustered sparse outliers (see Fig.1(h)). Furthermore, the statistic is determined by the sorted signed distance, as the undesired patches exhibit significantly higher spectral values than the surrounding patches.

More precisely, let $M : \Omega \to R$ be the frequency image of a sinogram data, where Ω is the image domain. Then, the value of PROD at $i \in \Omega$ is defined as

$$\mathrm{PROD}(i) = PR_1(i) + PR_2(i) + \cdots + PR_m(i), \quad (2)$$

where $PR_k(i)$ is the kth largest term in $\{d_P(i,j)|j \in N_{i,s} \setminus \{i\}\}$ with

$$\begin{aligned} d_P(i,j) &= d_s(M(N_{i,d}), M(N_{j,d})) \\ &= \min_{\theta \in \{\frac{2\pi k}{n}\}_{k=0}^{k=n}} \sum_{y \in N_{i,d}} (M(y) - M_\theta(T_{i,j}y)), \end{aligned} \quad (3)$$

where $N_{i,s} \subseteq \Omega$ represents a square window centered at i with a radius of s, $M(N_{i,d}) = \{M(t)|t \in N_{i,d}\}$ denotes a patch, and $d_P(i,j) = d_s(M(N_{i,d}), M(N_{j,d}))$ quantifies the signed distance between patches $M(N_{i,d})$ and $M(N_{j,d})$. $T_{i,j}$ is the translation that maps $N_{i,d}$ to $N_{j,d}$ by $T_{i,j}y = j + (y - i)$, and m is usually set as $m = \left\lfloor \frac{Card(N_{i,s})}{2} \right\rfloor$. We also consider the influence of rotation on $d_P(i,j)$ because rotation does not change image similarity from a geometric perspective. We obtain $M_\theta(T_{i,j}y)$ by rotating $M(T_{i,j}y)$ with an angle θ. The PROD(i) can detect significantly different patches in frequency space using their local information. The PROD(i) is a large positive value when patch $M(N_{i,d})$ consists of sparse outliers corresponding to stripes in sinogram data, as each $PR_k(i)$ will be a positive value.

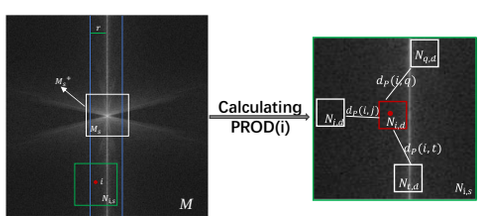


Fig. 2. The procession of calculating PROD (Eq.2) and several limitations proposed in the work. The right image is the enlarged green framed region in the left image and shows the calculation of PROD(i).

The proposed PROD detector can be applied to the entire image as well as to specific regions, as each element is measured using its local information. Taking into account the characteristics of the frequency data and the geometric attributes of undesired elements, several constraints are imposed when calculating the PROD statistic in this paper. **First**, the PROD is performed in the exterior region $M_s^- = M(\Omega \setminus N_s)$ of a rectangular region $M_s = M(N_s)$ centered at the frequency data M because the central region M_s always contains the most information and structures of an image, but the central value is extremely large. **Second**, the PROD statistic can be

limited to a disturbance area centered around the centerline with a radius r, indicated by the blue lines in Fig. 2. This is because the undesired components are perpendicular to the horizontal direction and form a straight line-shaped area, almost centered at the centerline in the frequency data (see Fig.1(h)), while most other elements remain unchanged. The calculation of the $\mathrm{PROD(i)}$ statistic and its limitations are shown in Fig.2.

C. The adaptive patch filter

The analysis of the degraded sinogram data in different domains indicates that the ring artifacts are attributed to abnormal elements (shaped like lines and with high magnitudes) in the frequency space (Fig.1h), while the majority of other components remain unaffected.

Therefore, we define the adaptive patch filter using the PROD statistic (Eq.2) to remove the ring artifacts by filtering out the undesired patches that cause the line-shaped elements in the frequency image of the spinogram data. There are three important properties of the proposed filter. **First**, the proposed filter can adaptively detect and process the elements that cause the stripes in the sinogram data. **Second**, the fundamental unit of the novel filter is the patch rather than the pixel, as the undesired elements in the frequency data may be clustered outliers. **Third**, the undesired units are filtered out by focusing on units with small values, as undesired elements typically have high magnitudes.

More precisely, the proposed filter is defined as

$$\hat{M}(N_{i,d}) = \begin{cases} M(N_{i^*,d}), & \text{if } \mathrm{PROD(i)} > \mathrm{S} \cdot \max\{\mathrm{M}\} \\ M(N_{i,d}), & Otherwise \end{cases}, \quad (4)$$

with

$$N_{i^*,d} = arg \max_{\{N_{j,d}\}_{j \in N_{i,s}}} \{d_s(M(N_{i,d}), M(N_{j,d}))\}, \quad (5)$$

where M and $\hat{M}$ respectively represent the frequency data of the sinogram data and the filtered result, the signed distance $d_s(\cdot,\cdot)$ between patches is defined in Eq.3. The parameters s and d are the same as those that define the PROD detector (Eq.2). $S \cdot \max\{M\}$ represents a fixed threshold, where $\max\{M\}$ denotes the largest magnitude of M and the scale S is set within the range of 0.02 to 0.5. In our experiments, we have $S = 0.05$. The PROD is used to identify the patch $M(N_{i,d})$ that potentially causing the undesired line-shaped elements, and then filter it using $M(N_{i^*,d})$. Choosing the filtered elements by Eq.5 can ensure that no other high-valued abnormal elements will be introduced, and the image information will be restored as much as possible.

Using the proposed filter (Eq.4), the patches that cause undesired stripes in the sinogram data, as well as ring artifacts in CT images, can be adaptively detected and filtered, while preserving the normal information effectively. As a result, the ring artifacts are effectively reduced, and the essential structures in the images are accurately restored.

III. Experiments

The proposed algorithm will be evaluated in this section. The most widely used method, known as W-FFT, combines wavelet and Fourier filtering [6] and be used as the comparative method. The effectiveness of different methods in various domains is tested using cases of weak and strong ring artifacts. The FBP algorithm is utilized to reconstruct the CT image.

a) ***Week artifacts correction****:* We will first evaluate the effectiveness of the proposed algorithm in correcting weak artifacts. The experimental data is the scanning data of a mouse skull, with 720 rotation angles and 512 detector pixels. The results are shown in Fig.3, where the red and green colored regions in Fig.3(a) are magnified and respectively placed in the bottom-right and bottom-left corners of the image to emphasize the details. Fig.3(d) is the original degraded image, and the result in blue framed region is shown in Fig.3(a), Fig.3(b), and Fig.3(e).

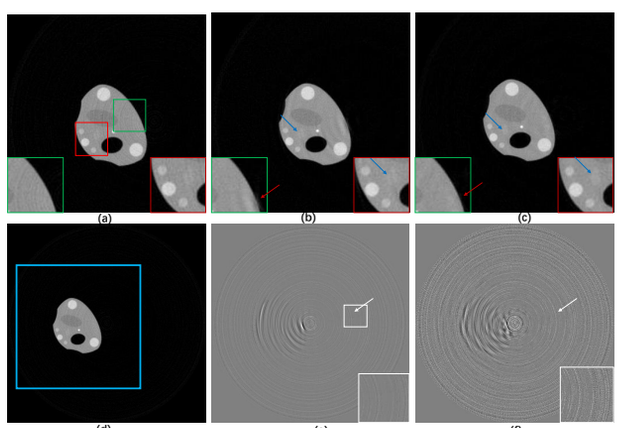


Fig. 3. The reduction of weak ring artifacts: (a), (b), and (c) are the enlarged blue framed region in the original degraded image (d), with (a) degraded data, (b) the results of W-FFT, (c) the proposed method, (e) and (f) are respectively the difference image of W-FFT and our method corresponding to (d).

It can be found that the proposed method (Fig.3(c)) performs better in correcting weak ring artifacts and accurately restoring image structures. More precisely, the zoomed green areas in Fig.3(b) and Fig.3(c) show that our method can effectively avoid edge blurring when removing artifacts (as indicated by the red arrows). In addition, the enlarged red areas demonstrate that our method performs better in accurately restoring image information. Specifically, the proposed method can effectively restore the gray area (indicated by the blue arrows in the large image), while the W-FFT method (Fig.3(b)) severely loses such information. Finally, the difference images Fig.3(e) and Fig.3(f) in the second row show that our method can remove more undesired artifacts from the reconstructed image, as indicated by the white arrows and enlarged white areas.

b) ***Sharp artifacts correction****:* We assessed the efficacy of the proposed algorithm in correcting sharp ring artifacts using test data of pork placed in a cup, with 720 rotation angles and 2060 detector pixels. The results are shown in Fig.4, and the red and green areas are enlarged. In Fig. 4(a), the white area within the green frame represents bones, and the gray area within the red frame represents meat in this data.

It shows that the proposed method (Fig.4(c)) can effectively correct sharp ring artifacts and accurately restore important

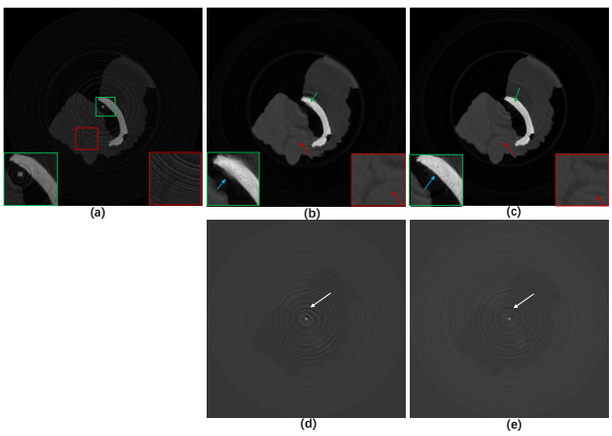


Fig. 4. The reduction of sharp ring artifacts: (a) represents the degraded data, (b) and (c) show the results of W-FFT and the proposed method, and (d) and (e) are respectively the difference image of (b) and (c) in relation to (a).

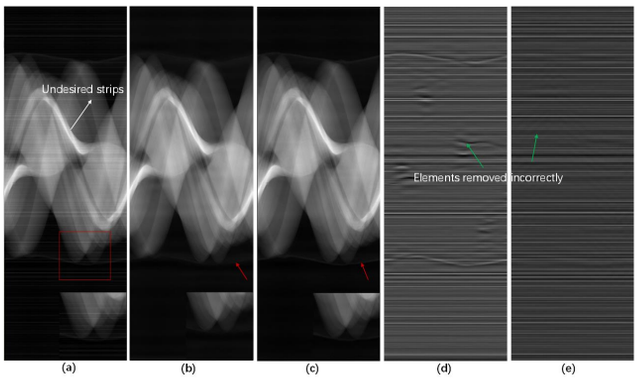


Fig. 5. The corrected results of Fig.4(a) in the sinogram domain, where (a) represents the degraded data, (b) and (c) show the results of W-FFT and the proposed method, and (d) and (e) are respectively the difference images of (b) and (c) in comparison to (a).

structures in CT images. Specifically, the correction results in the meat area (as indicated by the enlarged red area) show that our method can effectively prevent edge blurring when removing artifacts, especially as seen in the red arrows. The corrected results of the bone area (as shown in the enlarged green area) indicate that our method performs better in restoring more image structures. Our method accurately restores the small holes in the bone, particularly the black dots indicated by the green arrows. In contrast, the W-FFT method (Fig. 4(b)) is unable to restore such information and results in more severe blurring (as indicated by the blue arrows). Finally, the difference images Fig.4(d) and Fig.4(e) indicate that our method can accurately remove unwanted artifacts from the reconstructed image, as indicated by the white arrows.

c) ***Sinogram domain results****:* Finally, we evaluated the proposed algorithm in the sinogram domain, and the experimental image is the sinogram data of Fig.4(a). The results are shown in Fig. 5, where the line-shaped elements (indicated by the white arrows) cause ring artifacts in the reconstructed image (Fig.4(a)), and the red area is enlarged and displayed. The proposed method effectively corrects the undesired line-shaped elements in the sinogram data while preserving the normal structures, as demonstrated in the comparison of difference images Fig.5(d) and Fig. 5(e), particularly in the area indicated by the green arrows. Furthermore, this conclusion can also be inferred from the comparison of the red areas and the areas indicated by the red arrows in Fig.5(b) and Fig.5(c).

IV. Conclusion

This paper proposes an effective method for reducing ring artifacts by processing the sinogram data in the frequency domain. This approach is based on the observation that the undesired elements in the sinogram data, which cause the ring artifacts, exhibit more significant and simpler features in the frequency domain. We have developed an adaptive patch filter to correct abnormal patches in the frequency image, reducing undesired components in the sinogram image and thereby minimizing ring artifacts in the CT image. During the filtering process, the patches that cause artifacts can be effectively identified and measured using the PROD statistic, and then adaptively corrected by the proposed filter, while ensuring that the clear patches are sufficiently maintained. The experimental results show that the proposed method performs better in terms of correcting ring artifacts and recovering the natural image structure.

References

[1] J. Barrett and K. N., "Artifacts in CT: recognition and avoidance," *Radiographics*, vol. 24, no. 6, pp. 1679–1691, 2004.

[2] S. Rashid, S. Lee, and M. Hasan, "An improved method for the removal of ring artifacts in high resolution ct imaging," *Eurasip Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 1–18, 2012.

[3] E. Mohamed, H. Mohamed, C. Yang, M.and Myung, C. Min, and L. Soo, "A ring artifact correction method: Validation by micro-CT imaging with flat-panel detectors and a 2d photon-counting detector," *Sensors*, vol. 2017, no. 17, pp. 269–279, 2017.

[4] S. Rudin, A. R. Podgorsak, S. V. S. Nagesh, D. R. Bednarek, and C. N. Ionita, "Use of a cmos-based micro-CT system to validate a novel ring artifact correction algorithm on low-dose image data," in *Physics of Medical Imaging*, 2018.

[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," *Optics Express*, vol. 27, no. 10, pp. 14 231–14 245, 2019.

[6] 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.

[7] A. Matani and K. Terakawa, "Artifact reduction filtering method for CT images," in *Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology*, vol. 2, 2002, pp. 1035–1036.

[8] D. Prell, Y. Kyriakou, and W. A. Kalender, "Comparison of ring artifact correction methods for flat-detector CT," *Physics in Medicine & Biology*, vol. 54, no. 12, pp. 3881 – 3895, 2009.

[9] F. Brun, G. Kourousias, D. Dreossi, and L. Mancini, "An improved method for ring artifacts removing in reconstructed tomographic images," in *World Congress on Medical Physics and Biomedical Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 926–929.

[10] 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," *Medical physics*, vol. 46, no. 11, pp. 4803–4815, 2019.

[11] D. Hein, K. Liappis, F. Grnberg, A. Eguizabal, and M. Persson, "Ring artifact correction in photon-counting spectral CT using a convolutional neural network with spectral loss," *arXiv e-prints*, 2023.

Rawdata-Based Higher Order Deep Beam Hardening Correction (DBHC)

Fabian Jäger, Pascal Paysan, Michał Walczak, and Marc Kachelrieß

***Abstract*—Background: CBCT images, essential for image-guided radiotherapy, must display precise CT values. Artifacts like streaks or cupping, caused by the polychromatic radiation source, are common. While water precorrection helps to reduce soft tissue artifacts, it is less effective for materials vastly different from water, such as bone.**
Purpose: The reliance on a *a priori* reconstruction can be problematic if this first pass reconstruction contains too many artifacts, such as truncation. In this study we perform a first test to use a deep learning-based method to directly correct for beam hardening in the rawdata domain. In addition, it is applicable in real time and can be applied on single projections.
Methods: A convolutional neural network (CNN) is used to predict a boneless projection. The bone free projection with the measured projection can be used to extract a ratio from bone to soft tissue. This ratio is used for a higher order beam hardening correction, similar to a water precorrection.
Results: A first simulation study was performed to proof the proposed method. Different clinical CT scans were used to simulate polychromatic CBCT projections which are the input for our CNN. The projections are reconstructed and compared to a water precorrection and the monochromatic reconstruction.
Conclusions:The proposed method can perform a higher order beam hardening correction. The difference to a monochromatic scan is much lower than for a water precorrected reconstruction, drastically decreasing the difference in the bones and reducing the streak artifacts caused by the beam hardening.

***Index Terms*—CBCT, Beam Hardening, Deep Learning.**

I. Introduction

CANCER is one of the leading causes for death world wide. There is countless research done to improve the outcome for patients, by making treatment better and saver. In radiation therapy, the attenuation μ can be used to estimate the stopping power. Hence, a mapping as close as possible to the real object is necessary to archive a reasonable treatment plan. The current cone-beam computed tomography (CBCT) scanners do not have sufficient image quality to optimize treatment plans for radiotherapy. Different artifacts during image acquisition and reconstruction are the cause for the image quality loss.

The log attenuation reaching a detector element, is given by

$$q(L) = -\ln \int dE\, w(E,L)\, e^{-\int d\lambda\, \mu(E,\boldsymbol{s}(L)+\lambda\Theta(L))} \tag{1}$$

Here, L is the line of integration corresponding to a ray starting at $\boldsymbol{s}(L)$ going along the directional vector $\Theta(L)$. $\mu(E,\boldsymbol{r})$ is the energy-dependent spatial distribution of the linear attenuation coefficient. E is the photon energy and $w(E,L)$ is the detected spectrum normalized to one, which is angle-dependent, and hence depends on L. Note, the dependence on L is dropped for the sake of readability. The spectrum includes the emitted x-ray distribution and attenuation aspects such as the prefiltration and the sensitivity of the detector. Furthermore, the attenuation $\mu(E,\boldsymbol{r})$ can be split into a spatial and an energy-dependent part $\mu(E,\boldsymbol{r}) = \psi(E)f(\boldsymbol{r})$. Then, we can define:

$$p = \int d\lambda\, f(\boldsymbol{s}+\lambda\Theta) \tag{2}$$

Evidently, there is a non-linear relation $q = q(p)$ between q and p.

In CT $\mu(\boldsymbol{r})$ is assessed by reconstructing the attenuation measured along various lines L. During reconstruction, the non-linearity is neglected by dropping the energy dependence of q. The relationship $q(p)$ is concave and strictly monotonically increasing. Hence, the measured attenuation q increases slower than the line integral, i.e. if the line integral p is doubled, the attenuation q will not double. Consequently, the attenuation underestimates the true line integral p which results in cupping artifacts. In addition, objects with a higher atomic number than water induce streak artifacts.

For a first-order beam hardening correction, of some object $f_0(\boldsymbol{r})$, we assume the object to consists of only one material, with energy-dependency $\psi_0(E)$. Then, the measured attenuation q is connected to $f_0(\boldsymbol{r})$ by

$$q = -\ln \int dE\, w(E) e^{-p_0\psi_0(E)}, \tag{3}$$

with p_0 being the line integral through $f_0(\boldsymbol{r})$. In the preprocessing before the image reconstruction the correction inverts the function $q(p_0)$ to obtain p_0. In clinical CT commonly water is used ($\psi_0(E) = \psi_{\text{water}}(E)$), due to its abundance in the human body[1]. Hence, the name water precorrection is used in the following. Whenever there is tissue which is not water-equivalent (i.e. bones), beam hardening artifacts still appear after water precorrection. There are plenty of higher order beam hardening correction algorithms based on a prior reconstruction [1], [2], [3], [4].

II. Material & Methods

This section provides the theory for an higher order beam hardening correction and discusses the use of a deep learning based method.

Fabian Jäger, and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls-University, Heidelberg, Germany. Dr. Pascal Paysan and Dr. Michał Walczak are with Varian Medical Systems Imaging Lab, GmbH, Baden-Dättwil, Switzerland. Corresponding author: fabian.jaeger@dkfz.de

A. Higher Order Hardening Correction

The attenuation q can be rewritten for a higher order beam hardening approach as

$$q = -\ln \int dE\, w(E) e^{-p_{\text{soft}}\psi_{\text{soft}}(E) - p_{\text{bone}}\psi_{\text{bone}}(E)} \quad (4)$$

Now, consisting of two tissues (bone and soft tissue) with their respective energy dependence. The line integrals p_{soft} and p_{bone} can be estimated from a prior reconstruction [2]. This is prone to artifacts such as truncation, due to the missing information in the reconstruction. From Eq. (4) we know, that a higher order beam hardening correction was possible if we knew the line integral in soft and bone tissue. The aim of the deep beam hardening correction is to calculate p_{soft} and p_{bone} from the prediction of the neural network (NN). This allow us to calculate virtual monoenergetic images by calculating the monochromatic attenuation as

$$p_{\text{mono}}(E_0) = p_{\text{soft}}\psi_{\text{soft}}(E_0) + p_{\text{bone}}\psi_{\text{bone}}(E_0). \quad (5)$$

B. Deep Beam Hardening Correction

At first glance, the obvious step would be to train a network to predict p_{soft} and p_{bone} directly from the attenuation q. As seen in Fig. 1, p_{soft} has a dark shadow at the position of the bones, which we will call "holes" in the following. Moreover, bones generate high-frequency structures in both p_{soft} and p_{bone}, the correct prediction of which may be more difficult to learn. In addition, the NN has to learn a mapping from the polychromatic q to the spectrum-independent line integral. Our proposed method, relies on predicting q_{water}. Similar to Eq. (3), q_{water} is calculated as

$$q_{\text{water}} = -\ln \int dE w(E) e^{-p_{\text{water}}\psi_{\text{soft}}(E)} \quad . \quad (6)$$

$p_{\text{water}} = p_{\text{soft}} + p_{\text{waterbones}}$ is the line integral over the soft tissue and the bones replaced by standard water (see Fig. 1). Further, we define

$$q = q_{\text{water}} + q_{\text{bone}} \quad (7)$$

The declarations are summarized in Table I.

TABLE I
SUMMARY OF THE MOST IMPORTANT DEFINED VARIABLES.

p_{soft}	Monochromatic forward projection (FP), soft tissue only
p_{bone}	Monochromatic FP, bone tissue only
p_{water}	Monochromatic FP, bones replaced with water and soft tissue
q_{water}	Polychromatic FP, bones replaced with water and soft tissue
q	Polychromatic FP

With the assumption

$$r = \frac{p_{\text{bone}}}{p_{\text{soft}}} \approx \frac{q_{\text{bone}}}{q_{\text{water}}} \quad (8)$$

we can rewrite Eq. (4) as

$$q = -\ln \int dE\, w(E) e^{-p_{\text{soft}}\psi_{\text{soft}}(E) - r p_{\text{soft}}\psi_{\text{bone}}(E)} \quad (9)$$

With the known spectrum $w(E)$ and good approximations for the energy dependencies of the soft and bone tissue Eq. (9) can be solved with any common root finding algorithm. We decided to use the Newton-Raphson method. Ultimately, the NN performs a deep bone extraction (DBE) in the prediction of q_{water}.

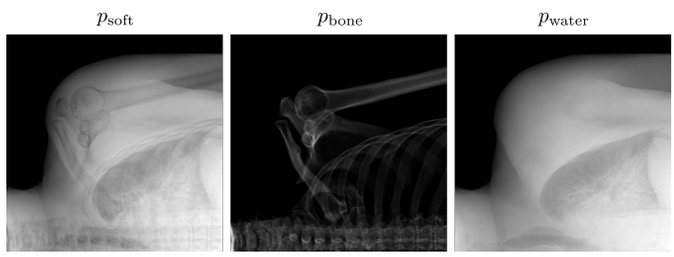


Fig. 1. The line integrals in the soft tissue (left), bone tissue (middle) and soft tissue with bones replaced by standard water (right), p_{soft}, p_{bone} and p_{water} respectively.

C. Training Data

To calculate data for training, validation and testing we prepare clinical CT scans by first creating a segmentation (see II-C1) and then calculating the NN inputs q and labels q_{water} (see II-C2).

1) Segmentation: A binary segmentation with a threshold of 200 HU was used. In addition, morphological operations were used to eliminate singular voxels without a neighbour bone voxel to eliminate voxel which would be classified as bone due to noise. The segmentation is used to substitute bones with water with an attenuation of $\mu_{\text{water}} = 0.0192/\text{mm}$.

2) Data Generation: To calculate q and q_{water} we calculate the line integrals p_{soft}, p_{bone} and p_{water}. In addition with a spectrum and the energy dependence of the soft and the bone tissue we can use Eq. (4) to generate training data for the DBE. For the polychromatic spectrum, a semiempirical model was used to generate tungsten target X-ray spectra [5]. With a tube voltage of 125 kV and a prefiltration of 0.4 mm tin. For our simulation we used $\psi_{\text{soft}}(E) = \psi_{\text{water}}(E)$. Since $\mu(E, \boldsymbol{r})$ is the product of $f(\boldsymbol{r})$ and $\psi(E)$, one can trade an arbitrary constant between the two functions. In our example the spatial function $f(\boldsymbol{r})$ is defined as

$$f(\boldsymbol{r}) = \frac{\text{CT}(\boldsymbol{r}) + 1000\ \text{HU}}{1000\ \text{HU}} \mu_{\text{water}}(70\ \text{keV})$$

with $\text{CT}(\boldsymbol{r})$ being the CT value distribution. We define $\psi(E)$ for tissue i as

$$\psi_i(E) = \frac{(\mu/\rho)_i(E)}{(\mu/\rho)_{\text{water}}(70\ \text{keV})}.$$

The mass attenuation values $(\mu/\rho)(E)$ for different elements and tissues as well as the material compositions of different tissue types can be found in the literature [6], [7].

A set of nine full body CT scans was used for the data generation. The scans are from cadaver measurements and have been acquired in scope of a forensic study in close collaboration with the Institute of Forensic and Traffic Medicine (Prof. Sarah Heinze), Heidelberg University, Heidelberg, Germany, after being approved by the local ethics review board (S 388/2014).The patients were split between training set (seven patients) and one patient each for testing and validation. For

each patient different points were selected (excluding regions with artifacts such as metal implants). Each point is used as the rotation center for a CBCT scan with a shifted detector. 72 projections were evenly simulated across a 360° range per scan. This ensures that the network is trained with as much different data as possible in order to make it more robust and better able to generalize. As a result, the training set comprised approximately 7500 projections, while the validation set contained roughly 1000 projections. The projection number does not represent a clear 7:1 ratio because some patients have more or less regions which got excluded due to artifacts.

D. Network and Training

Here, we use the U-net architecture [8]. It consists of an encoding path and a decoding path which are connected by a skip connection. There are six stages in the encoding path, going from 16 to 1024 feature channels. Each stage applies two 3×3 convolutions, followed by a rectified linear unit (ReLU) activation. The encoding path uses max pooling and the decoding path employs a bilinear upsampling to restore the spatial resolution. Corresponding layers in the encoding and decoding paths are linked by skip connections, helping to preserve fine-grained details. In the final layer of the U-net, a 1×1 convolution is applied to map the 16-component feature vector to the desired output channel. The network was initialized with the He initialization, also known as Kaiming initialization [9] and the bias with zeros. As optimizer the AdamW [10] a modified version of the Adam algorithm [11] was chosen with the L1 loss. The network was trained on patches of 640×640 pixel (slightly larger than the receptive field of the U-net) and with a batch size of 16. Entire projections were used as input for validation and testing. The network was trained until the validation loss no longer decreased for 100 epochs with a learning rate of 10^{-5}.

E. Simulation Study

First, let us verify the method by comparing the beam hardening correction from a set of labels, i.e., the best possible beam hardening correction we can achieve with the ratio of $q_{\text{water}}/q_{\text{bone}}$.

Second, we want to evaluate the NN predictions by calculating the mean absolute percentage error (MAPE) to the label for a test scan:

$$\text{MAPE} = \frac{1}{K} \sum_{n,\boldsymbol{u}} \left| \frac{\hat{y}(n, \boldsymbol{u}, \boldsymbol{w}) - y(n, \boldsymbol{u})}{y(n, \boldsymbol{u}) + \epsilon} \right| \tag{10}$$

where n is the sample index, $\boldsymbol{u}$ the pixel coordinate, K the product of batch size and pixel number, $\hat{y}$ the prediction with weights $\boldsymbol{w}$, y the label and $\epsilon = 10^{-5}$ a small factor to avoid division by zero.

Then, the reconstruction of water-precorrected projections is used as reference method and for both methods the mean absolute error (MAE) to a monochromatic reconstruction is calculated.

III. Results

In the initial part of this section, our objective is to explore the optimal beam hardening correction (BHC) based on the presupposition that $r = p_{\text{bone}}/p_{\text{soft}}$ is approximately equal to $q_{\text{bone}}/q_{\text{water}}$. In the second part we investigate a thorax scan of the test patient.

A. Theoretical Achievable BHC

Figure 3 displays the estimated values for p_{soft} and p_{bone}, determined from q and q_{water}, and compares them with the ground truth (GT), the actual p_{soft} and p_{bone} that were used to compute q. The shown absolute percentage error shows a large deviation at the bone regions in both line integral estimations. The calculated p_{soft} exhibits the same feature as the p_{water} used to calculate q_{water}. The "holes" from the bones are not visible. This is somewhat expected, because the ratio $q_{\text{bone}}/q_{\text{water}}$ also considers "filled" bones in q_{water}. More important is the reconstruction of a monochromatic scan and the comparison with the reconstruction of the GT line integrals. For the shown narrow window of 50 HU the deviations show improvement to a water precorrection (see Fig. 2).

B. Deep BHC

Now we want to perform the beam hardening correction with the predicted q_{water}. In a first step we calculated the MAPE for the predictions on a fully simulated scan of the test patient. The MAPE was calculated in the patient region excluding air and was found to be $1.21 \pm 0.31\,\%$. The error is given by the standard deviation of the absolute percentage error.

Fig. 2 shows an example slice for the different reconstructions and compares them to a monochromatic reconstruction. The improvement from polychromatic to water-precorrected and then to further improvement of the bone tissue by the DBHC is clearly visible.

Furthermore, the mean absolute error (MAE) across all the reconstructed slices, comparing the performance of the water precorrection, our proposed method, and our method utilizing the q_{water} label against the monochromatic reconstruction is 3.4 HU, 2.3 HU and 1.7 HU, respectively. The MAE is reduced by over $30\,\%$ if compared to the water precorrection. The small deviation between the results from the NN and the labels highlights the good performance of the NN while having a MAPE of over one percent between the label and the predicted q_{water}. DBHC also reduced the maximum deviation to the ground truth by over $45\,\%$ from 91 HU to 48 HU.

IV. Discussion

The deep beam hardening correction in rawdata domain out performs the water precorrection if the respectively corrected reconstructions are compared to the monochromatic reconstructions. The proposed method is real time applicable and can be applied on a single projection, with no need of a prior reconstruction.

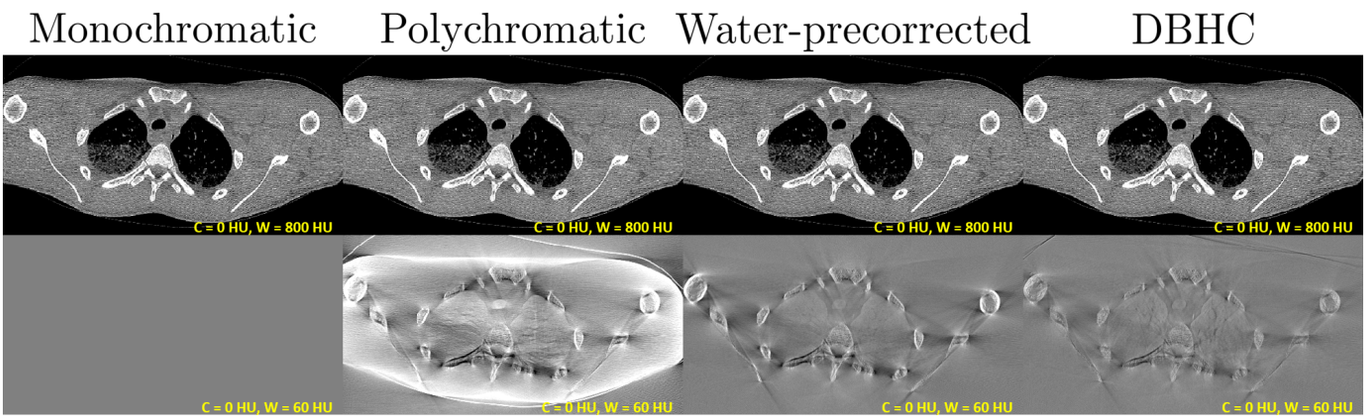


Fig. 2. The evaluation contrasts various polychromatic reconstruction approaches: one without any correction, another with a water precorrection, and a third using our proposed method, all against a monochromatic reconstruction. The bottom row shows the difference to the monochromatic reconstruction. Reconstruction: $C = 0\,\mathrm{HU}$, $W = 800\,\mathrm{HU}$; Differences: $C = 0\,\mathrm{HU}$, $W = 50\,\mathrm{HU}$; Difference Polychromatic: $C = 0\,\mathrm{HU}$, $W = 500\,\mathrm{HU}$

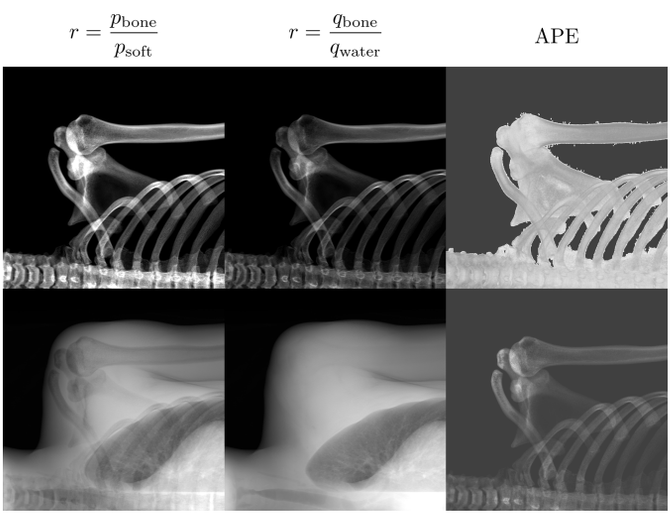


Fig. 3. The left column shows the ground truth (GT) p_{bone} (top) and p_{soft} (bottom). The second column gives the line integrals as estimated from q and q_{water}. The third column shows the absolute percentage (APE) between the estimation and the GT. p_{soft}: $C = 0.21$, $W = 0.42$; p_{bone}: $C = 0.5$, $W = 1.0$; APE: $C = 0.25$, $W = 1.0$

The only drawback is that the spectrum needs to be known for a precise beam hardening correction. Further experiment are needed to investigate how prone this method is if the wrong spectrum is assumed.

V. Conclusions

This work shows a possible deep higher order beam hardening correction, outperforming a water precorrection. So far it was evaluated only on one anatomic region and only on a singular scan. Both of these factors can be easily addressed. Another point to be addressed is the testing on real measurements.

Acknowledgement

Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany. This project was funded by Varian Medical Systems, a Siemens Healthineers Company. Prof. Dr. Sarah Heinze from the Institute of Forensic and Traffic Medicine and their collaboration on a forensic study.

References

[1] Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelrieß, "Empirical beam hardening correction (EBHC) for CT," *Med. Phys.*, vol. 37, no. 10, pp. 5179–5187, Oct. 2010.

[2] M. Kachelrieß and W. A. Kalender, "Improving PET/CT attenuation correction with iterative beam hardening correction," *IEEE Medical Imaging Conference Program*, pp. M04–5, Oct. 2005.

[3] D. Oh, S. Kim, D. Park, S. Choi, H. Song, Y. Choi, S. Moon, J. Baek, and D. Hwang, "Correction of severe beam-hardening artifacts via a high-order linearization function using a prior-image-based parameter selection method," *Medical Physics*, vol. 45, pp. 4133–4144, 9 2018.

[4] J. Levi, B. L. Eck, R. Fahmi, H. Wu, M. Vembar, A. Dhanantwari, A. Fares, H. G. Bezerra, and D. L. Wilson, "Calibration-free beam hardening correction for myocardial perfusion imaging using ct," *Medical Physics*, vol. 46, pp. 1648–1662, 4 2019.

[5] D. M. Tucker, G. T. Barnes, and D. P. Chakraborty, "Semiempirical model for generating tungsten target x–ray spectra," *Med. Phys.*, vol. 18, no. 3, pp. 211–218, Mar./Apr. 1991.

[6] D. E. Cullen, J. H. Hubbell, and L. Kissel, "EPDL97: The evaluated photon data library, '97 version," *UCRL-50400*, vol. 6, 1997.

[7] D. R. White, J. Booz, R. V. Griffith, J. J. Spokas, and I. J. Wilson, "4. the composition of body tissues," *Reports of the International Commission on Radiation Units and Measurements*, vol. os-23, no. 1, pp. 20–23, 1989.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," vol. 9351. Springer Verlag, 2015, pp. 234–241.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," 2 2015.

[10] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 11 2017.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 12 2014.

Task-Specific Data Preparation for Deep Learning to Reconstruct Structures of Interest from Severely Truncated CBCT Data

Yixing Huang*, Fuxin Fan*, Ahmed Gomaa, Andreas Maier, Rainer Fietkau, Christoph Bert, and Florian Putz

***Abstract*—Cone-beam computed tomography (CBCT) is widely used in interventional surgeries and radiation oncology. Due to the limited size of flat-panel detectors, anatomical structures might be missing outside the limited field-of-view (FOV), which restricts the clinical applications of CBCT systems. Recently, deep learning methods have been proposed to extend the FOV for multi-slice CT systems. However, in mobile CBCT system with a smaller FOV size, projection data is severely truncated and it is challenging for a network to restore all missing structures outside the FOV. In some applications, only certain structures outside the FOV are of interest, e.g., ribs in needle path planning for liver/lung cancer diagnosis. Therefore, a task-specific data preparation method is proposed in this work, which automatically let the network focus on structures of interest instead of all the structures. Our preliminary experiment shows that Pix2pixGAN with a conventional training has the risk to reconstruct false positive and false negative rib structures from severely truncated CBCT data, whereas Pix2pixGAN with the proposed task-specific training can reconstruct all the ribs reliably. The proposed method is promising to empower CBCT with more clinical applications.**

***Index Terms*—Field-of-view extension, computed tomography, deep learning, data truncation, task-specific learning.**

I. Introduction

CONE-BEAM computed tomography (CBCT) imaging is widely used in interventional surgeries [1], [2] and radiation oncology [3]–[5]. Data truncation is a common problem in practical CBCT applications, which arises in two scenarios: a) Application of collimators in interior tomography [6]–[8] to reduce dose exposure to sensitive organs, e.g., protecting eyes in head imaging; b) The flat-panel detectors with a limited size are not large enough to cover entire patients or additional structures of interest (SOI) [2]–[4], e.g., distorted fiducial markers in spine surgery and missing shoulders in head & neck adaptive radiotherapy. To extend the field of view (FOV) in the latter scenario, special scan trajectories have been developed, including traverse-continuous-rotate scan [9], elliptical trajectory [10], and multiple source-translation scan [11]. Nevertheless, the circular trajectory is the most widely applied trajectory in practice.

Y. Huang, A. Gomaa, R. Fietkau, C. Bert, and F. Putz are with Department of Radiation Oncology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany. They are also with Comprehensive Cancer Center Erlangen-EMN (CCC ER-EMN), 91054 Erlangen, Germany.

F. Fan and A. Maier are with Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91058 Erlangen, Germany.

Y. Huang and F. Fan contribute equally. Correspondence: yixing.yh.huang@fau.de, fuxin.fan@fau.de

To extend the FOV of CBCT with a conventional circular trajectory, conventional extrapolation methods such as cosine function fitting [12] and water cylinder extrapolation (WCE) [13] have been applied. Compressed sensing techniques are effective to reduce cupping artifacts inside the FOV, but are not sufficient to restore anatomical structures outside the FOV [6], [8]. Recently, the advent of advanced deep learning techniques has opened up new possibilities for extending the FOV [4], [8], [14]–[16], which has the potential to empower CBCT with a broader range of clinical applications. Fonseca *et al.* has demonstrated the superiority of a U-Net based deep learning method called HDeepFov to a recently released Siemens commercial algorithm HDFov [4]. In our previous work [8], a general plug-and-play framework has been proposed to extend the FOV from projection data extrapolated from deep learning prior reconstruction images, which guarantees data consistency inside the FOV and makes the most of the deep learning models to estimate structures outside the FOV. Xu *et al.* have proposed a two-step method for body composition analysis [15]: one network for FOV border extension and the other network for missing tissue imputation. Xie *et al.* has demonstrated that CT images inpainted by their proposed gated convolutional (GatedConv) network can achieve more accurate dose planning than other networks for radiotherapy [16].

Note that multi-slice CT systems are considered in the previous work [4], [14]–[16], which typically have a larger FOV size than CBCT systems. Especially, mobile C-arm CBCT systems can have a very small FOV of 16 cm in diameter [2]. As a consequence, a large portion of anatomical structures are severely truncated, posing a challenge for neural networks to estimate all missing structures. In some clinical applications, only certain structures outside the FOV are of interest, for example, fiducial markers in spine surgery [2] and skeletal structures for path planning in image-guided needle biopsy in lung/liver cancer diagnosis [3], [17]. In this work, a task-specific data preparation method for deep learning to reconstruct SOI outside FOV from severely truncated CBCT data is proposed.

II. Methods

A. Task-Specific Data Preparation

In the scenarios with severely truncated data, it is very challenging for a deep learning model to restore complete structures outside the FOV accurately due to the large amount

of missing data. In this work, we propose a task-specific learning strategy to let neural networks focus on reconstructed SOI only, since only certain structures outside the FOV are of interest in certain applications.

1) Conventional data preparation for FOV extension: To prepare training data, it is straightforward to use images reconstructed from truncated data as the neural network input and images reconstructed from untruncated (complete) data as the neural network output [15], [16], [18]. Such a conventional data preparation way can be represented as follows,

$$\begin{aligned} \boldsymbol{f}_{\text{input}} &= \mathcal{R}(\boldsymbol{A}_{\text{TP}} \cdot \boldsymbol{f}), \\ \boldsymbol{f}_{\text{label, conventional}} &= \mathcal{R}(\boldsymbol{A}_{\text{UTP}} \cdot \boldsymbol{f}), \end{aligned} \tag{1}$$

where $\boldsymbol{f}$ is the complete reference 3D image, $\boldsymbol{f}_{\text{input}}$ is the network input, and $\boldsymbol{f}_{\text{label, conventional}}$ is the conventional network output. The operation $\mathcal{R}$ means image reconstruction, which is FDK reconstruction from water cylinder extrapolated data (WCE) in particular in this work [8]. $\boldsymbol{A}_{\text{TP}}$ and $\boldsymbol{A}_{\text{UTP}}$ are paired truncated and untruncated forward projection operators. In practice, it is challenging to acquire paired truncated and untruncated projection data in real CBCT systems. As deep learning models trained by realistic synthetic data can generalize to real data [19], paired truncated and untruncated forward projection operators (and data) are easily accessible via simulation.

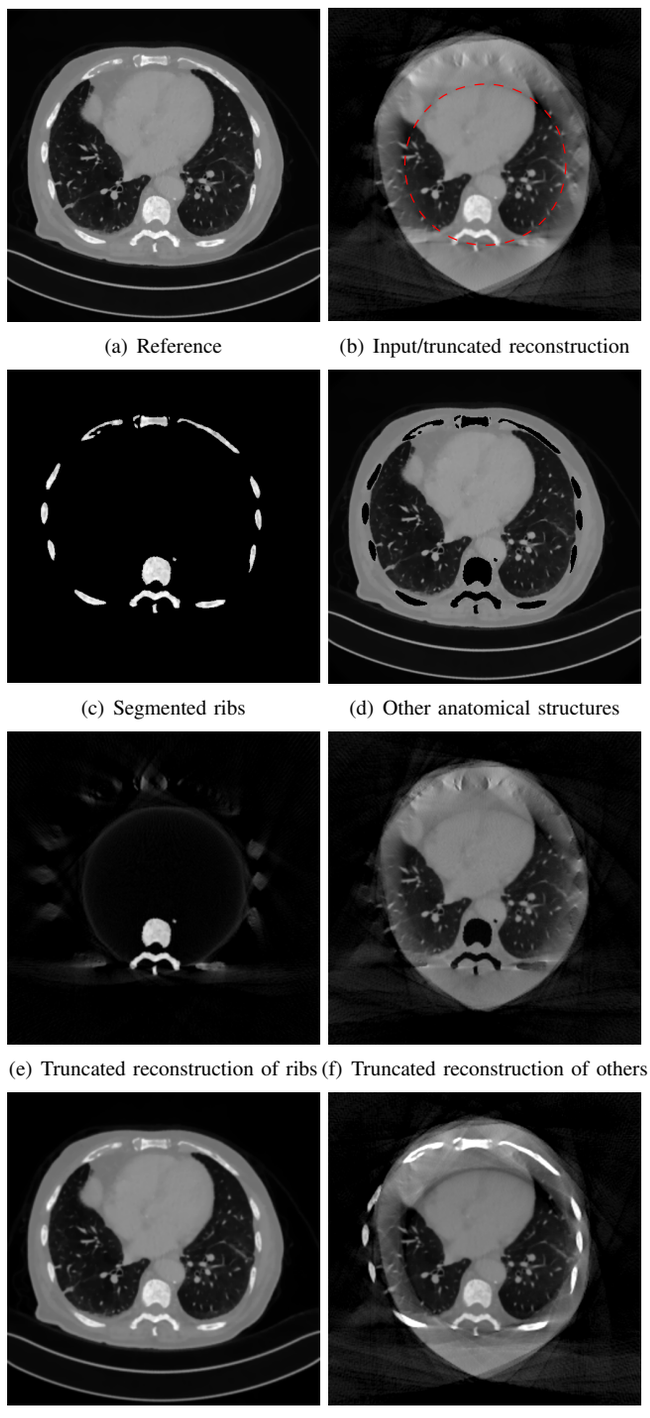
2) Task-specific data preparation for FOV extension: To enable neural networks to concentrate on the SOI, a possible method is to assign greater importance to these structures compared to others through increased weighting. However, in images that are reconstructed from incomplete data, artifacts such as streaks exist globally. As a result, segmenting these areas for weight application becomes impractical. Moreover, if the weighting is restricted solely to the SOI region, the artifacts originating from the SOI will persist outside this specifically weighted zone.

To avoid such difficulty, a task-specific data preparation has been successfully applied to fiducial marker recovery in navigation assisted spine surgery with severely truncated data [2]. In this work, the task-specific data preparation strategy is extended to the following general form:

$$\begin{aligned} \boldsymbol{f}_{\text{input}} &= \mathcal{R}(\boldsymbol{A}_{\text{TP}} \cdot \boldsymbol{f}) = \mathcal{R}(\boldsymbol{A}_{\text{TP}}(\boldsymbol{f}_{\text{Others}}) + \mathcal{R}(\boldsymbol{A}_{\text{TP}}(\boldsymbol{f}_{\text{SOI}})), \\ \boldsymbol{f}_{\text{label, task-specific}} &= \mathcal{R}(\boldsymbol{A}_{\text{TP}}(\boldsymbol{f}_{\text{Others}})) + \mathcal{R}(\boldsymbol{A}_{\text{UTP}}(\boldsymbol{f}_{\text{SOI}})), \end{aligned} \tag{2}$$

where $\boldsymbol{f}_{\text{SOI}}$ denotes the segmented SOI and $\boldsymbol{f}_{\text{others}}$ denotes the other structures. Therefore, our task-specific data preparation relies on the segmentation of SOI as a prerequisite.

Compared with Eqn. (1), the inputs of both data preparations are fundamentally the same. The complete reconstruction can be partitioned into two parts when a linear reconstruction operator like FDK is used. In Eqn. (2), only the SOI are reconstructed from untruncated data in the neural network output. With such data preparation, the difference between an input image and its corresponding output image is originated from SOI only. Therefore, neural networks only need to focus on learning such difference. The task-specific learning makes the network neglect unimportant structures and focus on SOI only, hence it has the potential to generate more robust predictions for complex clinical data than conventional data preparation.



(a) Reference (b) Input/truncated reconstruction

(c) Segmented ribs (d) Other anatomical structures

(e) Truncated reconstruction of ribs (f) Truncated reconstruction of others

(g) Conventional label for training (h) Task-specific label for training

Fig. 1. Intermediate reconstructions under different data preparations. (a) is the reference slice with all structures. (b) is the input reconstruction from truncated data. (c) and (d) contain ribs and other anatomical structures, respectively. (e) and (f) are truncated reconstruction for ribs and other anatomical structures, respectively. (g) shows the label by conventional data preparation. (h) is the label by task-specific data preparation.

3) Exemplification with rib reconstruction in image-guided needle biopsy: To show the generalizability of the task-specific learning for image reconstruction tasks from severely truncated

data, potential applications to image-guided needle biopsy in lung cancer diagnosis [17] and interventional oncology of liver [3] are considered, in addition to navigation assisted spine surgery [2]. In such applications, a 3D volume is needed to establish the safest needle path to the target lesion, where impenetrable obstacles like ribs need to be avoided. Nowadays such a 3D volume is typically acquired by multi-slice CT systems instead of CBCT systems because a large FOV is necessary to cover the lung/liver and ribs. CBCT systems with a small FOV can visualize the target region with high image quality, but the ribs are located outside the FOV, as displayed in the example of Fig. 1(b).

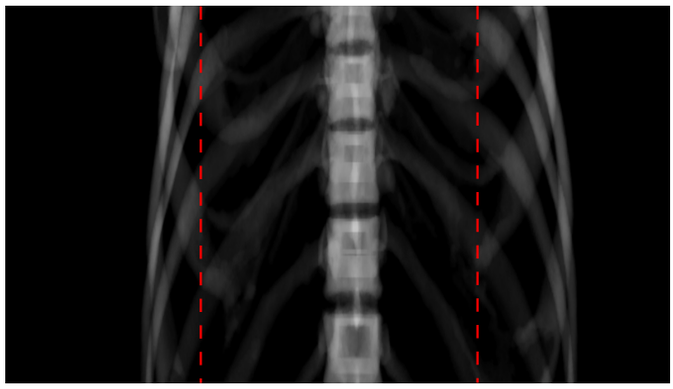
With our task-specific learning, CBCT systems with a small FOV are potentially capable of such an application. To achieve this goal, the skeletal structures including ribs, vertebra and scapula of CT volumes are segmented first for training, as displayed in Fig. 1(c). The soft tissues are displayed in Fig. 1(d). Their projections are displayed in Fig. 2(a) and Fig. 2(b), respectively, and their corresponding truncated reconstructions are displayed in Fig. 1(e) (i.e., $\mathcal{R}(\boldsymbol{A}_{\mathrm{TP}}(\boldsymbol{f}_{\mathrm{SOI}}))$ in Eqn. (2)) and Fig. 1(f) (i.e., $\mathcal{R}(\boldsymbol{A}_{\mathrm{TP}}(\boldsymbol{f}_{\mathrm{Others}})$ in Eqn. (2)), respectively. In Fig. 1(e), the ribs outside the FOV are severely distorted with low contrast, but we can still see their presence. The conventional output for training is Fig. 1(g), which is the reconstruction from nontruncated projection data. In contrast, the task-specific label for training is Fig. 1(h), which is the combination of Fig. 1(c) and Fig. 1(f), whereas the input image Fig. 3(b) is the combination of Fig. 1(c) and Fig. 1(e).

B. Experimental Setup

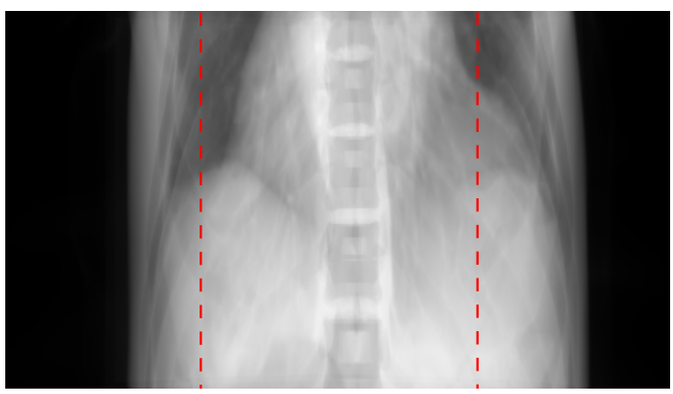
To demonstrate this, the skeletal structures including ribs, vertebra and scapula of 18 AAPM Low Dose CT Grand Challenge volumes are segmented with 3D Slicer [20] segmentation tools. A CBCT system with parameters in Tab. I is used to simulate projection data. The simulated CBCT system parameters are aligned with a Siemens mobile C-arm system. Artificial Poisson noise is added considering a photo number of 10^6 for each incident X-ray. 1360 slices from 17 volumes are used for training Pix2pixGAN models with either conventional data preparation or task-specific data preparation, and the 18th volume is used for test. The models are trained for 300 epochs with an initial learning rate of 0.001 and a decay rate of 0.97. The Adam optimizer is used.

TABLE I
PARAMETERS FOR THE C-ARM CBCT SYSTEM AND RECONSTRUCTION.

Parameter	Value
Scan angular range	200°
Angular step	0.5°
Source-to-detector distance	1164 mm
Source-to-isocentor distance	622 mm
Detector size	500×680
Extended virtual detector size	1200×680
Detector pixel size (mm)	0.608×0.608
FOV diameter	16.2 cm
Reconstruction volume size	$512 \times 512 \times 512$
Reconstruction voxel size (mm)	$0.625 \times 0.625 \times 0.625$



(a) Projection of ribs

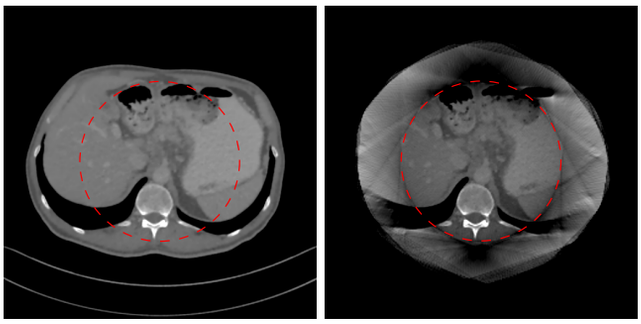


(b) Projection of other anatomical structures

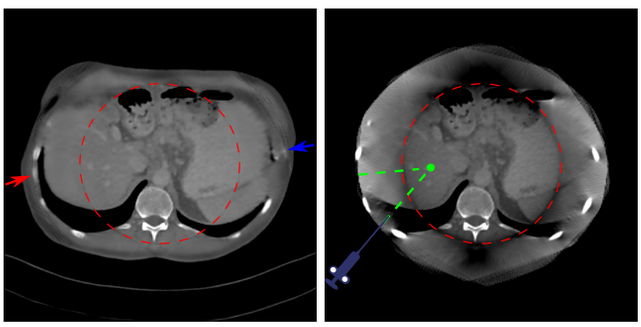
Fig. 2. The projections of one patient's ribs solely (a) and other anatomical structures (b), where the dashed lines indicate the truncation boundaries, display windows [0, 4.5] for (a) and [0, 9] for (b).

III. RESULTS AND DISCUSSION

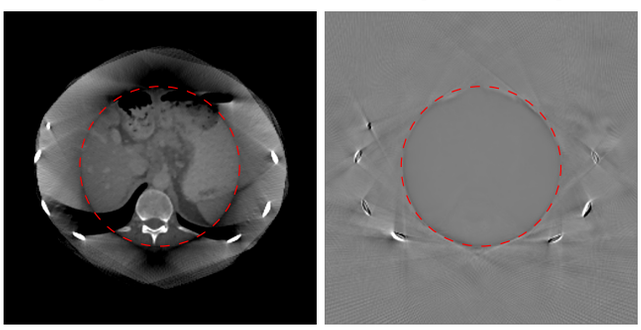
To show the efficacy of the task-specific learning for image reconstruction from severely truncated data, the results of an exemplary slice are displayed in Fig. 3, where the structures inside the FOV and the ribs outside the FOV are of interest. In the input image (Fig. 3(b)), the ribs are almost not visible at all due to severe truncation. With the conventional data preparation, Pix2pixGAN reconstructs all the structures with a realistic appearance outside the FOV in Fig. 3(c). As a result, the complete liver is visible. However, the body outline is inaccurate at the top left side. More importantly, the ribs of interest are reconstructed incorrectly. A false positive rib is reconstructed, as indicated by the red arrow in Fig. 3(c). In addition, the rib indicated by the blue arrow has very low contrast and wrong shape. A binary rib mask is manually segmented from Fig. 3(c) and its Dice score is 0.906 with respect to the refence rib mask. In contrast, the ribs in the Pix2pixGAN output (Fig. 3(d)) with our proposed task-specific learning are well reconstructed with a high Dice score of 0.958, where the number of ribs is correct and the positions are accurate, although their intensities are higher than those in the reference (Fig. 3(a)) (because the rib regions are overlapped with artifacts caused by the truncated reconstruction of other structures). With the known rib positions, two potential needle paths for the target (marked by the green point) can be found in Fig. 3(d).



(a) Reference (b) Input



(c) Pix2pixGAN, conventional (d) Pix2pixGAN, task-specific



(e) Tasks-specific label (f) Difference, task-specific

Fig. 3. A potential application to rib reconstruction from severely truncated data for image-guided needle biopsy, where the structures inside the FOV (indicated by the dashed circle) and the ribs outside the FOV are of interest, window: [-600, 600] HU. (b) is the input image. (c) is the Pix2pixGAN output with conventional data preparation, where a false positive rib indicated by the red arrow is reconstructed and a rib indicated by the blue arrow has a low contrast and incorrect shape. (d) is the Pix2pixGAN output with our proposed task-specific data preparation, where the number of ribs is correct and the positions of ribs are accurate. The two green dash lines are two potential needle paths for the target (marked by the green point) between ribs. Its error image with respective the task-specific label (e) is displayed in (f).

IV. Conclusion

In this work, a task-specific data preparation method is proposed to reconstruct SOI outside the FOV from severely truncated CBCT data. In the application to path planning in image-guided needle biopsy, Pix2pixGAN with the proposed method can reconstruct the number of ribs correctly and their locations accurately, whereas Pix2pixGAN with conventional training has the risk to reconstruct false positive and false negative ribs, although the images looks realistic. The proposed method has the potential for SOI reconstruction in other CT reconstruction tasks from severely insufficient data, e.g., extremely limited-angle and sparse-view tomography, which will be our future work.

References

[1] R. C. Orth, M. J. Wallace, M. D. Kuo, T. A. C. of the Society of Interventional Radiology *et al.*, "C-arm cone-beam CT: general principles and technical considerations for use in interventional radiology," *J. Vasc. Interv. Radiol.*, vol. 20, no. 7, pp. S538–S544, 2009.

[2] F. Fan, B. Kreher, H. Keil, A. Maier, and Y. Huang, "Fiducial marker recovery and detection from severely truncated data in navigation-assisted spine surgery," *Med. Phys.*, vol. 49, no. 5, pp. 2914–2930, 2022.

[3] B. Bapst, M. Lagadec, R. Breguet, V. Vilgrain, and M. Ronot, "Cone beam computed tomography (CBCT) in the field of interventional oncology of the liver," *Cardiovasc. Intervent. Radiol.*, vol. 39, pp. 8–20, 2016.

[4] G. P. Fonseca, M. Baer-Beck, E. Fournie, C. Hofmann, I. Rinaldi, M. C. Ollers, W. J. C. van Elmpt, and F. Verhaegen, "Evaluation of novel AI-based extended field-of-view CT reconstructions," *Med. Phys.*, May 2021.

[5] A. Karius, V. Strnad, M. Lotter, S. Kreppner, and C. Bert, "First clinical experience with a novel, mobile cone-beam CT system for treatment quality assurance in brachytherapy," *Strahlentherapie und Onkologie*, vol. 198, no. 6, pp. 573–581, 2022.

[6] H. Yu and G. Wang, "Compressed sensing based interior tomography," *Phys. Med. Biol.*, vol. 54, no. 9, p. 2791, 2009.

[7] Y. Han and J. 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.

[8] Y. Huang, A. Preuhs, M. Manhart, G. Lauritsch, and A. Maier, "Data extrapolation from learned prior images for truncation correction in computed tomography," *IEEE Trans. Med. Imaging*, pp. 1–12, 2021.

[9] D. Nahamoo, C. R. Crawford, and A. C. Kak, "Design constraints and reconstruction algorithms for traverse-continuous-rotate CT scanners," *IEEE Trans. Biomed. Eng.*, no. 2, pp. 79–98, 1981.

[10] T. Li, X. Li, Y. Yang, D. E. Heron, and M. S. Huq, "A novel off-axis scanning method for an enlarged ellipse cone-beam computed tomography field of view," *Med. Phys.*, vol. 37, no. 12, pp. 6233–6239, 2010.

[11] H. Yu, S. Ni, J. Chen, W. Ge, L. Zhang, and F. Liu, "Analytical reconstruction algorithm for multiple source-translation computed tomography (mSTCT)," *Appl. Math. Model.*, vol. 117, pp. 251–266, 2023.

[12] K. Sourbelle, M. Kachelriess, and W. A. Kalender, "Reconstruction from truncated projections in CT using adaptive detruncation," *Eur. Radiol.*, vol. 15, no. 5, pp. 1008–1014, May 2005.

[13] 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," *Med. Phys.*, vol. 31, no. 9, pp. 2385–2391, 2004.

[14] B. S. Khural, M. Baer-Beck, E. Fournié, K. Stierstorfer, Y. Huang, and A. Maier, "Deep learning-based extended field of view computed tomography image reconstruction: influence of network design on image estimation outside the scan field of view," *Biomed. Phys. Eng. Express*, vol. 8, no. 2, p. 025021, 2022.

[15] K. Xu, T. Li, M. S. Khan, R. Gao, S. L. Antic, Y. Huo, K. L. Sandler, F. Maldonado, and B. A. Landman, "Body composition assessment with limited field-of-view computed tomography: A semantic image extension perspective," *Med. Image Anal.*, p. 102852, 2023.

[16] K. Xie, L. Gao, H. Zhang, S. Zhang, Q. Xi, F. Zhang, J. Sun, T. Lin, J. Sui, and X. Ni, "Inpainting truncated areas of ct images based on generative adversarial networks with gated convolution for radiotherapy," *Med. Biol. Eng. Comput.*, pp. 1–16, 2023.

[17] C. De Margerie-Mellon, C. De Bazelaire, and E. De Kerviler, "Image-guided biopsy in primary lung cancer: Why, when and how," *Diagn. Interv. Imaging*, vol. 97, no. 10, pp. 965–972, 2016.

[18] Y. Huang, L. Gao, A. Preuhs, and A. Maier, "Field of view extension in computed tomography using deep learning prior," in *Bildverarbeitung für die Medizin 2020*, 2020, pp. 186–191.

[19] C. Gao, B. D. Killeen, Y. Hu, R. B. Grupp, R. H. Taylor, M. Armand, and M. Unberath, "Synthetic data accelerates the development of generalizable learning-based algorithms for x-ray image analysis," *Nat. Mach. Intell.*, vol. 5, no. 3, pp. 294–308, 2023.

[20] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka *et al.*, "3d slicer as an image computing platform for the quantitative imaging network," *Magn. Reson. Imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

Segmentation-Guided Knee Radiograph Generation using Conditional Diffusion Models

Siyuan Mei, Fuxin Fan, Fabian Wagner, Mareike Thies, Mingxuan Gu, Yipeng Sun, and Andreas Maier

***Abstract*—Deep learning-based medical image processing algorithms require representative data during development. In particular, surgical data might be difficult to obtain, and high-quality public datasets are limited. To overcome this limitation and augment datasets, a widely adopted solution is the generation of synthetic images. In this work, we employ conditional diffusion models to generate knee radiographs from contour and bone segmentations. Remarkably, two distinct strategies are presented by incorporating the segmentation as a condition into the sampling and training process, namely, conditional sampling and conditional training. The results demonstrate that both methods can generate realistic images while adhering to the conditioning segmentation. The conditional training method outperforms the conditional sampling method and the conventional U-Net.**

***Index Terms*—Radiograph synthesis, diffusion models, conditional image generation.**

I. Introduction

RADIOGRAPHY is one of the most commonly used medical imaging techniques for diagnosis and surgical interventions, capturing 2D projection images of patients using X-ray. The advent of deep learning-based techniques has enabled automated and precise processing of X-ray images, including organ segmentation [1], motion compensation [2], and denoising [3]. However, many of these methods heavily rely on extensive training data which is challenging to collect at scale [4]. In particular, recent research demands atypical data, such as weight-bearing imaging of knees [5], [6].

To address this challenge, recent works proposed synthesizing simulated data as a substitute for clinical data [7]. Traditional forward projection methods create digitally reconstructed radiographs (DRRs) from 3D computed tomography (CT) volumes using Radon transform [8], which guarantees geometrical accuracy. However, such methods require original volumetric CT scans, and sophisticated forward models are needed to capture all properties of realistic-looking X-rays (e.g., energy-dependent effects, noise, scatter, etc.). In contrast, deep generative models leverage 2D X-ray datasets to generate radiographs. For instance, Weber et al. [9] employed generative adversarial network (GAN)-based models to augment chest X-ray images.

Recently, diffusion models have emerged as a powerful technique for data generation, demonstrating competitive performance compared to GANs [10]. In addition, diffusion models are also successfully applied in conditional radiograph generation, e.g., projection inpainting conditioned on masked projection [11] and class-conditional chest radiograph synthesis [12]. Despite achieving convincing performance, previous work has mainly focused on generating radiographs under specific conditions. In this work, we focus on knee imaging and extend the conditional generation of knee radiographs to include more general conditions, such as simple segmentation with contour and bone information. To learn the conditional distribution given the segmentation, we propose two distinct pipelines of conditional diffusion models that incorporate conditional images into the sampling and training processes, respectively.

All authors are with the Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. Contact E-mail: siyuan.mei@fau.de.

II. Methods and Materials

A. Method of Conditional Sampling

Diffusion models employ a forward diffusion process and a reverse diffusion process for image generation. In the forward process, data points gradually diffuse into random Gaussian noise over time, implying the transition from a complex to a simple data distribution. Conversely, the reverse diffusion process generates new data samples by progressively removing noise, starting from a Gaussian prior. Elegantly, the forward perturbation can be modeled as a stochastic differential equation (SDE), which is tractable throughout the reverse process [13].

Let $\mathbf{x}_0 \in \mathbb{R}^d$ denote the data sample, $\mathbf{x}_t$ denote the perturbed data at time point $t \in (0, 1]$, and $p(\mathbf{x}_t)$ denote the corresponding probability density function of $\mathbf{x}_t$. The forward SDE is defined by

$$\mathrm{d}\mathbf{x}_t = \mathbf{f}(\mathbf{x}_t, t)\mathrm{d}t + g(t)\mathbf{z}_t, \quad t : 0 \to 1, \tag{1}$$

where $\mathbf{f}(\mathbf{x}_t, t) \in \mathbb{R}^d$ and $g(t) \in \mathbb{R}$ are the drift and diffusion coefficients of $\mathbf{x}_t$, and $\mathbf{z}_t \sim \mathcal{N}(0, \mathbf{I})$ defines a random variable drawn from a standard normal distribution approximating the standard Wiener process. The corresponding reverse-time SDE has the form

$$\mathrm{d}\overline{\mathbf{x}}_t = [\mathbf{f}(\mathbf{x}_t, t) - g(t)^2 \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)]\mathrm{d}t + g(t)\mathbf{z}_t, \quad t : 1 \to 0, \tag{2}$$

where $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ represents the gradient of the logarithmic probability density, also known as the score function.

In this paper, we adopt the variance exploding (VE) SDE setting [13], defining the growing standard deviation of the noise as

$$\sigma_t = \sigma_{min} \left(\frac{\sigma_{max}}{\sigma_{min}} \right)^t, \tag{3}$$

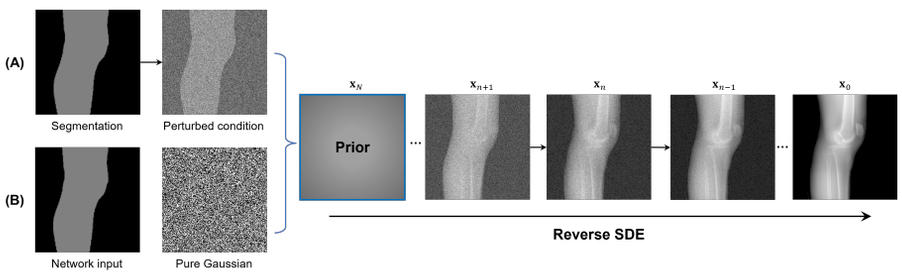


Fig. 1: Overview of the sampling processes. (A) illustrates the sampling process of the conditional sampling method; (B) illustrates the sampling process of the conditional training method.

which determines the drift and diffusion coefficients in (1) and (2) as

$$\mathbf{f}(\mathbf{x}_t, t) = 0, \quad g(t) = \sigma_t \sqrt{2 \log \frac{\sigma_{max}}{\sigma_{min}}}. \tag{4}$$

Crucially, the last unknown term $\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t)$ in (2) can be estimated by training a time-conditional neural network $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$. The optimal parameter $\boldsymbol{\theta}^*$ is obtained by minimizing denoising score matching [14]

$$\begin{aligned} \boldsymbol{\theta}^* &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathbb{E}_{t\sim U[0,1]} \left[\| \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) \|_2^2 \right] \\ &\approx \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathbb{E}_{t\sim U[0,1]} \left[\| \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_0) \|_2^2 \right], \end{aligned} \tag{5}$$

where $p(\mathbf{x}_t|\mathbf{x}_0)$ is the perturbation kernel of the VE SDE.

With segmentation as known information (denoted as $\mathbf{y}$), the conditional sampling method (CSM) follows the SDEdit algorithm [15] to synthesize knee radiographs. As depicted in Fig 1 (A), CSM commences with a perturbed leg contour segmentation, and realistic details are generated through iterative denoising while retaining the desired shape. Notably, the initial perturbing noise is appropriately reduced to preserve conditional information by setting the starting time point t_0 of the reverse diffusion process smaller than 1. Therefore, the prior sample becomes

$$\mathbf{y}_{t_0} \sim \mathcal{N}(\mathbf{y}, \sigma_{t_0}^2 \mathbf{I}). \tag{6}$$

The pipeline of CSM is described in algorithm 1.

Algorithm 1 Sampling algorithm of CSM

Require: N (number of sampling steps), $\mathbf{y}$ (segmentation guide), t_0 (starting time of the reverse diffusion process)

$\mathbf{x}_N \sim \mathbf{y} + \mathcal{N}(0, \sigma_{t_0}^2 \mathbf{I})$
for $n = N-1$ to 0 **do**
  $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$
  $t_n \leftarrow \frac{n}{N} t_0$
  $\Delta \bar{\mathbf{x}}_n \leftarrow g(t_n)^2 \, \mathbf{s}_{\boldsymbol{\theta}^*}(\mathbf{x}_n, t_n) + g(t_n)\mathbf{z}$
  $\mathbf{x}_n \leftarrow \mathbf{x}_{n+1} + \Delta \bar{\mathbf{x}}_n$
end for
return $\mathbf{x}_0$

B. Method of Conditional Training

An alternative to the CSM is integrating the conditions into the training process, thereby directly estimating the score function of the conditional distribution [16]. We concatenate the condition and the perturbed image along the channel dimension as network input. The structure of the score-based network is detailed in section II-D. Surprisingly, this conditional score network can be trained following the same form of denoising score matching [17], which is

$$\begin{aligned} \boldsymbol{\theta}^* &= \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathbb{E}_{t\sim U[0,1]} \left[\| \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, \mathbf{y}, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}) \|_2^2 \right] \\ &\approx \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathbb{E}_{t\sim U[0,1]} \left[\| \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x}_t, \mathbf{y}, t) - \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{x}_0) \|_2^2 \right]. \end{aligned} \tag{7}$$

After training the conditioned network, the score function in (2) can be directly substituted by $\mathbf{s}_{\boldsymbol{\theta}^*}(\mathbf{x}_t, \mathbf{y}, t)$ for the provided conditional image in each sampling step. As shown in Fig. 1 (B), the generated sample is controlled even though the sampling process starts from random Gaussian noise. Algorithm 2 outlines the sampling process of the conditional training method (CTM), where the modifications with respect to algorithm 1 are highlighted in orange.

Algorithm 2 Sampling algorithm of CTM

Require: N (number of sampling steps), $\mathbf{y}$ (segmentation guide)

$\mathbf{x}_N \sim \mathcal{N}(0, \sigma_{max}^2 \mathbf{I})$
for $n = N-1$ to 0 **do**
  $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$
  $t_n \leftarrow \frac{n}{N}$
  $\Delta \bar{\mathbf{x}}_n \leftarrow g(t_n)^2 \, \mathbf{s}_{\boldsymbol{\theta}^*}(\mathbf{x}_n, \mathbf{y}, t_n) + g(t_n)\mathbf{z}$
  $\mathbf{x}_n \leftarrow \mathbf{x}_{n+1} + \Delta \bar{\mathbf{x}}_n$
end for
return $\mathbf{x}_0$

C. Dataset Preparation

To obtain knee radiographs, we selected 55 leg CT volumes from the public SICAS medical image repository [18]. Each CT volume was simulated over 360° with an angular increment of 6° and projected onto the detector of the size 256×256 using CONRAD [19], resulting in 60 DRRs per volume. All projections were normalized to the range of [0,1].

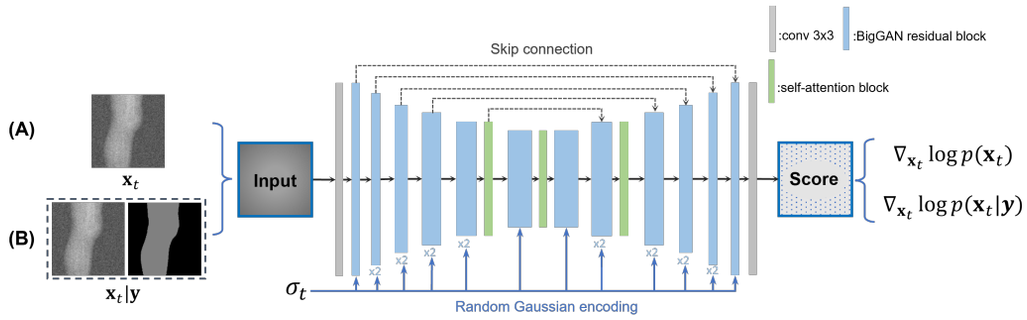


Fig. 2: Structure of noise-conditional score network. (A) represents the network input of CSM; (B) illustrates the network input of CTM.

In addition, two different segmentations for each DRR were automatically generated as follows. The first segmentation extracted the leg contour using a threshold of 0.1, having a value of 0.5 for the contour and 0 for the background (refer to a(0)-c(0) in Fig. 3). The second segmentation extracted bones by thresholding the original CT volume, followed by forward projecting the bone. The bone projection was set to 0.5, and adding it to the contour segmentation resulted in the second segmentation, with bones having a value of 1 and contour a value of 0.5 (refer to d(0)-f(0) in Fig. 3). In total, 3300 radiographs are generated for each type of segmentation. They were randomly split into a 9:1:1 ratio for training, validation, and testing.

D. Network Structure and Hyperparameters

The backbone of the neural network employs the noise-conditional score network++ (NCSNpp) [13]. As illustrated in Fig. 2, we configure six resolution levels of $(256, 128, 64, 32, 16, 8)$ with a corresponding number of channels of $(64, 128, 128, 128, 128, 256)$. Furthermore, the time-conditional noise scale σ_t is encoded to random Gaussian feature [20] and embedded into all residual blocks. Importantly, for CSM only the perturbed X-ray images are input to the network. When conducting the CTM, additional segmentations are concatenated in the input. To compare the diffusion models with naive image-to-image models, the noise-encoding module is removed to form an improved U-Net model and then trained using the $L1$ loss.

In our experiments, the parameters σ_{min} and σ_{max} are set to 0.01 and 128. We use a batch size of 16 and the Adam optimizer with learning rate 2×10^{-4} for training. During the sampling process, a Langevin dynamic corrector with a signal-to-noise ratio of 0.4 is supplied after reverse SDE at each sampling step, and the number of sampling steps N is chosen as 500 to improve sampling speed. Moreover, the hyperparameter t_0 for CSM is set to 0.4. All models are trained on four Nvidia A100 GPUs on a single cluster node with a cap of 300 epochs.

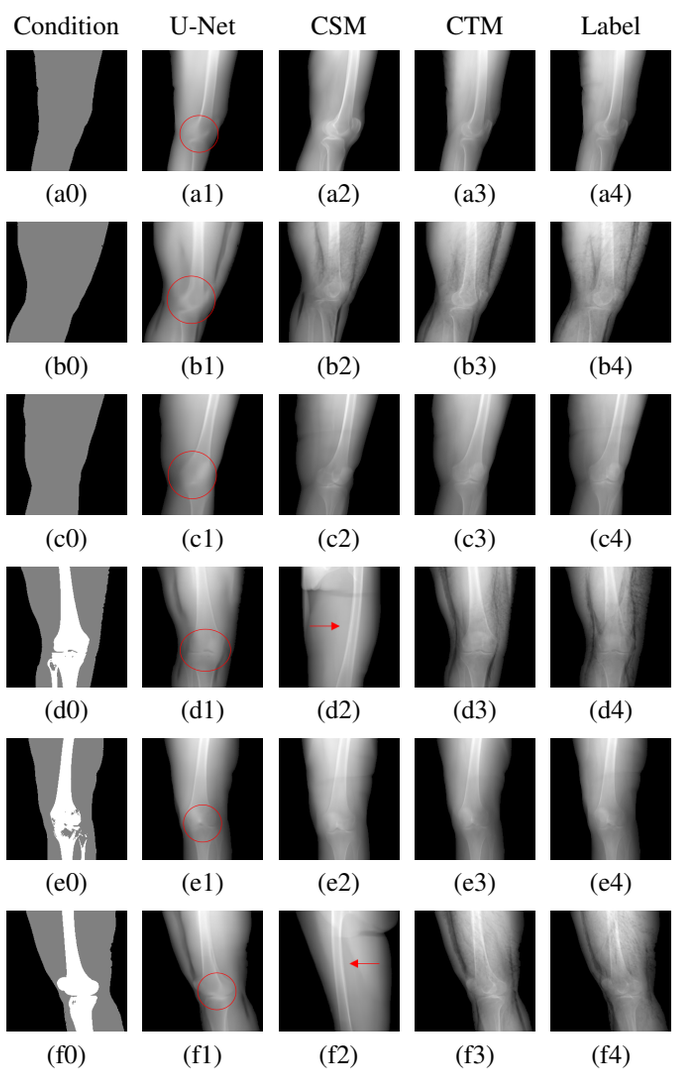


Fig. 3: The generated samples under different conditions. Rows (a)-(c): contour segmentation. Rows (d)-(f): contour and bone segmentation.

III. Results and Discussion

In this section, we provide qualitative and quantitative results of the two proposed diffusion-based methods and compare them with the baseline U-Net model. The first column of Fig. 3 showcases six randomly selected conditions: (a0)-(c0) show contour segmentations, and (d0)-(f0) denote seg-

mentations containing contour and bones. In Fig. 3 (a1)-(f1), the images generated by U-Net contain blurred fine details in locations where bones overlap, despite maintaining the given shape, as highlighted by the red circle. In contrast, the results from CSM appear more realistic than the U-Net. However, their quality decreases with introduced constraints, as indicated by the red arrow in Fig. 3 (d2) and (f2). The results from CTM not only achieve nearly the same level of fineness as the labels but also provide reasonable results with respect to the given conditions as illustrated in the fourth column.

Condition		Contour	Contour+bone
U-Net	MAE	0.0209 ± 0.007	0.0188 ± 0.006
	PSNR (dB)	29.188 ± 2.22	30.304 ± 2.45
CSM	MAE	0.0395 ± 0.010	0.0507 ± 0.010
	PSNR (dB)	22.911 ± 1.89	21.350 ± 1.49
CTM	MAE	$\mathbf{0.0193\pm0.005}$	$\mathbf{0.0152\pm0.007}$
	PSNR (dB)	$\mathbf{29.498\pm1.91}$	$\mathbf{31.680\pm1.76}$

TABLE I: Quantitative model comparison.

Table I summarizes the quantitative results averaged across all testing data. The evaluation metrics include mean absolute error (MAE) and peak signal-to-noise ratio (PSNR). We observed that CTM performs substantially better than U-Net and CSM under both segmentation-based conditions, and CSM performs worse than the U-Net.

Unlike the U-Net which learns a mapping function between input and output, the diffusion models can implicitly capture the underlying data distribution from the training data and then sample it, preventing the loss of fine details on the pixel level. However, in CSM, conditions are incorporated only at the first sampling step while being perturbed, which results in imprecise conditional information. Instead, CTM provides an estimated score function of the conditional distribution for each sampling step, accommodating both reliability and realism. Nonetheless, presently generated X-ray images only encompass independent 2D conditional information, which may introduce geometric inconsistencies between a set of projections. Future research will focus on modeling 3D probabilistic distributions with the provided 2D conditions to enable CT reconstruction from the generated projections. In addition, clinical datasets will also be incorporated.

IV. Conclusion

In this work, we explored two different pipelines of diffusion models to generate segmentation-conditioned knee X-ray data. The results demonstrate that both methods can generate realistic radiographs under the given conditions, with the method of conditional training achieving more stable performance. Ultimately, these high-quality synthetic medical images have the potential to benefit the development of data-driven research and educational applications in the medical field.

Acknowledgment

This work was supported by the European Research Council (ERC Grant No. 810316). The authors gratefully acknowledge the HPC resources provided by NHR@FAU using hardware funded by the German Research Foundation (DFG).

References

[1] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert *et al.*, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint arXiv:1809.10486*, 2018.

[2] M. Thies, F. Wagner, N. Maul, L. Folle, M. Meier, M. Rohleder, L.-S. Schneider, L. Pfaff, M. Gu, J. Utz *et al.*, "Gradient-based geometry learning for fan-beam ct reconstruction," *Physics in Medicine & Biology*, vol. 68, no. 20, p. 205004, 2023.

[3] F. Wagner, M. Thies, M. Gu, Y. Huang, S. Pechmann, M. Patwari, S. Ploner, O. Aust, S. Uderhardt, G. Schett *et al.*, "Ultralow-parameter denoising: Trainable bilateral filter layers in computed tomography," *Medical Physics*, vol. 49, no. 8, pp. 5107–5120, 2022.

[4] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.

[5] B. Bier, K. Aschoff, C. Syben, M. Unberath, M. Levenston, G. Gold, R. Fahrig, and A. Maier, "Detecting anatomical landmarks for motion estimation in weight-bearing imaging of knees," in *Machine Learning for Medical Image Reconstruction: First International Workshop, MLMIR 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1*. Springer, 2018, pp. 83–90.

[6] J.-H. Choi, R. Fahrig, A. Keil, T. F. Besier, S. Pal, E. J. McWalter, G. S. Beaupré, and A. Maier, "Fiducial marker-based correction for involuntary motion in weight-bearing c-arm ct scanning of knees. part i. numerical model-based optimization," *Medical physics*, vol. 40, no. 9, p. 091905, 2013.

[7] C. Gao, B. D. Killeen, Y. Hu, R. B. Grupp, R. H. Taylor, M. Armand, and M. Unberath, "Synthetic data accelerates the development of generalizable learning-based algorithms for x-ray image analysis," *Nature Machine Intelligence*, vol. 5, no. 3, pp. 294–308, 2023.

[8] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*. SIAM, 2001.

[9] T. Weber, M. Ingrisch, B. Bischl, and D. Rügamer, "Implicit embeddings via gan inversion for high resolution chest radiographs," in *MICCAI Workshop on Medical Applications with Disentanglements*. Springer, 2022, pp. 22–32.

[10] P. Dhariwal and A. Nichol, "Diffusion models beat gans on image synthesis," *Advances in neural information processing systems*, vol. 34, pp. 8780–8794, 2021.

[11] S. Mei, F. Fan, and A. Maier, "Metal inpainting in cbct projections using score-based generative model," in *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2023, pp. 1–5.

[12] 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*, 2022, pp. 10 684–10 695.

[13] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.

[14] P. Vincent, "A connection between score matching and denoising autoencoders," *Neural computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[15] C. Meng, Y. He, Y. Song, J. Song, J. Wu, J.-Y. Zhu, and S. Ermon, "Sdedit: Guided image synthesis and editing with stochastic differential equations," *arXiv preprint arXiv:2108.01073*, 2021.

[16] 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*, vol. 45, no. 4, pp. 4713–4726, 2022.

[17] G. Batzolis, J. Stanczuk, C.-B. Schönlieb, and C. Etmann, "Conditional image generation with score-based diffusion models," *arXiv preprint arXiv:2111.13606*, 2021.

[18] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, and P. Büchler, "The virtual skeleton database: an open access repository for biomedical research and collaboration," *Journal of medical Internet research*, vol. 15, no. 11, p. e245, 2013.

[19] A. Maier, H. G. Hofmann, M. Berger, P. Fischer, C. Schwemmer, H. Wu, K. Müller, J. Hornegger, J.-H. Choi, C. Riess *et al.*, "Conrad—a software framework for cone-beam imaging in radiology," *Medical physics*, vol. 40, no. 11, p. 111914, 2013.

[20] 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," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

Noise-Augmented Deep Denoising of CT Images

Gernot Kristof, Achim Byl, Elias Eulig, and Marc Kachelrieß

***Abstract*—Denoising low dose CT images can have great advantages for the aim of minimizing patient risk, as it can help lower the effective dose to the patient, while providing constant image quality. Conventional deep denoising algorithms cannot handle the correlation between neighboring pixels or voxels, because the noise structure in CT is a resultant of the global attenuation properties of the patient and because the receptive field of denoising approaches is rather small. In this work additional noise realizations were generated, reconstructed, and used as additional input into a denoising network to guide the denoising process. The network was compared to a similar network, without additional noise augmentation. It was shown, that the noise-augmented deep denoising network outperformed the conventional denoising network. Due to the additional information about the noise correlation it was able to correctly identify noise structures, and remove them from images, while the conventional network often failed to distinguish the noise structure from anatomical structures.**

***Index Terms*—Diagnostic CT, noise reduction, dose reduction, neural networks, deep learning.**

I. Introduction

Image noise in CT is correlated between neighboring pixels, i.e. the covariance between pixels is not zero. The correlation depends on the patient's attenuation properties and on the tube current curve that was used to generate the images. For eccentric cross–sections of the patient the correlated noise appears as streaks oriented along the direction of highest x-ray attenuation. For example consider the shoulder region of a patient where noise appears as streaks oriented along the lateral direction.

Typical neural networks used for noise reduction, such as those in references [1]–[8], for example, have a small receptive field, i.e. they do not know where in the patient cross-section they currently operate. Consequently, they have no information about the spatially dependent covariance matrix and about the resulting noise correlation. Even if the receptive field was large it may be difficult for the network to leverage the physics behind the CT image formation to conduct the structure of the covariance matrix.

We here aim for a simple method to provide such covariance information in image domain and to develop and train a neural network that leverages this information. This is achieved by providing the denoising network with additional noise realizations. Our noise-augmented deep denoising (NADD) algorithm is evaluated using simulated data. We compare the results to the results of well-known denoising networks that do not use noise augmentation. All networks are trained using the same training data. Our experiments are currently limited to 2D images but will be extended to 3D volumes in the near future.

Gernot Kristof, Achim Byl, Elias Eulig, and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls-University, Heidelberg, Germany.
Corresponding author: gernot.kristof@dkfz.de

II. Method

The idea of this project is to generate, by noise injection into the rawdata followed by image reconstruction, a multitude of new noise realizations $h_c(\boldsymbol{r})$ of the CT image $g(\boldsymbol{r})$ that is to be denoised. Here we use $C = 10$ noise realizations. These are input to the NADD network, together with the original CT image that is to be denoised. With these ten additional input channels NADD can learn about the noise correlation and thereby improve its denoising capabilities.

A. Simulated Noise-Free Rawdata

Starting from 13 normal dose, i.e. low noise, clinical CT patient data sets we perform a forward projection to obtain simulated rawdata $p(\alpha, u, v)$, with α being the projection angle and u, and v being the two dimensions of the detector. We will drop the arguments of p whenever they are not necessary and whenever we feel that it is convenient for the reader. We assume p to be noise-free.

B. Noise Injection

Given a projection value p we convert it to intensities as $I = I_0 \exp(-p)$ with I_0 denoting the number of photons incident on the patient. For convenience we assume I_0 to be independent on α, u, and v, i.e. no heel effect, no bowtie filter, no tube current modulation etc.

We now inject Poisson noise by calculating a random variable N such that $I + N$ is Poisson-distributed. To mimick a realistic detector, we have to introduce correlations between neighboring detector elements. This is done by convolving $N(\alpha, u, v)$ along the u and v direction with the kernel

$$k(u,v) = \frac{1}{994}\begin{pmatrix} 1 & 2 & 4 & 2 & 1 \\ 2 & 25 & 80 & 25 & 2 \\ 3 & 100 & 500 & 100 & 3 \\ 2 & 25 & 80 & 25 & 2 \\ 1 & 2 & 4 & 2 & 1 \end{pmatrix}.$$

We had determined the kernel about one decade ago by analyzing projection data of a Siemens Somatom Flash CT system that was equipped with a Stellar detector. We denote the resulting noise image as $\hat{N}(\alpha, u, v) = N(\alpha, u, v) * k(u, v)$.

The noise-injected projection data are now given as

$$\hat{p}(\alpha, u, v) = -\ln(I(\alpha, u, v) + \hat{N}(\alpha, u, v))/I_0.$$

We denote the whole noise injection process as $\mathcal{N}(p)$. This means, we have $\hat{p}(\alpha, u, v) = \mathcal{N}(p(\alpha, u, v))$.

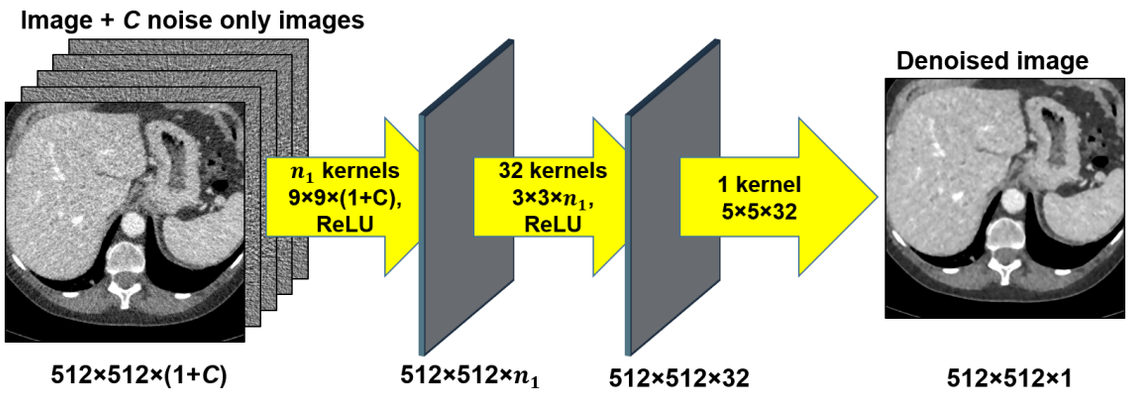


Fig. 1. Network architecture used in this work

C. Noisy Rawdata and Additional Input Channels

The noisy rawdata, that we later want to denoise in image domain, are computed as

$$q = \mathcal{N}(p),$$

i.e. by injecting noise into the simulated (nearly) noise-free rawdata that we had generated by forward projection. These noisy rawdata q can be considered to be very similar to a real measurement. The final aim would be the rawdata q to be the actually measured sinograms.

From q we generate C noise only realizations as

$$r_c = \mathcal{N}(q) - q$$

for $c = 1, \ldots, C$. Reconstructing $q(\alpha, u, v)$ yields the image $g(\boldsymbol{r})$ that we want to denoise with our network. We denote the reconstruction operator as X^{-1}, which is the inverse x-ray transform. We have $g = \mathsf{X}^{-1}q$. Here we use FBP as the reconstruction algorithm.

Reconstructing the C sinograms r_c yields C noise only images $h_c = \mathsf{X}^{-1}r_c$ that we use as an additional input channels to our network. From these images NADD shall deduce the required information about the shift-variant covariances.

For this conference contribution we use $C = 10$, which we denote as NADD_{10}. We compare the results with a conventional network, i.e. one that uses $C = 0$ but otherwise has a very similar structure and a similar number of open parameters. We denote the conventional network as NADD_0. In general, NADD_C has $1 + C$ input channels, the image g to be denoised and the C noise augmentation images h_c. The networks are detailed in the next section.

D. Network

The network architecture we used, was originally proposed by Chen et al. [1]. It is a very simple, but rather effective convolutional denoising network. The network consists of three convolutional layers and the network architecture is shown in figure 1 (if you assume $C = 0$). Originally, the network had convolution kernels of size $s_1 = 9$, $s_2 = 3$, and $s_3 = 5$ and $n_1 = 64$ and $n_2 = 32$ kernels in the first and second layer respectively.

In order to compare NADD_0 and NADD_{10} as fair as possible, we chose to adjust the number of kernels n_1 in the first layer, such that both networks have approximately the same number of trainable parameters. The total number of trainable parameters for a network with the depicted architecture and $1 + C$ input channels is given as

$$n_1 s_1^2 (1 + C) + n_1 + n_2 s_2^2 n_1 + n_2 + s_3^2 n_2 + 1.$$

For $C = 0$ we choose $n_1 = 205$, which is larger than the value of the original paper. It results in 76683 open parameters for NADD_0. For $C = 10$ we choose $n_1 = 64$, which is the value of the original paper. This yields 76353 open parameters for NADD_{10}. This a little less than for the $C = 0$ case and thus results in a slightly harder comparison for our proposed NADD_{10} network.

Both networks were trained for 50 epochs, as after that there was no significant decrease in the train loss. The loss function used was MSE loss and we chose the Adam optimizer with learning rate $\eta = 0.0001$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-9}$. For a larger sample size we trained on 32×32 patches, with a batch size of 64.

III. Results

Figure 2 shows selected slices of three different patients and zoomed in regions thereof. The two bottom slices represent the same patient at different z-positions. The normal dose images are the images $f = \mathsf{X}^{-1}p$, that have been reconstructed from the (nearly) noise-free rawdata. The low dose images are the noisy images $g = \mathsf{X}^{-1}q$. The two rightmost columns show NADD's denoising results, once without noise augmentation (NADD_0) and once with ten–fold noise augmentation (NADD_{10}).

We intentionally selected slices corresponding to quite eccentric patient cross-sections. The reason becomes clear when looking at the low dose images: Strong noise correlations are visible, appearing as streaks along the lateral direction. Sometimes the noise introduces significant artifacts that look like small hyper- or hypodense anatomical structures that are not visible in the normal dose images. The yellow arrows point towards these artifacts in the low dose and NADD_0 images and towards the corresponding spot in the normal dose and NADD_{10} images where they are not visible. NADD_0 does not recognize these structures as noise and thus does not remove them when denoising the image. In contrast,

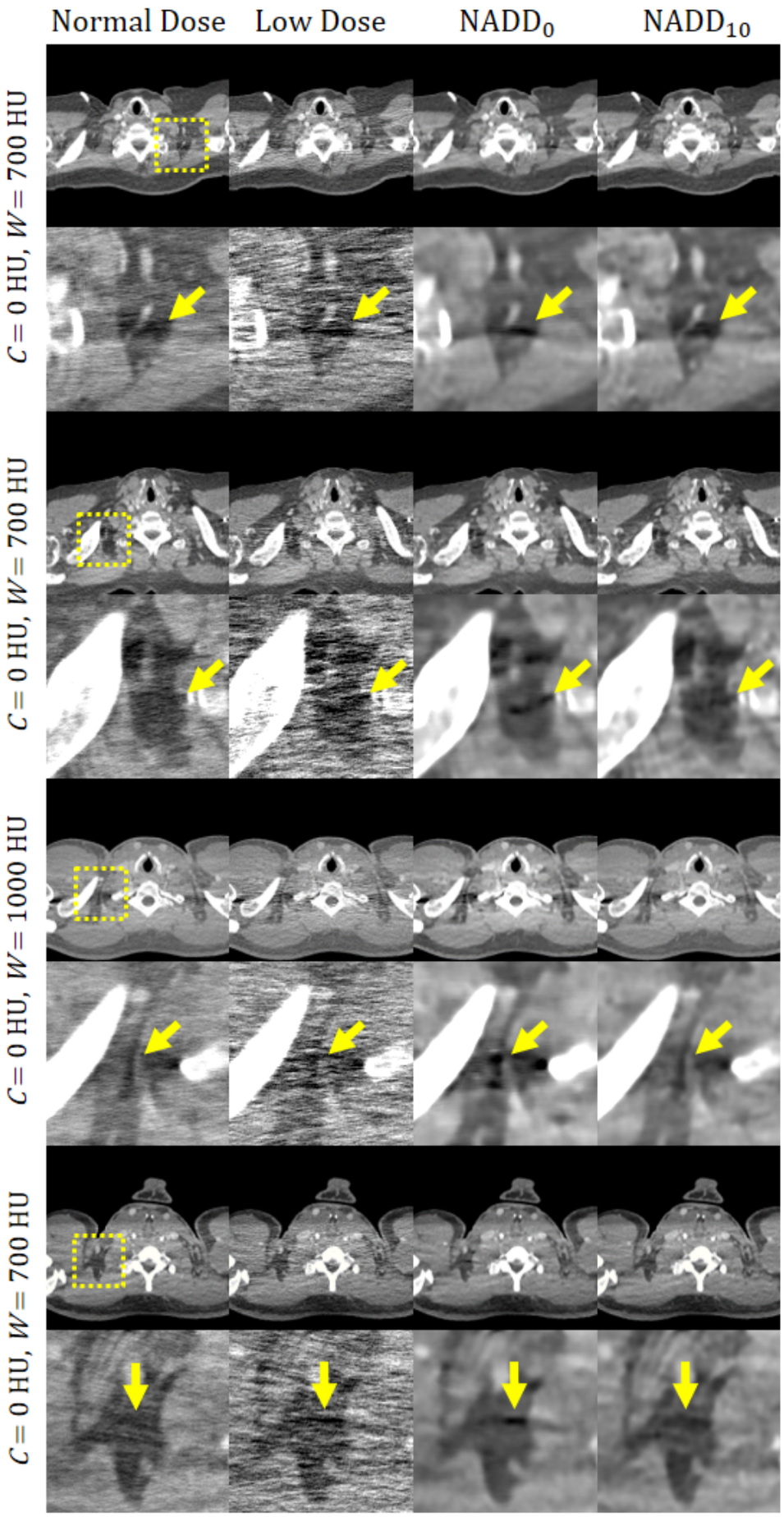


Fig. 2. Ground truth, input and predictions for the different networks. The zoom-ins show regions where the noise correlations introduce artifacts, mostly appearing as black spots (yellow arrows). NADD$_0$, which is the conventional deep denoising approach, is not able to remove those artifacts. However, NADD$_{10}$ benefits from the ten noise augmentations it receives as input.

NADD_{10} leverages the information contained in the ten noise augmentations and outperforms the conventional algorithm. The artifacts are no longer visible.

IV. Discussion

Providing the denoising network with additional noise only images helps the network to distinguish between anatomical details and noise structures. Such noise artifacts cannot be removed by conventional denoising approaches. They are, in contrast, leveraged to a great extent by NADD.

While we have analyzed the noise augmentation technique in combination with one specific denoising network it should be noted that the approach can be combined with any kind of image-based denoising network in a straight forward manner. An extension of this method towards multiple denoising network architectures will be conducted shortly.

Our study currently has several limitations. It is based on simulated projection data only. We do not yet know, how NADD performs on actual measurements. Further, we did not investigate yet how different denoising networks perform with the noise augmentation technique. Moreover, we trained our networks for the specific (medium smooth) reconstruction kernel that we had used to generate the images. It is unclear whether a new training is required for each combination of reconstruction kernel, reconstruction slice thickness and voxel size. Moreover, tube current modulation (TCM) was switched off. Since TCM is known to mitigate the strong noise correlations to some extent [9]–[11] we may find less benefit for NADD in scans with TCM. Regarding the choice of C, an augmentation study is at hand, that shall be conducted in the near future.

A potential downside of NADD is that it requires $1 + C$ image reconstructions.

V. Conclusions

Noise-augmented denoising has the potential to outperform other denoising techniques due to the added knowledge of the noise texture and the covariances.

VI. Acknowledgements

This study was supported in parts by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection (BMUV) under grant 67KI2036B, and by the Helmholtz International Graduate School for Cancer Research, Heidelberg, Germany. The projection simulation software RayConStruct-PS and the image reconstruction software RayConStruct-IR were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] H. Chen, Y. Zhang, W. Zhang, P. Liao, K. Li, J. Zhou, and G. Wang, "Low-dose CT denoising with convolutional neural network," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 143–146.

[2] H. Chen, M. K. Zhang Y. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Whang, "Low–dose CT with a residual encoder–decoder convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[3] F. Fan, H. Shan, M. K. Kalra, R. Singh, G. Qian, M. Getzin, Y. Teng, J. Hahn, and G. Wang, "Quadratic autoencoder (Q–AE) for low–dose CT denoising," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 2035–2050, Jun. 2020.

[4] Z. Huang, J. Zhang, Y. Zhang, and H. Shan, "DU–GAN: Generative adversarial networks with dual–domain U–Net–based discriminators for low–dose CT denoising," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2022.

[5] F. Wagner, M. Thies, M. Gu, Y. Huang, S. Pechmann, M. Patwari, S. Ploner, O. Aust, S. Uderhardt, G. Schett, S. Christiansen, and A. Maier, "Ultralow–parameter denoising: Trainable bilateral filter layers in computed tomography," *Med. Phys.*, vol. 49, no. 8, pp. 5107–5120, 2022.

[6] Z. Zhang, L. Yu, X. Liang, W. Zhao, and L. Xing, "TransCT: Dual-path transformer for low dose computed tomography," *MICCAI*, 2021.

[7] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and W. G., "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, Jun. 2018.

[8] A. D. Missert, S. Leng, L. Yu, and C. H. McCollough, "Noise subtraction for low–dose CT images using a deep convolutional neural network," *Proceedings of the Fifth International Conference on Image Formation in X-Ray Computed Tomography*, pp. 399–402, May 2020.

[9] 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.

[10] W. A. Kalender, H. Wolf, and C. Suess, "Dose reduction in CT by anatomically adapted tube current modulation. II. Phantom measurements," *Med. Phys.*, vol. 26, no. 11, pp. 2248–2253, Nov. 1999.

[11] 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.

Integer Optimization of CT Trajectories using a Discrete Data Completeness Formulation

Linda-Sophie Schneider, Gabriel Herl, Andreas Maier, *Member, IEEE*

***Abstract*—X-ray computed tomography (CT) plays a key role in digitizing three-dimensional structures for a wide range of medical and industrial applications. Traditional CT systems often rely on standard circular and helical scan trajectories, which may not be optimal for challenging scenarios involving large objects, complex structures, or resource constraints. In response to these challenges, we are exploring the potential of twin robotic CT systems, which offer the flexibility to acquire projections from arbitrary views around the object of interest. Ensuring complete and mathematically sound reconstructions becomes critical in such systems. In this work, we present an integer programming-based CT trajectory optimization method. Utilizing discrete data completeness conditions, we formulate an optimization problem to select an optimized set of projections. This approach enforces data completeness and considers absorption-based metrics for reliability evaluation. We compare our method with an equidistant circular CT trajectory and a greedy approach. While greedy already performs well in some cases, we provide a way to improve greedy-based projection selection using an integer optimization approach. Our approach improves CT trajectories and quantifies the optimality of the solution in terms of an optimality gap.**

***Index Terms*—Data completeness condition, arbitrary scan geometry, robot-supported CT, Integer Optimization.**

I. Introduction

X-RAY computed tomography (CT) is an important tool in both medical and industrial domains for digitizing three-dimensional structures. It facilitates critical applications such as medical diagnostics and industrial quality control. Traditional industrial CT systems often use circular and helical scan trajectories due to their simplicity and compatibility with fast reconstruction algorithms. However, these standard trajectories may not be ideal for large-scale objects, complex structures, or resource constraints [5], [6]. Furthermore, mechanical instabilities further complicate the acquisition [2]. Combining standard circular or helical trajectories with cone beam setups can lead to artifacts due to violations of the Tuy-Smith condition [1]. These degrade CT reconstruction quality, especially for complex or highly attenuating materials. This highlights the need for more adaptive and optimized CT trajectory planning approaches that consider the unique characteristics of the object being scanned.

Recent advances in CT technology have introduced twin robotic CT systems, where individual robots control the X-ray source and detector. This innovation provides flexibility by allowing projections from arbitrary views around the object of interest [3]. Robot-supported CT systems offer the potential for more tailored and complex CT trajectories, enabling efficient scanning of large objects [4], metal artifact reduction [5], and shorter scan times [6].

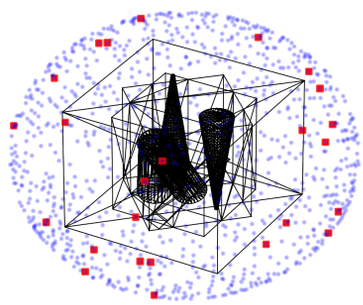


Fig. 1. Optimized Projections for Test Specimen in Section III-B. Red projections represent the 25 selected ones, while blue indicates others.

The flexibility of robotic CT systems presents challenges. Ensuring mathematically complete reconstructions becomes crucial, requiring sufficient information from the acquired projections. In the standard CT process, users select a continuous curve, often circular or helical, around the region of interest and sample it into equidistant views based on the desired number of projections. Focusing on specific regions of objects typically requires manual CT trajectory planning. While effective, this manual approach is time-consuming, subjective, and may result in suboptimal image quality.

A promising approach for performing CT trajectory optimization in terms of data completeness was introduced by Herl et al. [9]. Their approach merges the Tuy-Smith condition with the Nyquist-Shannon theorem to create a comprehensive data completeness condition. This condition can evaluate the completeness of any set of projections, making it suitable for complex scanning scenarios. However, since its calculation of data completeness is based on a greedy approach, it is highly dependent on the projections already chosen.

To overcome this limitation, we introduce an integer optimization-based (IP) approach to assess data completeness, focusing on optimizing the coverage of discretely sampled unit spheres around voxels of interest (VOIs). We employ a branch-and-cut optimization technique to improve solution quality. In addition, we can quantify the solution quality by an "optimality gap", which is the gap between the best possible objective and the best found objective. Our experiments highlight the advantages of our automated IP-based projection selection method over traditional approaches. It offers the possibility to optimize heuristically found solutions and provides a deep insight into the complexity of CT trajectory optimization.

L. Schneider and A. Maier are with the Pattern Recognition Lab, Friedrich- Alexander University Erlangen-Nuremberg, Erlangen, Germany (e-mail: linda-sophie.schneider@fau.de)

G. Herl is with the Deggendorf Institute of Technology

II. Methods

A. Continuous CT Data Completeness Condition

Computed Tomography (CT) relies on the Radon transform, a mathematical tool for translating spatial data into the Radon space. To ensure complete CT reconstruction, it is essential for the Radon space to exhibit continuity, a concept initially introduced by Radon.

In 1983 [7], Tuy introduced data completeness conditions for CT reconstruction based on continuous curves. A curve C is defined as a continuous function $C : \Theta \to \mathbb{R}^3$, where Θ represents an interval in $\mathbb{R}$. The unit sphere $\mathbb{S} \subset \mathbb{R}^3$ and Ω, a compact region encompassing the object characterized by the density function f, play crucial roles in this context.

Tuy's conditions provide the foundation for achieving data completeness in CT reconstruction. However, these conditions rely on two assumptions that, while theoretically valid, pose practical challenges. The first assumption is the *Detector Resolution Assumption* [9], in which Tuy assumes precise knowledge of the impact position of each measured photon, implying the need for a detector with infinitely small pixels. The second assumption, known as the *Source Assumption* [9], postulates that X-rays are emitted at every point along the source trajectory, effectively resulting in infinite projections.

Three critical conditions must be met to ensure accurate reconstruction of an object within the defined region Ω using data gathered along curve C. Firstly, curve C should avoid entering or intersecting Ω, remaining entirely outside of it. Secondly, curve C should exhibit desirable characteristics such as continuity, boundedness, and smoothness to ensure its predictable behavior. Lastly, at any point within Ω and in any given direction represented by the unit sphere $\mathbb{S}$, a parameter θ must exist linked to curve C. This parameter aligns with the object's information in the specified direction. Together, these conditions facilitate precise object reconstruction from projection data.

In 1985 [1], Smith introduced a simplified condition known as the *Tuy-Smith Condition*, which remains sufficient for CT reconstruction under assumptions 1 and 2 (Detector Resolution and Source Assumption). According to the Tuy-Smith Condition, an object within region Ω can be accurately reconstructed from projection data generated along curve C if, for all $(x, u) \in \Omega \times \mathbb{S}$, there exists $\theta \in \Theta$ such that $x \cdot u = C(\theta) \cdot u$.

The Tuy-Smith Condition asserts that every plane intersecting the region of interest must also intersect the source trajectory. Given that each point in the Radon space corresponds to a plane in the spatial domain, adhering to this condition guarantees comprehensive sampling of the Radon space.

As both the detector resolution assumption and the source assumption are impractical, the projections, as well as the continuous trajectory, require sampling. The Nyquist-Shannon sampling theorem is applied to ensure adequate information. This theorem dictates that, for faithful CT reconstruction, the sampling rate must be at least twice the highest frequency within the object, corresponding to fine details or sharp transitions in the object's density distribution.

B. Discrete CT Data Completeness Condition

For archiving a discrete formulation of CT data completeness, we follow the ideas of [8], which were extended by [9]. First, we require a discrete sampling of the unit sphere in vectors $(u) \in (\mathbb{S}_u)$. Following ideas from [11], we sample the sphere as uniformly as possible. It's important to note that we only need to sample half of the unit sphere, as the plane integrals exhibit symmetry for the top and bottom portions.

Next, we need to compute the data completeness in terms of Radon sphere coverage for every voxel of interest (VOI). Our objective is to identify vectors from the unit sphere u that are covered by the normal vectors of the Radon planes, i.e., vectors that are perpendicular to the current viewing direction $d \in D$, which represents the connection between the source and the current voxel. We only consider those vectors d that actually hit the detector.

The Nyquist-Shannon sampling theorem states that in Radon space, the angular separation between adjacent planes need only be less than the maximum angular gap $\Delta\gamma$ [9]. In addition, since the inner product of two unit vectors describes the cosine of the angle between them, and we want to find vectors that deviate by less than $\Delta\gamma$, the following, already reformulated, condition holds for the vectors of interest [10]:

$$|d^T u| < \sin(\Delta\gamma) \tag{1}$$

It's worth noting that the sine function can be precomputed since $\Delta\gamma$ remains constant in our case. The above equation is evaluated for every unit vector and every viewing direction through the VOI.

C. Integer Program for CT Trajectory Optimization

The discrete formulation of a CT data completeness condition is used to model an integer program, which allows us to find an optimized set of k viewing directions $D_k \subseteq D$. Therefore, we sample a unit sphere in vectors $u \in \mathbb{S}_u^V$ for a VOI V. Using Equation (1), we precompute for every direction $d \in D$ which unit vectors $u \in \mathbb{S}^V$ fulfills Equation (1). With that information, we build up a matrix $S_V \in \{0, 1\}^{|D| \times |\mathbb{S}^V|}$, where an entry of 1 indicates that Equation (1) holds. Generally, this formulation can be extended to multiple VOIs by sampling a unit sphere for each VOI.

Additionally, we use the idea of [5] to limit the number of selectable viewing directions d by incorporating an absorption-based metric into our model. Therefore, we select a threshold α indicating the percentage of photons still reaching the detector after passing through the object. The absorption for a viewing direction is evaluated in the projection domain by calculating the average normalized pixel value in the regions of interest indicated by the voxels of interest. This helps us assess the viewing direction's reliability for the respective VOIs. This metric can be precomputed, and the results are saved in a $|D| \times 1$-matrix, denoted by A_D. Instead of absorption, alternative projection-based metrics constrained by threshold values can be employed, increasing the optimization process's flexibility and enabling task-specific adjustments.

Combining all the information acquired, we formulate the selection x of k viewing directions as a convex integer optimization program. Our objective is to maximize the coverage $C = \sum_{i=1}^{|\mathbb{S}^V|} c_i$ of the unit sphere.

$$\text{Maximize:} \quad \sum_{i=1}^{|\mathbb{S}^V|} c_i \tag{2}$$

$$\text{Subject to:} \quad \sum_{i=1}^{|D|} x_i = k \tag{3}$$

$$A_D \cdot x \leq \alpha \tag{4}$$

$$c \leq {S_V}^T \cdot x \tag{5}$$

$$x, c \in \{0, 1\}. \tag{6}$$

D. Reference Methods

The effectiveness of the proposed method is compared to two reference methods. First, we use an equidistant circular CT trajectory as a reference since this is one of the state-of-the-art CT trajectories for non-destructive testing. To make a fair comparison, we chose the same number k of viewing directions as the IP-based approach. Because of the equidistant sampling, the absorption-based metric is not considered. The second reference method selects k viewing direction with a greedy approach, trying to maximize the coverage of the unit sphere vectors $\mathbb{S}_u^V$ by selecting the viewing direction d, which contributes most to the overall coverage C.

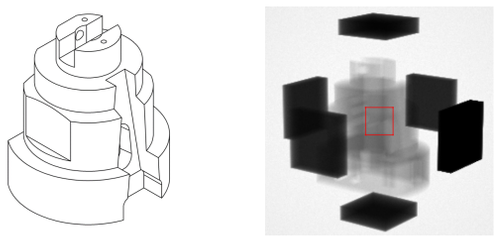
III. Experiments and Results

We used Python and the Gurobi optimizer for our CT trajectory optimization. Given the potential exponential runtime of integer programs, we incorporated a callback method into the optimizer. This method monitors the progress of the optimization and intervenes if necessary. Specifically, it checks whether significant progress has been made in reducing the optimality gap within the last $20\,\mathrm{s}$. If the improvement fell below 1×10^{-8}, the optimization was terminated prematurely, even if this resulted in a suboptimal solution. This decision is consistent with the observation that extended optimization efforts often yield negligible improvements within a reasonable timeframe, emphasizing timely results while maintaining acceptable solution quality.

Both experiments were evaluated using identical metrics, comparing the reconstructed volume between the sparse CT trajectory and the trajectory including all available projections. For reconstruction we used the Algebraic Reconstruction Technique because of its capability to reconstruct arbitrary orbits. The quantitative evaluation included the calculation of the Structural Similarity Index (SSIM), the contrast-to-noise ratio (CNR) and the Peak Signal to Noise Ratio (PSNR) within the ROI.

A. CT Trajectory Optimization for Metal Artifact Reduction

In this study, we aimed to visualize the internal structure of a $3\,\mathrm{cm}$ carbon sample, as shown in Figure 2. To add complexity, we placed the carbon sample between six $1.5\,\mathrm{cm}$ wide and



(a) (b)

Fig. 2. Overview of the test specimen of Experiment III-A. (a) Sketch of the carbon test specimen. (b) Simulated projection of the test specimen, red box markes the projected VOI [10].

$0.4\,\mathrm{cm}$ thick iron plates, resulting in a highly attenuating and symmetrical configuration. This setup, similar to a previous experiment in [10], aimed to showcase the effectiveness of a CT trajectory optimization approach. Our dataset consisted of 51 tilt angles ranging from $-90°$ to $90°$ relative to the horizontal axis, enabling projections over a half-sphere. For each tilt angle, we simulated 61 evenly spaced projections within a $216°$ range. The simulations incorporated a polychromatic spectrum, simulated noise, and excluded scattering effects. Scan parameters included a $150\,\mathrm{kV}$ voltage, no prefilter, a pixel size of $700\,\mu\mathrm{m}$, a $18\,\mathrm{cm} \times 18\,\mathrm{cm}$ detector, and a magnification factor of 3.

After examining the simulated projections, we adjusted the α parameter to exclude X-rays with excessive attenuation, particularly those passing through metal plates in our experimental setup. This modification set α to a value of 0.7, indicating that, on average, at least 30% of the photons in our region of interest (ROI) successfully penetrated the object. Consequently, this constraint reduced the available selectable projections from 3111 to 1851. For our projection selection method, we aimed to select a set of 61 projections, corresponding to the number of viewpoints in each $216°$ circle. To facilitate comparative analysis, we chose the untilted circle trajectory as the most fundamental within our projection set. We chose the VOI as the center of the object and sampled a unit sphere around it, resulting in $|\mathbb{S}_u^V| = 2000$ unit sphere points and a maximum angular distance of $\Delta\gamma = 0.01$, a choice informed by the methodology outlined in [9].

TABLE I
Comparison of Quality Metrics for 61 Projections

Approach	SSIM↑	PSNR↑	CNR↑	Coverage↑
Circular	0.60	118.7db	1.12	45%
Greedy	0.62	118.8db	3.3	64%
IP	0.72	119.9db	4.4	91%

The IP approach performs best in terms of image quality and data completeness and has a higher SSIM (0.72) and PSNR (119.9 dB) than the other methods. In addition, it is better at selecting projections, covering 91% of the unit sphere, while Greedy achieves only 64%. The small optimality gap of 1.79% underlines the IP approach's potential as a promising choice for CT trajectory design optimization.

B. CT Trajectory Optimization for an Arbitrary VOI

We artificially generated an empty cube for our second experiment and placed multiple geometric shapes inside, as visible in Figure 1. We simulated 1000 projections of a full sphere with the object placed in the center to ensure variable viewing positions. Our simulations incorporated a polychromatic spectrum and simulated noise, omitting the simulation of scattering effects. Additional scan parameters included a voltage of $150\,\mathrm{kV}$, no pre-filter, a pixel size of $600\,\mu\mathrm{m}$, a detector size of $15.4\,\mathrm{cm}\times15.4\,\mathrm{cm}$, and a magnification factor of 2. The imaging task was to enhance reconstruction quality for the cylindrical object inside the cube. After simulating the projections, we selected $\alpha = 0.84$ to exclude high attenuation X-rays. This resulted in a reduction of the available projections from 1000 to 535.

We evaluated our method on $k = 25$ projections. We simulated an equidistantly sampled 360° circular trajectory for comparison. To test the performance of the method on non-central VOIs, we chose a voxel on the cylinder shaft as the VOI. We sampled a unit sphere around the VOI with $|\mathbb{S}_u^V| = 1500$ points and a maximum angular distance of $\Delta\gamma = 0.015$.

TABLE II
COMPARISON OF QUALITY METRICS FOR 25 PROJECTIONS

Approach	SSIM↑	PSNR↑	CNR↑	Coverage↑
Circular	0.5	113.7db	19	34.5%
Greedy	0.57	115.3db	26	48.1%
IP	0.59	113.9db	48	73.6%

The results presented in Table II and Figure 3 show the performance of the IP approach in a specific region of interest and thereby do not reflect the overall image quality. The IP approach achieves an SSIM of 0.59 and a PSNR of 113.9dB, suggesting a competitive image quality compared to the reference methods. In addition, it outperforms its counterparts in terms of CNR, indicating improved image contrast. It also shows better coverage, encompassing 73.6% compared to 48.1% of the unit sphere. With a small optimality gap of 0.75%, the IP approach demonstrates its ability to provide near-optimal solutions.

IV. CONCLUSION

In this study, we introduced an IP-based CT trajectory optimization method to address the challenges of data completeness and projection selection in CT. Our approach uses discrete data completeness conditions to formulate an optimization problem for selecting an optimized set of viewing directions. This method ensures data completeness while considering absorption-based metrics for reliable evaluation of the selected projections.

Experimental results demonstrate the effectiveness of our approach in improving the quality of CT reconstructions. In both experiments, our optimized CT trajectory outperformed the standard and greedy trajectory regarding image quality. In addition, we used the concept of an optimality gap to quantify optimality in terms of our discrete data completeness formulation. The experiments emphasized the importance of careful projection selection.

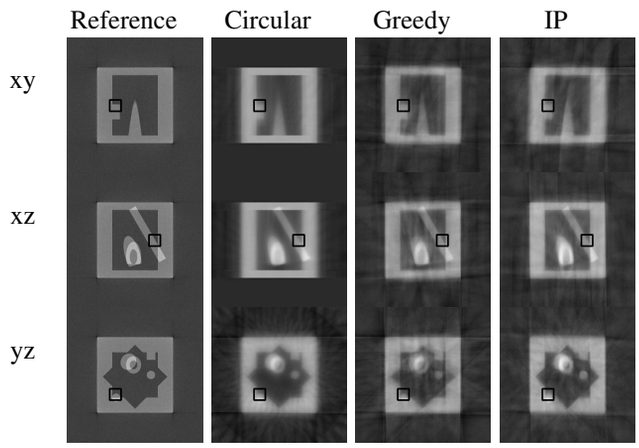


Fig. 3. Optimized CT trajectory for the marked region. The optimized and circular CT trajectories consist of $k = 25$ projections, while the reference data was a spherical CT trajectory with 1000 projections.

Future work includes studying parameter influence, refining the IP formulation for faster convergence, and exploring additional projection-based metrics for CT trajectory optimization.

ACKNOWLEDGMENT

This research was financed by the "SmartCT – Methoden der Künstlichen Intelligenz für ein autonomes Roboter-CT System" project (project nr. DIK-2004-0009).

REFERENCES

[1] B. D. Smith, "Image reconstruction from cone-beam projections: necessary and sufficient conditions and reconstruction methods," *IEEE Transactions on Medical Imaging*, vol. 4, no. 1, pp. 14–25, 1985.

[2] A. Maier, J-H. Choi, A. Keil, C. Niebler, M. Sarmiento, A. Fieselmann, G. Gold, S. Delp, and R. Fahrig, "Analysis of vertical and horizontal circular C-arm trajectories," in *Medical Imaging 2011: Physics of Medical Imaging*, vol. 7961, pp. 602–609, SPIE, 2011.

[3] P. Landstorfer, G. Herl, and J. Hiller, "Investigation of Non-circular Scanning Trajectories in Robot-based Industrial X-ray Computed Tomography of Multi-material Objects," in *16th International Conference on Informatics in Control, Automation and Robotics*, 2019.

[4] W. Holub, F. Brunner, and T. Schön, "Roboct-application for in-situ inspection of join technologies of large scale objects," in *International Symposium on Digital Industrial Radiology and Computed Tomography*, 2019.

[5] G. Herl, J. Hiller, and A. Maier, "Scanning trajectory optimization using a quantitative Tuy-based local quality estimation for robot-based X-ray computed tomography," *Nondestructive Testing and Evaluation*, vol. 35, no. 3, pp. 287–303, 2020.

[6] F. Bauer, M. Goldammer, and C. U. Grosse, "Scan time reduction by fewer projections-an approach for part-specific acquisition trajectories," in *20th World Conference on Non-Destructive Testing*, 2020.

[7] H. K. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM Journal on Applied Mathematics*, vol. 43, no. 3, pp. 546–552, 1983.

[8] A. Maier, P. Kugler, G. Lauritsch, and J. Hornegger, "Discrete Estimation of Data Completeness for 3D Scan Trajectories with Detector Offset," in *Bildverarbeitung für die Medizin 2015*, Springer, 2015, pp. 47–52.

[9] G. Herl, A. Maier, and S. Zabler, "X-ray CT Data Completeness Condition for Sets of Arbitrary Projections," in *7th International Conference on Image Formation in X-Ray Computed Tomography*, 2022.

[10] G. Herl, "Multipositional X-ray Tomography for Avoidance and Reduction of Image Artefacts," Ph.D. thesis, 2022.

[11] Á. González, "Measurement of Areas on a Sphere Using Fibonacci and Latitude–Longitude Lattices," *Mathematical Geosciences*, vol. 42, no. 1, pp. 49–64, 2009.

Enhancing Cross-Modality Synthesis: Subvolume Merging for MRI-to-CT Conversion

Fuxin Fan, Jingna Qiu, Yixing Huang, and Andreas Maier

***Abstract*—Providing more precise tissue attenuation information, synthetic computed tomography (sCT) generated from magnetic resonance imaging (MRI) contributes to improved radiation therapy treatment planning. In our study, we employ the advanced SwinUNETR framework for synthesizing CT from MRI images. Additionally, we introduce a three-dimensional subvolume merging technique in the prediction process. By selecting an optimal overlap percentage for adjacent subvolumes, stitching artifacts are effectively mitigated, leading to a decrease in the mean absolute error (MAE) between sCT and the labels from 52.65 HU to 47.75 HU. Furthermore, implementing a weight function with a gamma value of 0.9 results in the lowest MAE within the same overlap area. By setting the overlap percentage between 50% and 70%, we achieve a balance between image quality and computational efficiency.**

***Index Terms*—CT synthesis, Swin transformer, subvolume merging.**

I. Introduction

COMPUTED Tomography (CT) and Magnetic Resonance Imaging (MRI) play crucial roles in radiotherapy planning [1]. CT images are essential for generating accurate electron density maps, which are vital for precise dose calculations [2]. Conversely, MRI is effective at offering superior soft tissue contrast, enhancing target segmentation in treatment planning [3]. Consequently, aligning MRI with CT images (MRI-to-CT registration) is a critical step. However, this process is susceptible to registration errors, potentially compromising treatment reliability [4]. To circumvent the MRI-to-CT registration errors, synthesizing CT from MRI images emerges as a viable alternative. Moreover, employing synthetic CT (sCT) instead of traditional CT scans can reduce additional radiation exposure to patients and lower treatment costs [5].

With the development of deep learning technologies, a diverse range of MRI-to-CT conversion algorithms has been proposed. The U-Net, known for its symmetrical design, stands out as an effective model for CT synthesis [6], [7]. Additionally, various Generative Adversarial Networks (GANs), which incorporate discriminative loss, have been developed specifically for MRI-to-CT translation [4], [8], [9]. Networks based on Vision Transformers (ViT) with attention mechanisms have also demonstrated impressive performance in this area [10], [11]. The recent SynthRad2023 challenge has contributed to this field by making registered MRI and CT images from multiple institutions available to researchers over the world [12], [13].

Despite these developments, most existing studies have primarily concentrated on network architecture, often overlooking the potential enhancements in the postprocessing stage. Our paper addresses this gap by introducing a novel approach to CT synthesis from MRI through subvolume training, utilizing the SwinUNETR structure [14]. Crucially, we present an innovative subvolume merging strategy during the prediction phase, aiming to further enhance the quality of the synthesized CT images.

F. Fan and A. Maier are with the Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. J. Qiu is with the Department of Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. Y. Huang is with the Department of Radiation Oncology, University Hospital Erlangen, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany. Contact E-mail: (fuxin.fan@fau.de).

II. Methods and Materials

A. Neural network

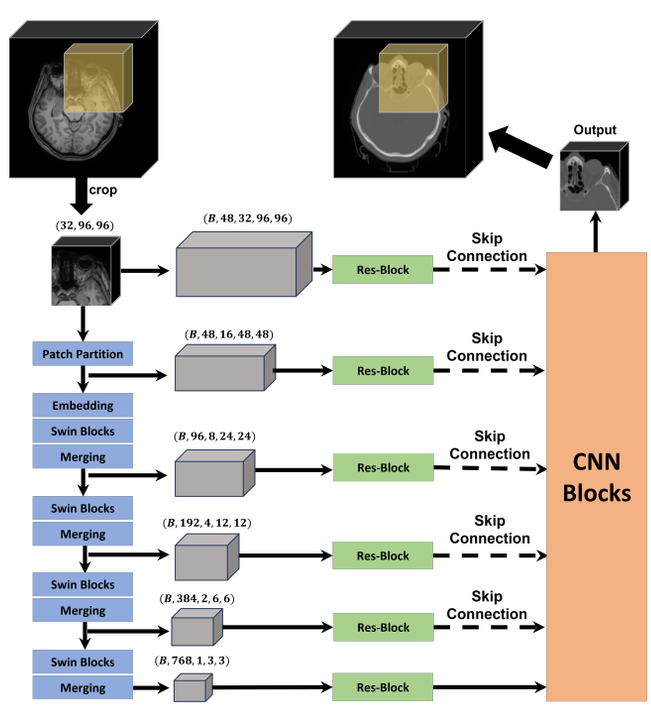


Fig. 1: The network structure of SwinUNETR.

In this work, we use a state-of-the-art network, SwinUNETR, for MRI-to-CT synthesis. The implementation of the SwinUNETR is available under the open-source framework MONAI [15]. The architecture of the SwinUNETR is shown in Fig. 1. This network consists of a Shift window (Swin) Vision transformer (ViT)-based encoder and a CNN-based

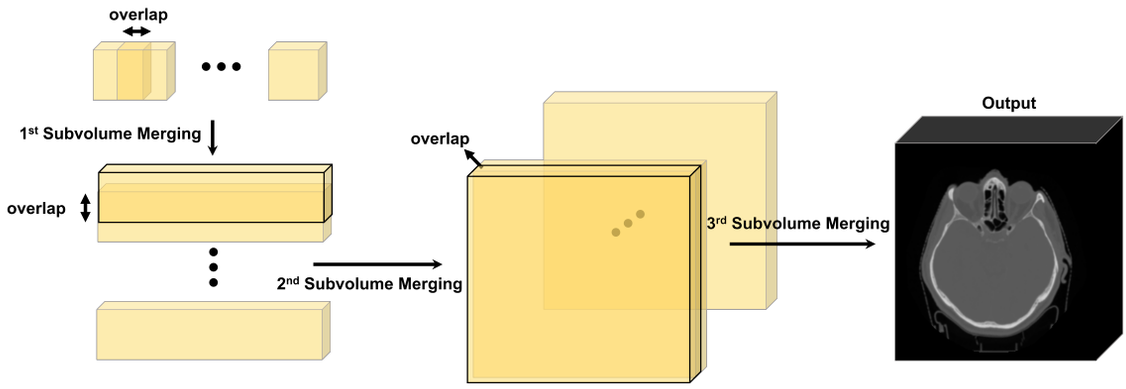


Fig. 2: 3D subvolume merging during the inference process.

decoder. A subvolume of size $32 \times 96 \times 96$ is randomly selected from an MRI volume and fed into the network. The SwinUNETR splits the subvolume into a sequence of patches, with the size of $2 \times 2 \times 2$. Each patch is embedded into a vector with the feature dimension of 48. The patch sequence then goes through four stages, and each stage has 2 Swin blocks followed by a patch merging operation. After patch merging, the side length of one patch is doubled. At the same time, the output dimension is also doubled. The output from each stage is reshaped and forwarded into a residual block before concatenating with CNN-based blocks. The residual block consists of two $2 \times 2 \times 2$ convolutional layers followed by an instance normalization layer. In each CNN block, the concatenated features are fed into another residual block and a deconvolutional layer. The feature size gets halved after the deconvolutional layer. The final outputs with a single channel are computed by using a $1 \times 1 \times 1$ convolutional layer.

The network was trained using an NVIDIA A100 GPU with 80 GB memory. The model is trained on 144 patient cases in each epoch. For each case, 20 subvolumes are randomly selected for each epoch. The predictions and labels are pixel-wisely multiplied with their corresponding binary masks before loss calculation. The L1 loss function and the Adam optimizer are used. The values for β_1 and β_2 are 0.9 and 0.999. The models are trained for maximum 1000 epochs, and the training stops when the validation loss is not decreased for three continuous epochs. The learning rate has stepwise decay from 0.0005 to 0.00005.

B. Subvolume merging

To reduce the inference time, only subvolumes within binary masks are predicted. The whole CT volume is constructed by merging adjacent subvolumes sequentially in three dimensions. The merging process is shown in Fig. 2. The smallest subvolumes construct long cuboids. Then long cuboids are connected with each other to build flat cubes. The CT volume is then obtained by merging all flat cubes together. The overlap areas of adjacent subvolumes are multiplied with two weight maps to keep smooth intensity transition. For the same voxel from two adjacent subvolumes a and b, the merged intensity I is the weighted summation of I_A and I_B:

$$I_j = (1 - w_j) I_{j,A} + w_j I_{j,B}. \tag{1}$$

The weight w_j satisfies an exponential function:

$$w_j = (j/N)^\gamma, \tag{2}$$

where N is the overlap length along one specific direction and j is the voxel index along the overlap direction. When $\gamma = 1$, a linear weight is used. As shown in Eq. (1), the weight for the former subvolume A decreases from 1 to 0 along the direction, whereas the weight for the latter increases from 0 to 1 complementarily. In addition, the overlap percentage ranging between 0 and 1 is defined as the ratio between the overlap length and the image length.

C. Data set preparation

In this study, we utilized 180 paired MRI and CT brain volumes sourced from three different institutions [12], [13]. These were divided into 144 cases for training the model, 18 for validation, and 18 for testing. To standardize the data, each volume from the brain and neck region was preprocessed to achieve a uniform voxel size of $1\,\text{mm} \times 1\,\text{mm} \times 1\,\text{mm}$. Furthermore, MRI and CT pairs were pre-registered by the organizers.

For data normalization, MRI intensity values were scaled down by a factor of 1000. CT values were first adjusted to be nonnegative by subtracting the minimum value of each volume (usually around -1000 HU) and then scaled down by a factor of 2000.

Subsequently, MRI subvolumes of size $32 \times 96 \times 96$, along with their corresponding binary masks and CT subvolumes, were randomly selected for network training. The MRI subvolumes were pixel-wisely multiplied with their binary masks before being fed into the network. These binary masks, provided for all cases, outline the patient's segmentation and the areas within these segmentations were the focus of our evaluation. We employed Mean Absolute Error (MAE) and Peak Signal-to-Noise Ratio (PSNR) as our primary evaluation metrics.

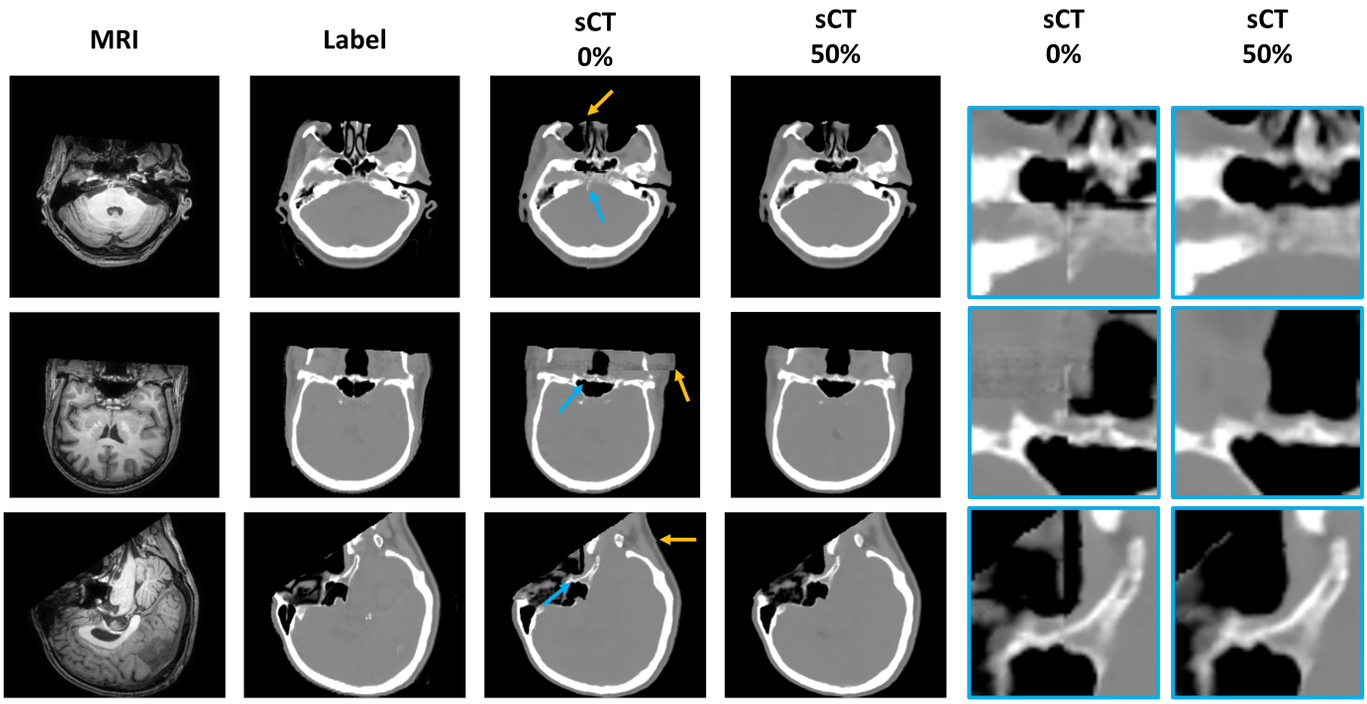


Fig. 3: The result of sCT before and after subvolume merging. The right two columns show the enlarged regions from the 3rd and 4th columns. Such regions are highlighted by blue arrows in the third column. The intensity window for CT images is [-500, 500] HU.

III. Results and Discussion

Fig. 3 illustrates the synthetic CT (sCT) images produced through subvolume merging with 0% and 50% overlap percentages. Notably, stitching artifacts, indicated by orange and blue arrows, are visible in the images with 0% overlap. In the enlarged regions near blue arrows, such artifacts are more distinctive. In contrast, increasing the overlap percentage to 50% effectively eliminates these artifacts.

The graph displayed in Fig. 4 depicts how the MAE varies with different values of γ. This analysis was performed with the overlap percentage fixed at 50%. Initially, as γ increases, there is a decrease in the MAE. However, beyond a certain point, the MAE begins to rise. The lowest MAE is observed when γ is set to 0.9.

Fig. 5 presents how image metrics and the average number of subvolumes for test cases vary with different overlap percentages. Generally, as the overlap percentage increases, there is a noticeable improvement in image quality: the MAE drops from 52.65 HU to 47.75 HU, and the PSNR rises from 27.84 to 28.65. However, this increase in overlap percentages also leads to a considerable rise in the average number of subvolumes, expanding from 53 to 36,318, thereby considerably increasing the computational time required for inference.

When models are trained using subvolumes, the process of subvolume merging plays a crucial role in shaping the final inference results. This method enhances image quality in both measurable and observable terms, which is advantageous for subsequent radiotherapy treatment planning. However, choosing a high overlap percentage can lead to an increase in computational time. Thus, finding a balance between image quality and inference speed is essential. Our experiments indicate that an overlap percentage between 50% and 70% is optimal. In this range, the MAE shows minimal variation, ranging from 48.12 HU to 48.96 HU. Moreover, the number of subvolumes falls between 178 and 508, which is 3.5 to 9.5 times bigger than that without overlap, but still manageable in terms of computational demand.

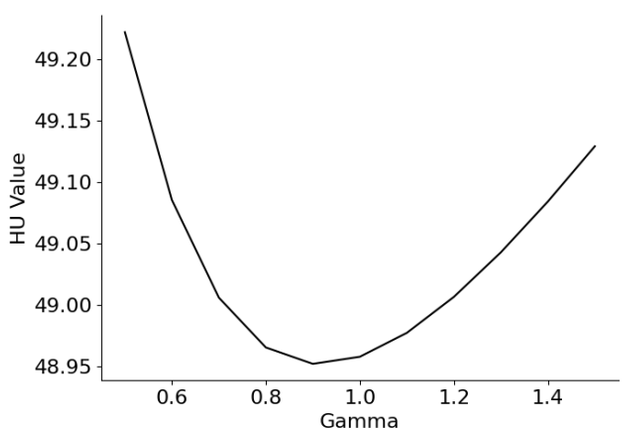


Fig. 4: MAE with respect to Gamma values.

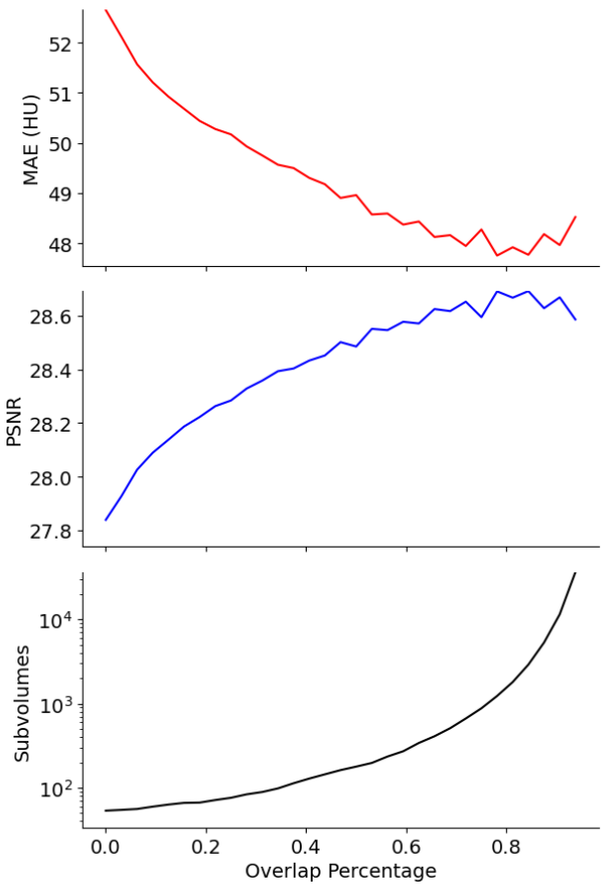


Fig. 5: Ablation study on overlap percentage.

IV. Conclusion

In this study, we introduce the use of subvolume merging in the inference process, a technique that produces higher quality images compared to subvolume joining methods. By opting for an overlap percentage between 50% and 70%, we are able to maintain computational efficiency while ensuring high image quality. This approach has broader applicability and could be effectively implemented in other regression tasks that require subimage or subvolume training.

Acknowledgment

The authors acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU).

References

[1] J. Fu, K. Singhrao, M. Cao, V. Yu, A. P. Santhanam, Y. Yang, M. Guo, A. C. Raldow, D. Ruan, and J. H. Lewis, "Generation of abdominal synthetic cts from 0.35 t mr images using generative adversarial networks for mr-only liver radiotherapy," *Biomedical Physics & Engineering Express*, vol. 6, no. 1, p. 015033, 2020.

[2] K. N. Brou Boni, J. Klein, A. Gulyban, N. Reynaert, and D. Pasquier, "Improving generalization in mr-to-ct synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data," *Medical physics*, vol. 48, no. 6, pp. 3003–3010, 2021.

[3] S. Dayarathna, K. T. Islam, S. Uribe, G. Yang, M. Hayat, and Z. Chen, "Deep learning based synthesis of mri, ct and pet: Review and analysis," *Medical Image Analysis*, p. 103046, 2023.

[4] Y. Lei, J. Harms, T. Wang, Y. Liu, H.-K. Shu, A. B. Jani, W. J. Curran, H. Mao, T. Liu, and X. Yang, "Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks," *Medical physics*, vol. 46, no. 8, pp. 3565–3581, 2019.

[5] A. M. Owrangi, P. B. Greer, and C. K. Glide-Hurst, "Mri-only treatment planning: benefits and challenges," *Physics in Medicine & Biology*, vol. 63, no. 5, p. 05TR01, 2018.

[6] M. F. Spadea, G. Pileggi, P. Zaffino, P. Salome, C. Catana, D. Izquierdo-Garcia, F. Amato, and J. Seco, "Deep convolution neural network (dcnn) multiplane approach to synthetic ct generation from mr images—application in brain proton therapy," *International Journal of Radiation Oncology* Biology* Physics*, vol. 105, no. 3, pp. 495–503, 2019.

[7] G. Dovletov, D. D. Pham, S. Lörcks, J. Pauli, M. Gratz, and H. H. Quick, "Grad-cam guided u-net for mri-based pseudo-ct synthesis," in *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2022, pp. 2071–2075.

[8] P. Qian, K. Xu, T. Wang, Q. Zheng, H. Yang, A. Baydoun, J. Zhu, B. Traughber, and R. F. Muzic, "Estimating ct from mr abdominal images using novel generative adversarial networks," *Journal of Grid Computing*, vol. 18, pp. 211–226, 2020.

[9] B. Sun, S. Jia, X. Jiang, and F. Jia, "Double u-net cyclegan for 3d mr to ct image synthesis," *International Journal of Computer Assisted Radiology and Surgery*, vol. 18, no. 1, pp. 149–156, 2023.

[10] Y. Li, S. Xu, Y. Lu, and Z. Qi, "Ct synthesis from mri with an improved multi-scale learning network," *Frontiers in Physics*, vol. 11, p. 1088899, 2023.

[11] B. Zhao, T. Cheng, X. Zhang, J. Wang, H. Zhu, R. Zhao, D. Li, Z. Zhang, and G. Yu, "Ct synthesis from mr in the pelvic area using residual transformer conditional gan," *Computerized Medical Imaging and Graphics*, vol. 103, p. 102150, 2023.

[12] A. Thummerer, E. van der Bijl, A. Galapon Jr, J. J. Verhoeff, J. A. Langendijk, S. Both, C. N. A. van den Berg, and M. Maspero, "Synthrad2023 grand challenge dataset: Generating synthetic ct for radiotherapy," *Medical Physics*, 2023.

[13] A. Thummerer, E. Huijben, M. Terpstra, O. Gurney-Champion, M. Afonso, S. Pai, P. Koopmans, M. van Eijnatten, Z. Perko, and M. Maspero, "SynthRAD2023 Challenge design: Synthesizing computed tomography for radiotherapy," Mar. 2023. [Online]. Available: https://doi.org/10.5281/zenodo.7781049

[14] A. Hatamizadeh, V. Nath, Y. Tang, D. Yang, H. R. Roth, and D. Xu, "Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images," in *International MICCAI Brainlesion Workshop*. Springer, 2021, pp. 272–284.

[15] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang *et al.*, "Monai: An open-source framework for deep learning in healthcare," *arXiv preprint arXiv:2211.02701*, 2022.

Noise Controlled CT Super-Resolution with Conditional Diffusion Model

Yuang Wang, Siyeop Yoon, Rui Hu, Baihui Yu, Duhgoon Lee, Rajiv Gupta, Li Zhang, Zhiqiang Chen, and Dufan Wu

***Abstract*—Improving the spatial resolution of CT images is a meaningful yet challenging task, often accompanied by the issue of noise amplification. This article introduces an innovative framework for noise-controlled CT super-resolution utilizing the conditional diffusion model. The model is trained on hybrid datasets, combining noise-matched simulation data with segmented details from real data. Experimental results with real CT images validate the effectiveness of our proposed framework, showing its potential for practical applications in CT imaging.**

***Index Terms*—Super-Resolution, Conditional Diffusion Model, Noise Controlling**

I. Introduction

MEDICAL imaging, especially Computed Tomography (CT), is crucial for diagnosing and treating health conditions. Achieving higher resolution in CT scans is an ongoing challenge, and super-resolution techniques play a vital role in enhancing spatial resolution. This improvement promises more detailed information for clinicians, leading to better diagnostic accuracy and improved patient care. Diffusion models [1], [2], rooted in probabilistic modeling and diffusion processes, offer significant advancements in natural imaging tasks, including super-resolution [3]–[5]. Their stability, unlike Generative Adversarial Networks (GANs), ensures reliable image generation without the pitfalls of mode collapse or unrealistic artifacts. In the context of CT super-resolution, diffusion models stand as a source of inspiration, paving the way for novel developments in medical imaging.

Achieving super-resolution in CT images presents a complex challenge in noise control, particularly when compared to natural images. The training of conditional diffusion models for super-resolution necessitates paired sets of high resolution (HR) and low resolution (LR) CT images. Achieving super-resolution without noise amplification requires meticulous matching of noise levels in HR and LR CT images. The scarcity of noise-matched pairs, prompted by radiation exposure, has led certain methodologies to resort to downsampling HR CT images in either the image domain or projection domain to obtain corresponding LR CT images for model training [6]. However, these approaches inadvertently exacerbate noise while improving spatial resolution. Some techniques introduce Gaussian noise to LR CT images [7] before or after downsampling to harmonize with the noise level, yet their efficacy in super-resolution may waver when applied to real data due to the introduction of unrealistically distributed noise.

In this article, we present an innovative framework for noise-controlled CT super-resolution, utilizing a conditional diffusion model trained on hybrid datasets. Numerical phantoms are employed to generate noise-matched simulation pairs of HR and LR CT images. Furthermore, details absent in numerical phantoms are segmented from noise-unmatched real pairs and integrated into the training process. Testing using real CT images validates the effectiveness of the proposed framework in real-world scenarios.

Y. Wang, S. Yoon, R. Hu, B. Yu, R. Gupta, and D. Wu are (were) with the Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston MA 02114, USA. E-mail: (dwu6@mgh.harvard.edu).

D. Lee is with Neurologica Corp., Danvers MA 01923, USA.

Y. Wang, L. Zhang, and Z. Chen are with the Department of Engineering Physics, Tsinghua University, Beijing 100084, China.

II. Method

A. Conditional Diffusion Model

In our article, we employ the Conditional Denoising Diffusion Probabilistic Model (DDPM) [4], [5] for super-resolution in CT images. Here, the set of LR images serves as the condition y, and the set of HR images forms the generation target x_0. The Conditional DDPM contains a forward and reverse processes. The forward process, a Markov chain, gradually introduces noise into the image until it becomes a standard Gaussian noise image, with the following process:

$$q\left(x_t|x_{t-1}\right) = N\left(x_t|\sqrt{1-\beta_t}x_{t-1}, \beta_t I\right), \tag{1}$$

where t is the diffusion time step from 0 to T, and β_t is a small positive hyperparameter determining the speed of the diffusion. x_0 is the original image and x_T approximately follows the standard normal distribution.

From (1), we can derive $q\left(x_t|x_0\right)$ and $q\left(x_{t-1}|x_0, x_t\right)$:

$$q\left(x_t|x_0\right) = N\left(x_t|\sqrt{\gamma_t}x_0, \left(1-\gamma_t\right) I\right), \tag{2}$$

$$q\left(x_{t-1}|x_t, x_0\right) = N\left(x_{t-1}|\widetilde{\mu}_t\left(x_t, x_0\right), \widetilde{\beta}_t I\right), \tag{3}$$

where

$$\gamma_t = \prod_{s=1}^{t} \alpha_s, \tag{4}$$

$$\alpha_t = 1 - \beta_t, \tag{5}$$

$$\widetilde{\mu}_t\left(x_t, x_0\right) = \frac{\sqrt{\gamma_{t-1}}\beta_t}{1-\gamma_t}x_0 + \frac{\sqrt{\alpha_t}\left(1-\gamma_{t-1}\right)}{1-\gamma_t}x_t, \tag{6}$$

and

$$\widetilde{\beta}_t = \frac{1-\gamma_{t-1}}{1-\gamma_t}\beta_t. \tag{7}$$

The reverse process, a Markov chain aimed at generating x_0 from $x_T \sim N(0, I)$, is modeled as a Gaussian process p_θ:

$$p_\theta\left(x_{t-1}|x_t, y\right) = N\left(x_{t-1}|\mu_\theta\left(x_t, y, t\right), \sigma_t^2\right), \tag{8}$$

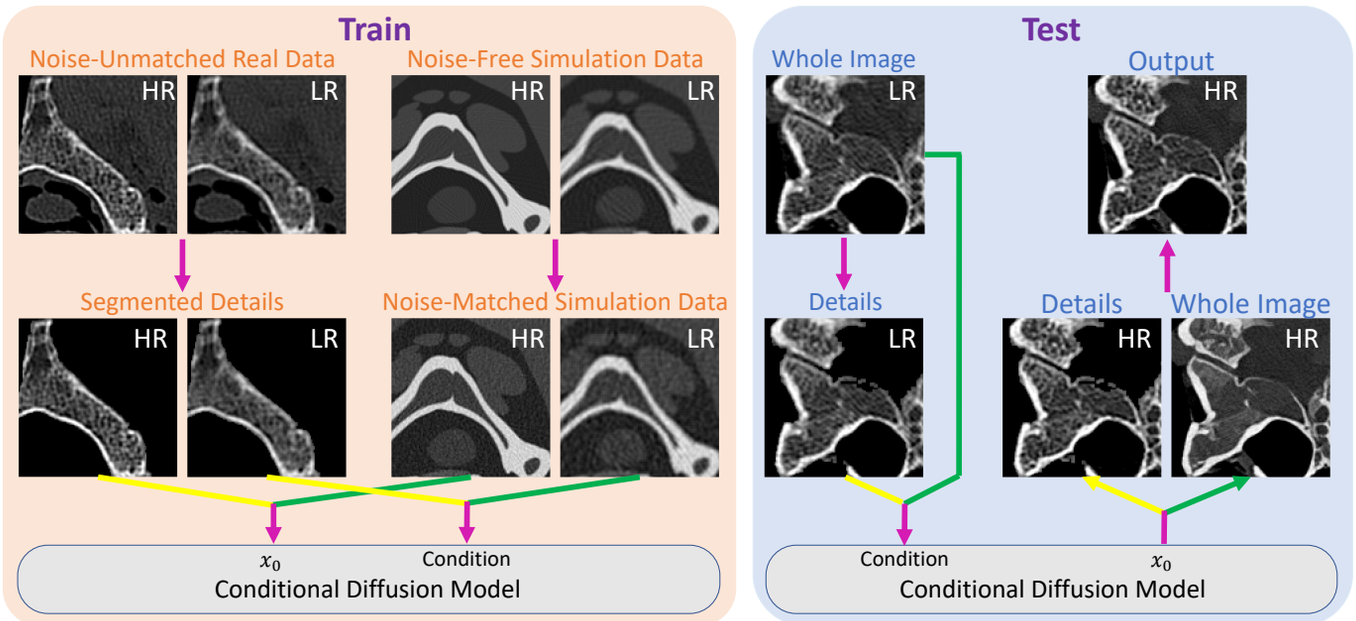


Fig. 1. Framework of Noise Controlled CT Super-Resolution

where σ_t^2 is a hyperparameter and is usually set as $\widetilde{\beta}_t$, and μ_θ is trained to match $\widetilde{\mu}_t$ to minimize the Kullback–Leibler (K-L) divergence between $p_\theta\left(x_{t-1}|x_t, y\right)$ and $q\left(x_{t-1}|x_t, x_0\right)$:

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, E_q \left\{ \frac{1}{2\sigma_t^2} \|\mu_\theta(x_t, y, t) - \widetilde{\mu}_t(x_t, x_0)\| \right\}. \quad (9)$$

To simplify the training, we can parameterize $\mu_\theta(x_t, y, t)$ as:

$$\mu_\theta(x_t, y, t) = \widetilde{\mu}_t(x_t, \widehat{x}_{0,\theta}(x_t, y, t)), \quad (10)$$

where

$$\widehat{x}_{0,\theta}\left(x_t, y, t\right) = \frac{1}{\sqrt{\gamma_t}} \left(x_t - \sqrt{1-\gamma_t}\epsilon_\theta\left(x_t, y, t\right) \right). \quad (11)$$

It leads to our training loss function

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, E_{x_0,y} E_{t,\epsilon} \|\epsilon_\theta\left(x_t, y, t\right) - \epsilon\|_2^2, \quad (12)$$

where

$$x_t = \sqrt{\gamma_t} x_0 + \sqrt{1-\gamma_t}\epsilon, \epsilon \sim N(0, I). \quad (13)$$

After the network ϵ_θ is trained, one can predict x_0 from $x_T \sim N(0, I)$ and the condition y following (8).

B. Framework of Noise-Controlled CT Super Resolution

The framework for noise-controlled CT super-resolution is illustrated in Fig. 1. The HR images are the labels x_0, and the corresponding LR images are the conditions y. Instead of directly training the conditional diffusion model using noise-unmatched real data, which is prone to amplifying noise while enhancing spatial resolution, our approach integrates training with hybrid datasets. These datasets comprise both noise-matched simulation data and segmented bones from real data.

Numerical phantoms are utilized to simulate CT scans with the same geometry parameters as real CT images, yielding noise-free projections. Assuming that the quantum noise is Gaussian, the following equations are used to inject correlated noises into the HR and LR images:

$$\widehat{p}_{HR} = p_{HR} + \frac{k_{HR}}{\sqrt{N_{HR}}} z \quad (14)$$

$$\widehat{p}_{LR} = p_{LR} + \frac{k_{LR}}{\sqrt{N_{LR}}} downsample\left(z\right), \quad (15)$$

where $z \sim N(0, I)$. $\widehat{p}$ and p are the noisy and noiseless post-log projections, respectively. N is the number of photons per ray that reached the detector, which is different for each detector. k_{HR} and k_{LR} are hyperparameters to match the noise levels of both simulated HR and LR images to that of the real LR image.

Because of the lack of details such as trabecular bone structures in the simulation phantoms, models trained with the simulation data alone would lead to significant oversmooth of the bony structures when applied to the real data. The bone regions from the real data are segmented using threshold followed by hole filling algorithms and opening and closing morphological operations, and serve as part of the training data, as demonstrated in Fig. 1. During testing, the trained model is used on both the original LR image and segmented bones. The pixels within the bone mask on the non-segmented super-resolved image are replaced by those in the bone-only super-resolution image.

In summary, the noise-matched simulated images provide the model the capability of super-resolution without amplification of the noise. Meanwhile, the segmented HR and LR bone pairs encourage the preservation of the detailed bony structures in reality without promoting noise amplification too much. By adopting both training data, the model is trained to enhance spatial resolution without noise amplification and good preservation of the bony structures.

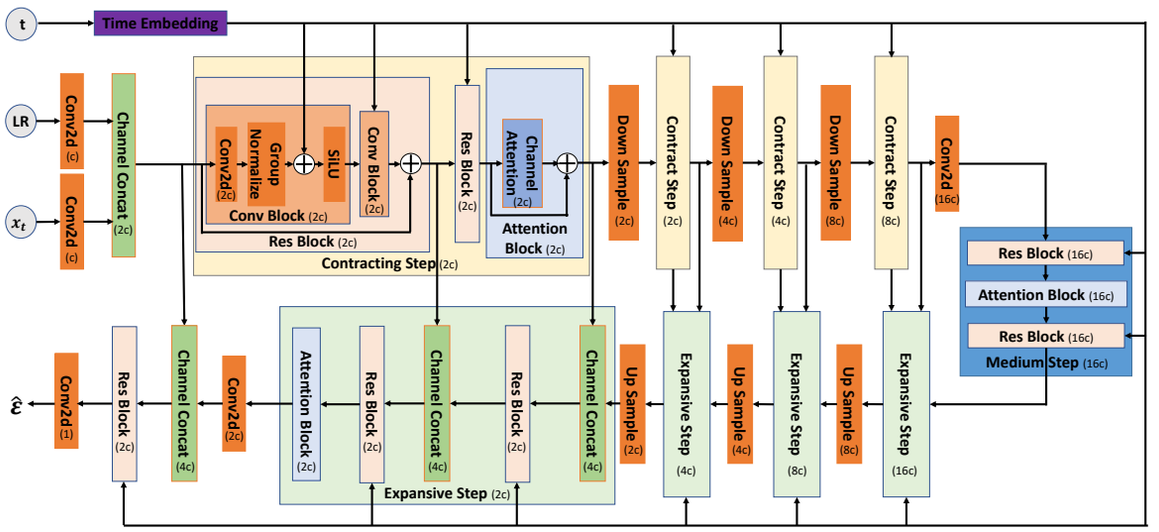


Fig. 2. Architecture of the conditional noise predictor. The content in parentheses (c, 2c, 4c, 8c and 16c) after the block name indicates the number of output channels of each block, and in our implementation c is set to 32.

C. Implementation Details

The real training and testing data were acquired using the OmniTom PCD portable photon counting CT system (Neurologica, Danvers USA) at Massachusetts General Hospital. Acquisition parameters included 120 kVp and 40 mAs for 1×1 binning data, with an effective detector size of 0.12 × 0.14 mm^2 at the isocenter. The projection data were rebinned to 3×3 and 6×5 to generate paired HR and LR real data. The real training dataset comprises 64 slices of a cadaver head, and the real testing dataset includes 32 slices of a separate temporal bone.

Simulation data were generated using the XCIST simulation tool [8], utilizing head parts from five whole-body NURBS phantoms it provides, resulting in a total of 1040 slices. The hyperparameters k_{HR} and k_{LR} in equations (14) and (15) are set to 0.12 and 0.60, respectively, to match the noise level of the simulation data to that of the real LR image.

Both real and simulation data were reconstructed into 0.3×0.3 mm^2 HR images and 0.6×0.6 mm^2 LR images using edge-enhancing bone filter. To align their sizes for the conditional diffusion model training, the LR images were upsampled by two times using sinc-interpolation before being fed into the network as condition y [9].

The architecture of the conditional noise predictor we employed is illustrated in Fig. 2. The core structure of the conditional noise predictor is a U-Net, similar to the one utilized in the DDPM [2]. y and x_t undergo a convolution layer separately and are then concatenated in the channel dimension. This structure is propagated through skip and residual connections across the entire U-Net, allowing the conditional noise predictor to extract information from y and x_t both separately and jointly.

We used $T = 1,000$ as the total number of diffusion steps, and β_t was selected according to the sigmoid schedule [10]. The model was trained on randomly cropped patches of 128 × 128, and tested on the full image. A batch size of 64 was used during the training, with 48 slices from the noise-matched simulation and 16 from the real segmented bone images. The network was trained by the Adam algorithm with a learning rate of 8×10^{-5} for 100,000 iterations.

III. Result

TABLE I
QUANTITATIVE ANALYSIS FOR ALL TESTED METHODS ON CASES AND ROIS IN FIG. 3.

Method	STD (HU)		Haralick feature distance		PSNR (dB)	
	ROI 1	ROI 2	ROI 3	ROI 4	Case 1	Case 2
HR	60.4	61.2	N/A	N/A	N/A	N/A
LR	45.7	38.1	1367	1148	23.88	24.30
Proposed	41.5	39.9	325	535	23.79	24.33
M1	43.3	40.3	857	976	21.46	21.62
M2	54.1	51.9	324	550	22.97	24.26

The representative results of noise-controlled super resolution from our proposed framework on LR CT images of a temporal bone are presented in Fig. 3, with quantitative analyses in Table I. The proposed method was further compared to two other methods: M1, which was trained with simulation data only; and M2, which was similar to the proposed method but without matched noise levels in the simulation.

Regions of interests (ROIs) are selected for quantitative analyses and marked in Fig. 3 using yellow ellipses and rectangles. Uniform ROIs 1 and 2 are used to calculate the standard deviation for evaluating the noise level, and ROIs 3 and 4, which contain intricate details, are used to calculate Haralick feature [11] distances to evaluate the similarity of the textures. Peak signal-to-noise ratio (PSNR) is also calculated to assess the consistency of the generated HR images.

Notably, M1 oversmoothes the trabecular bone structures due to the lack of such structures in the training data. Due to

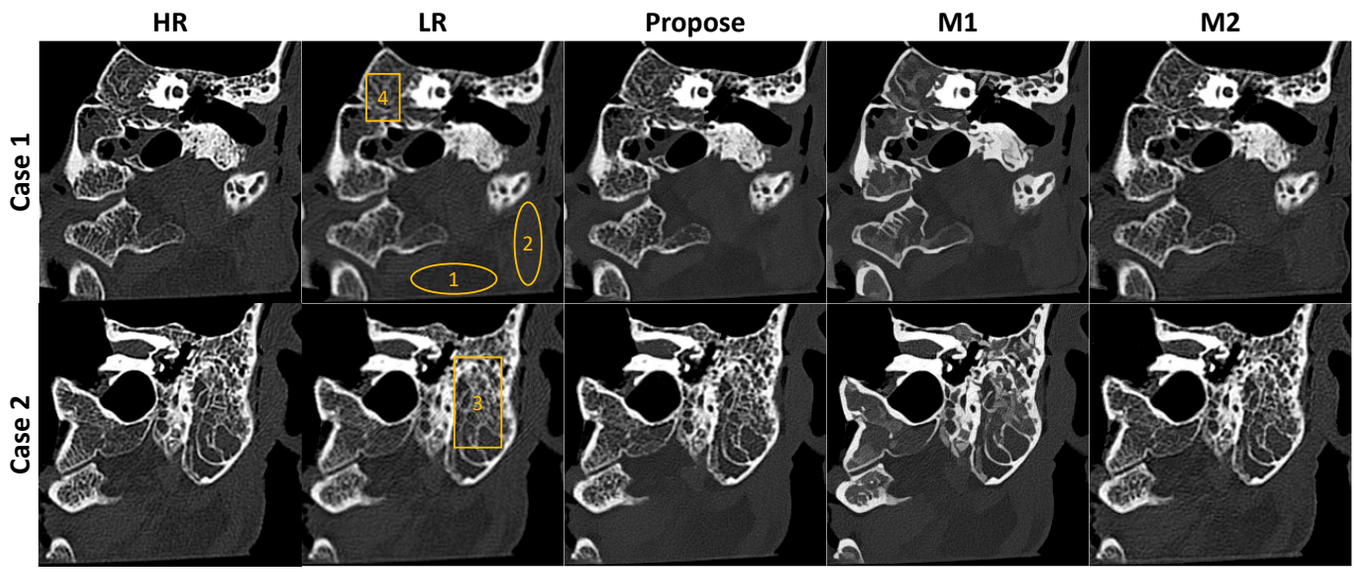


Fig. 3. Results on LR CT images of a temporal bone using our proposed framework and two comparison methods M1 and M2.

the oversmoothing, it has the worst Haralick feature distances in the bone ROIs (3 and 4) and the worst PSNR compared to the proposed method and M2. M2 has the similar enhancement of the spatial resolution and bony details compared to the proposed method, but it amplifies the noise as shown by the increased standard deviations in ROIs 1 and 2, leading to slightly worse PSNR compared to the proposed method. The proposed method achieved spatial resolution improvement without amplification of the noise. Both M2 and the proposed method had similar PSNR with the LR image, demonstrating that they are not drastically changing the structures as in M1. No significant improvement of the PSNR was observed due to the presence of noise in the HR images.

IV. Conclusion and Discussion

In conclusion, our proposed framework for noise controlled CT super resolution, leveraging the Conditional DDPM, presents a promising approach to enhance the spatial resolution of CT images while effectively controlling noise. By incorporating segmented details from real data and ensuring a matched noise level in simulation data during training, our model achieved image super-resolution without noise amplification. These results show its potential for practical applications in CT imaging, contributing to advancements in medical imaging technology.

One major limitation of the current method is that the proposed model depends on the precision of details segmentation, and errors in this process may introduce artifacts or result in the loss of details in the final output. Additionally, the training using segmented details featuring a completely black background may pose challenges to the performance of the conditional diffusion model. In our future work, we aim to investigate more robust and effective approaches to seamlessly integrate details unique to real data with noise-matched simulation data, thereby further enhancing the efficacy of noise-controlled CT super-resolution.

References

[1] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," *Advances in neural information processing systems*, vol. 32, 2019.

[2] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[3] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, "Ilvr: Conditioning method for denoising diffusion probabilistic models," *arXiv preprint arXiv:2108.02938*, 2021.

[4] 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*, vol. 45, no. 4, pp. 4713–4726, 2022.

[5] H. Li, Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, vol. 479, pp. 47–59, 2022.

[6] X. Zhang, C. Feng, A. Wang, L. Yang, and Y. Hao, "Ct super-resolution using multiple dense residual block based gan," *Signal, Image and Video Processing*, vol. 15, pp. 725–733, 2021.

[7] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong, M. W. Vannier, P. K. Saha, E. A. Hoffman, and G. Wang, "Ct super-resolution gan constrained by the identical, residual, and cycle learning ensemble (gan-circle)," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 188–203, 2020.

[8] 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.

[9] 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.

[10] A. Jabri, D. Fleet, and T. Chen, "Scalable adaptive computation for iterative generation," *arXiv preprint arXiv:2212.11972*, 2022.

[11] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

Tube Voltage Modulation for Reducing Noise in Photon Counting CT

Yirong Yang, Sen Wang, Grant M. Stevens, Adam S. Wang

***Abstract*—Photon counting detectors (PCDs) with energy discrimination capabilities enable quantitative material decomposition. The spectral image quality of different anatomical structures at a fixed dose depends on the source spectrum. For inhomogeneous object scans, such as thorax and pelvis, the thicknesses and material composition of the object varies greatly across views. A scan with a fixed tube spectrum leads to suboptimal image quality. To address this problem, tube *current* modulation has been introduced for decades, which allocates more dose to views with higher object thickness, albeit at the same tube *voltage*. In this work, we investigate another strategy that modulates the tube *voltage* angularly according to the object structure, and we evaluate its performance in reducing the image noise for multiple tasks at a fixed dose.**

We first evaluated the effect of incident tube voltages on noise performance in two material decomposition (MD) tasks by estimating the Cramér–Rao lower bound (CRLB) in the projection domain with a realistic silicon detector energy response under 80, 100, 120, and 140 kVp spectra at the same dose. We then investigated the performance of tube voltage modulation (TVM) by simulating a uniform ellipsoid phantom, where the tube voltage of each view angle was selected according to different optimization tasks. We finally compared the performance of TVM vs single-kVp scans for a more comprehensive task consisting of both MD and virtual monoenergetic imaging (VMI) on a simulated shoulder phantom.

The results show that for single tasks, TVM outperforms all single-kVp scans, with the image variance decreasing at least 11.29% for the calcium image, 5.39% for the water image, and 1.65% for VMI at 60 keV. For a combined task that optimizes both MD and VMI performance, TVM outperforms 80, 100, and 120 kVp scans with at least 30% lower image noise and is comparable with the 140 kVp scan in MD tasks, while outperforming the 140 kVp scan with 25.18% lower image noise for VMI at 55 keV. The results suggest that TVM can greatly improve the spectral image quality of inhomogeneous anatomical structures at a fixed dose compared with single-kVp scans.

***Index Term*—photon counting detector, inhomogeneous anatomy, tube voltage modulation, Cramér–Rao lower bound**

I. Introduction

ADDITIONAL spectral information provided by energy discriminating photon counting detectors (PCDs) improves diagnosis by providing MD images and VMIs in a single scan with high spatial resolution compared to conventional energy integrating detectors (EIDs). However, PCDs suffer from nonideal energy response such as charge sharing, Compton scattering, and pulse pileup, as well as finite energy resolution. Both degrade the spectral separation capability of PCDs and lead to less satisfactory spectral imaging performance. Previous work has shown that due to the nonideal properties of PCDs, single-source photon counting CT (PCCT) did not outperform dual-source EID CT in spectral tasks [1]. A later study also illustrated the dependence of MD noise on tube voltage selection and the benefit of applying fast kV switching to PCCT scans in reducing the noise in MD tasks [2]. While in most clinical CT scans the tube voltage is constant during the scan, angularly varied tube voltage is not a new concept. For example, fast kV switching has been used to provide extra spectral information in dual-energy CT (DECT). Recently, slow sinusoidal tube voltage modulation has been proposed to enable spectral imaging with EIDs [3]. In the field of cone-beam C-arm CTs with flat-panel detectors, due to the large variation of the desired irradiation and the limited achievable tube current, tube voltage is sometimes further automatically or manually adjusted by the system to achieve a satisfactory dose allocation and image quality [4]–[6]. However, in these prior works, the tube voltage has not been optimally modulated for each view, and the benefit of TVM in improving the spectral image quality for PCDs has not been discussed.

In this work, we investigated the optimal TVM strategies for different spectral imaging tasks for PCDs, such as MD and VMIs. We compared the image noise performance of TVM for a variety of tasks and different inhomogeneous phantoms with single-kVp scans at the same dose.

II. Methods

A. Experiment Settings

We used a previously published deep silicon detector model with $0.7268 \times 2.09\ \text{mm}^2$ pixels to reduce computational time, and 30 mm thickness to simulate a realistic energy response and resulting binned counts [7]. We used a typical CT geometry, with 900 views/rotation. We used four different incident spectra: 80, 100, 120, 140 kVp. The fluences were normalized with respect to CTDI to keep dose neutral and are equivalent to 120 kVp and 0.5 mAs per view. We set the energy thresholds to be 5, 18, 28, 32, 34, 46, 53, and 84 keV for the four spectra to maximize the spectral information contained in the binned counts of 120 kVp scans. These settings were used throughout our study.

This work was supported by GE HealthCare.

Yirong Yang (yryangd@stanford.edu) and Adam S. Wang are with the Department of Electrical Engineering, Stanford University.

Yirong Yang, Sen Wang, and Adam S. Wang are with the Department of Radiology, Stanford University.

Grant M. Stevens is with GE HealthCare, Waukesha, WI.

B. *Single kVp Comparison*

For single kVp comparisons, we compared the performance of MD and VMIs by evaluating a basis material space that spans 0 to 25 cm water and 0 to 1.5 cm of calcium using different tube voltages. We calculated their relative CRLB (rCRLB) to that of a 120 kVp scan at the same dose.

C. *Optimal TVM Strategy for Single Tasks*

To evaluate the performance of TVM for a specific single task, we simulated a uniform ellipsoid phantom filled with 60 mg/ml calcium in water with a 12.5 cm semi-major axis and a 2 cm semi-minor axis. The projections of this phantom intentionally span the positive diagonal line of the material space described in Sect. B.

We optimized the TVM strategies for the following 3 tasks: minimizing the average noise of the calcium image, water image, and VMI at 60 keV. To find the optimal tube voltage of each view, we first calculated the average line integral measured in each view, then selected the corresponding tube voltage that gives the lowest rCRLB for the specific task.

Due to the forward model for MD, which predicts PCD measurements from the attenuation materials and varies with different tube voltages, we first applied material decomposition in *projection* domain to evaluate the performance of TVM. Projection domain MD enables acquisition and processing of each view independently with arbitrary spectra if the forward model has been carefully calibrated for each spectrum. With estimated calcium and water thicknesses, corresponding basis material images and VMIs were reconstructed, and image noises were measured in corresponding ROIs for comparison.

D. *Optimal TVM Strategy for A Combined Task*

We finally evaluated the performance of TVM for a combined task that minimizes noise for both MD and VMI at 55 keV for a complex shoulder phantom.

The shoulder phantom comprises an ellipsoid with 20 cm semi-major axis and 7 cm semi-minor axis filled with water. Several locations are replaced with 60% cortical bone equivalent materials to mimic the bony structures of arms, clavicles, and the spine. A 2 mg/ml iodine insert was also simulated to represent a contrast enhanced lesion. The phantom design is presented in Figure 1.

To find the optimal tube voltage of each view for the combined task, the average projection domain rCRLBs for each task using different tube voltages were calculated for each view, and the tube voltage that achieves the minimum total rCRLB was selected. To evaluate the performance of TVM, *projection* domain MD was applied, and MD and VM images were reconstructed. Image noises were measured within the corresponding ROIs.

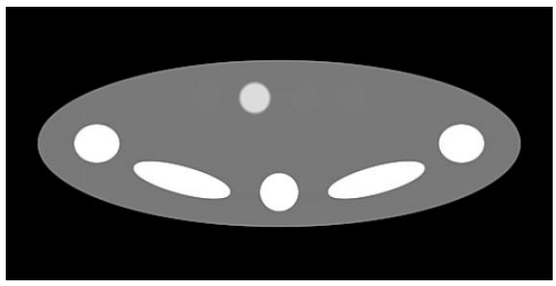


Figure 1. Simulated shoulder phantom with 60% cortical bone equivalent inserts and 2 mg/ml iodine insert.

III. Results

A. *rCRLB of Single-kVp Tasks*

We first show our analytical results of rCRLB (with respect to 120 kVp at the same dose) for MD and VMI tasks at different energies. The contour plots of rCRLB of different tasks are shown in Figure 2. Each sub-region has been colored according to the optimal tube voltage that yields the best image quality (lowest CRLB) among the four spectra. We observe that the optimal tube voltage varies with varied tasks and material thicknesses. This result indicates the significance of tube voltage modulation for scans of objects with strong angular variation in structures.

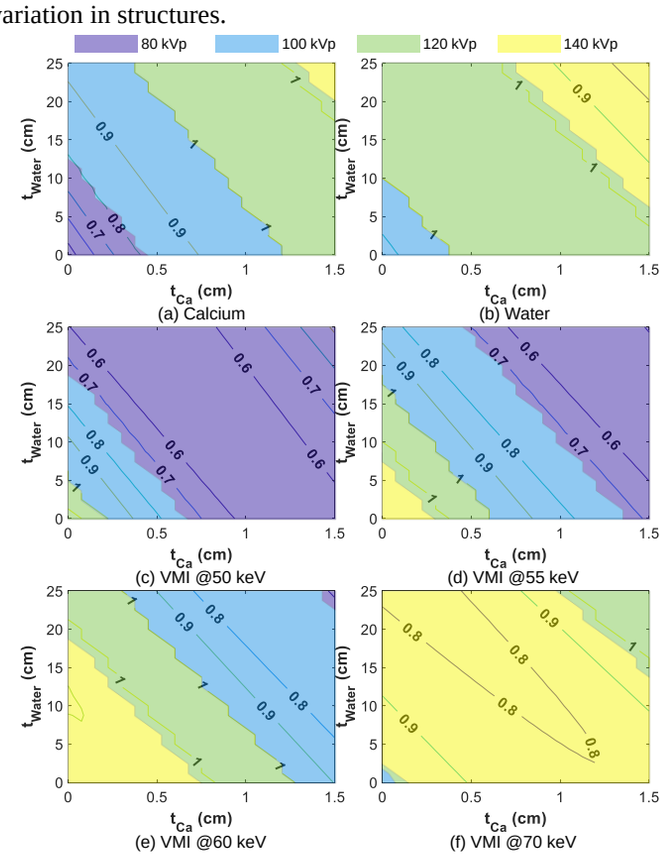


Figure 2. rCRLB w.r.t. 120 kVp at the same dose for (a) calcium thickness estimation, (b) water thickness estimation, VM estimation at (c) 50 keV, (d) 55 keV, (e) 60 keV, (f) 70 keV. The optimal tube voltage is shown in each region.

B. *TVM for Single Task*

1) *TVM for Calcium Thickness Estimation*

The TVM curve and the measured image domain variance relative to that of TVM are shown in Figure 3(a-b). Compared with TVM, the calcium image variance increased by 138.04% for 80 kVp, 42.64% for 100 kVp, 11.29% for 120 kVp, and 17.13% for 140 kVp. We also observe a strong correlation between average line integral and optimal tube voltage. The calcium images are shown in the first column of Figure 4.

2) *TVM for Water Thickness Estimation*

The TVM curve and the measured image domain variance relative to that from TVM are shown in Figure 3(c-d). Compared with TVM, the water image variance increased by 361.37%, 99.63%, 24.13%, and 5.39% for 80, 100, 120, and 140 kVp scans. The water images are shown in the second column of Figure 4. We observe the same trend between average line integral and corresponding optimal tube voltage, but the relationship between them is different from that of the calcium task.

3) *TVM for VMI at 60 keV Estimation*

The results are shown in Figure 3(e-f). Compared with TVM, the image variance of VMI at 60 keV increased by 9.04%, 1.65%, 12.17%, and 60.70% for 80, 100, 120, and 140 kVp scans. The VMIs are shown in the third column of Figure 4.

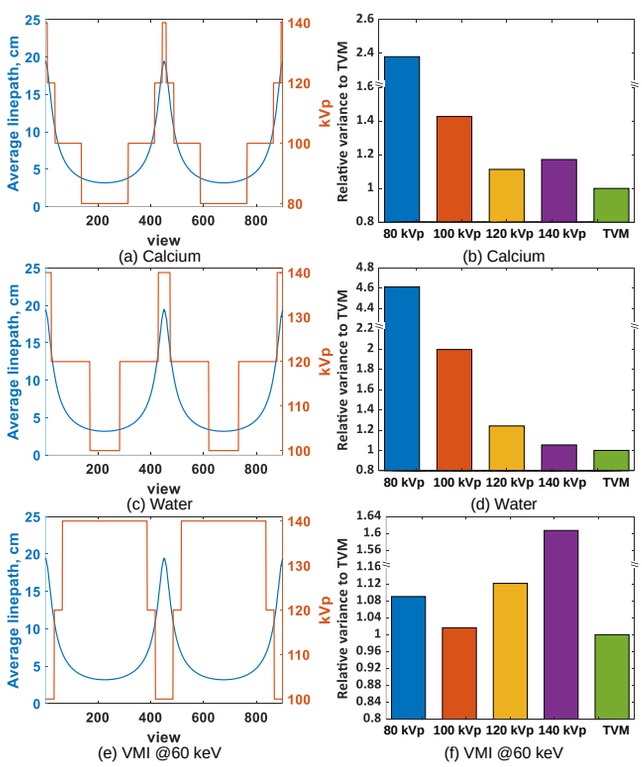


Figure 3. Average line integral of each view (blue) and corresponding optimal tube voltage (red) for (a) calcium thickness estimation, (c) water thickness estimation, (e) 60 keV VMI estimation; relative noise variance in (b) calcium image, (d) water image, and (f) VMI of different tube voltages to TVM.

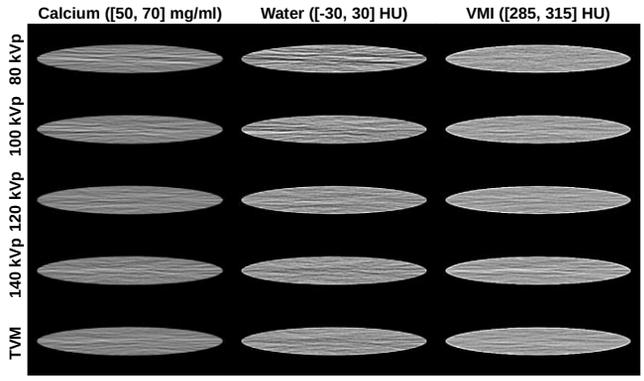


Figure 4. Reconstructed images of uniform ellipsoid phantom of different tasks: calcium image, water image, VMI at 60 keV.

C. *TVM for Combined Task*

The TVM curve is shown in Figure 5(a). The calcium image variance is measured within uniform regions of the bone inserts. The variance relative to TVM is calculated and presented in Figure 5(b). The relative variances of the water images are measured within the uniform water background and are presented in Figure 5(c). The relative VMI variances are measured within the 2 mg/ml iodine insert and are presented in Figure 5(d). The relative variances to TVM are summarized in Table 1.

We observe that with TVM, the MD noise of calcium and water are significantly decreased compared to single-kVp scans of 80, 100, and 120 kVp, and are comparable to the performance of 140 kVp. The variance of VMI at 55 keV with TVM is further decreased by 25.18% compared with the 140 kVp scan. Although 80 and 100 kVp scans outperform TVM in the VMI task, comparing the average performance of MD and VMI tasks demonstrates that TVM yields better image quality on average.

The calcium images, water images, and VMIs at 55 keV obtained using single tube voltages and TVM are shown in Figure 6. We observe severe streak artifacts in the MD images of 80 and 100 kVp scans due to bias from zero counts. These artifacts largely cancel each other in the VMI due to the negatively correlated noise between the two materials.

IV. Discussion

We have compared the noise performance of different tube voltages on a variety of spectral imaging tasks: MD and VMI at different energies. The rCRLB plot showed that the optimal tube voltages used for generating the incident spectra vary with different imaging tasks and different basis material thicknesses. This observation suggests TVM can benefit objects with large thickness variations across views.

The simulation study of the uniform ellipsoid phantom tested and validated our hypothesis that for an angularly inhomogeneous object, TVM will improve the spectral performance. We notice that although TVM provides substantial improvements of basis material image quality, the noise reduction of VMI at 60 keV is only 1.65% decrease compared to 120 kVp scan. This observation suggests that TVM fails to substantially improve the image quality for some tasks. We also observe that the TVM curves are different for the

same phantom when the tasks varied. This observation is consistent with the results in Figure 2 and suggests a task-specific TVM plan to achieve the optimal image quality.

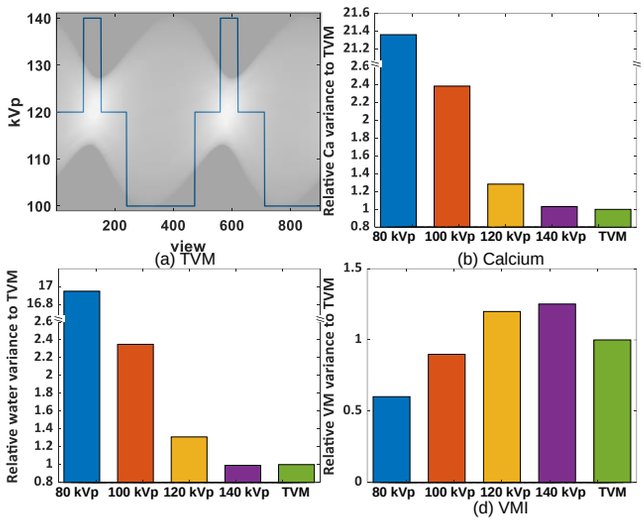


Figure 5. (a) Optimal tube voltage for each view of shoulder phantom, overlaid on sinogram. Relative variance to TVM of (b) calcium map, (c) water map, (d) VMI at 55 keV.

Table 1. Relative variance to TVM of calcium image, water image, and VMI at 55 keV in corresponding ROIs.

	Relative variance to TVM				
Task	80 kVp	100 kVp	120 kVp	140 kVp	TVM
Calcium	21.3619	2.3845	1.2835	1.0311	1.0000
Water	16.9468	2.3505	1.3108	0.9917	1.0000
VMI@ 55keV	0.6016	0.8988	1.1993	1.2518	1.0000

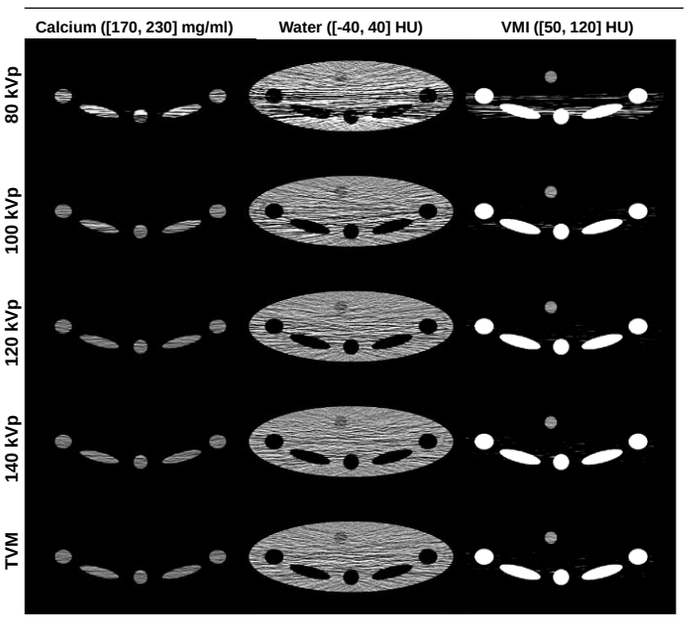


Figure 6. Calcium image, water images, and VMIs at 55 keV for single-kVp scans vs TVM.

The combined task results further present the benefits of TVM for spectral imaging tasks of inhomogeneous objects. Although TVM did not outperform all single-kVp scans for all tasks, TVM gave the best image quality if we consider the average performance of both MD and VMI.

In this study, we did not investigate the effect of continuous TVM. We used four different tube voltages with a step size of 20 kVp and assumed instantaneous switching. In practice, the optimal tube voltage is more likely to change smoothly. Considering the low frequency nature of overall human anatomy, a smoothly varying tube voltage could further improve the image quality, although it could complicate spectral calibrations.

We also did not consider other techniques, such as tube current modulation (TCM), in this work. However, to study the effect of TVM independently, we scaled the spectra of different tube voltages to keep dose equal for each view, which means the tube currents were scaled. We expect that applying both tube current and tube voltage modulation will further improve the spectral image quality of inhomogeneous objects. In clinical scans, TVM and TCM could be planned at the same time using 2D or 3D scout images.

In future work, we will conduct more experiments on our prototype silicon PCCT to investigate the performance of TVM, as well as the benefits of combined TVM and TCM.

V. Conclusion

We found that for PCCT scans, compared to using angularly constant tube voltage, using TVM with tube voltages of each view optimized for the specific spectral imaging tasks, we can substantially improve image quality of MD and VMIs simultaneously for objects with angularly inhomogeneous anatomical structures, such as the shoulder, thorax, and pelvis.

References

[1] A. Tao, R. Huang, S. Tao, G. J. Michalak, C. H. McCollough, and S. Leng, “Dual-source photon counting detector CT with a tin filter: a phantom study on iodine quantification performance,” *Phys. Med. Biol.*, vol. 64, no. 11, p. 115019, May 2019, doi: 10.1088/1361-6560/ab1c34.

[2] S. Wang, Y. Yang, D. Pal, N. J. Pelc, and A. S. Wang, “Fast kV switching for improved material decomposition with photon counting x-ray detectors,” in *Medical Imaging 2022: Physics of Medical Imaging*, SPIE, Apr. 2022, pp. 97–102. doi: 10.1117/12.2611601.

[3] T. Su, J. Zhu, J. Yang, D. Liang, and Y. Ge, “Multi-material decomposition in spectral CT with angular tube voltage modulation,” in *Medical Imaging 2022: Physics of Medical Imaging*, SPIE, Apr. 2022, pp. 685–690. doi: 10.1117/12.2611728.

[4] L. Ritschl, F. Bergner, C. Fleischmann, and M. Kachelrieß, “Water calibration for CT scanners with tube voltage modulation,” *Phys. Med. Biol.*, vol. 55, no. 14, p. 4107, Jul. 2010, doi: 10.1088/0031-9155/55/14/010.

[5] J. Zou, X. Hu, H. Lv, and X. Hu, “An Investigation of Calibration Phantoms for CT Scanners with Tube Voltage Modulation,” *International Journal of Biomedical Imaging*, vol. 2013, p. e563571, Dec. 2013, doi: 10.1155/2013/563571.

[6] H. Haseljić *et al.*, “The effect of x-ray tube voltage modulation to quality of perfusion images in cone beam C-arm CT,” in *Medical Imaging 2022: Physics of Medical Imaging*, SPIE, Apr. 2022, pp. 542–546. doi: 10.1117/12.2609652.

[7] 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, Jul. 2020, doi: 10.1117/1.JMI.7.4.043501.

A Dual-domain Regularization Method for Ring Artifact Removal of X-ray CT

Hongyang Zhu, Xin Lu, Yanwei Qin, Xinran Yu, Tianjiao Sun and Yunsong Zhao

***Abstract*—CT images often suffer from ring artifacts, which are mainly caused by inconsistent responses of detector units. These artifacts can substantially impair the quality of CT images and directly influence diagnostic accuracy. Consequently, it is imperative to effectively eliminate ring artifacts while preserving image quality. In this paper, we propose a dual-domain regularization model to remove ring artifacts in CT images. The proposed model addresses the correction of vertical stripe artifacts on the sinogram by updating the compensation coefficients for detector response inconsistency, which is accomplished by leveraging a group sparse prior for the vertical stripe artifact image and a sparse prior along the projection views. Simultaneously, the ring artifacts in the image domain are further rectified through the imposition of sparse constraints on the reconstructed image. A key advantage of the proposed method lies in considering the relationship between the response inconsistency compensation coefficients of the detector and the projection views, which allows for a more accurate correction of the response of detector. An alternately iterative solving algorithm is designed to solve the model. Experimental results on real data collected by a photon counting detector show that the proposed method is effective in removing ring artifacts from reconstructed CT images and preserving the structure and details of the CT images compared to competing methods.**

***Index Terms*—ring artifact removal, dual-domain regularization, compensation coefficients, group sparse.**

I. Introduction

PHOTON counting detector (PCD) CT have been one of the hot topics in X-ray imaging in recent years [1]. Though PCD have promising potential for offering significant improvements to existing CT imaging techniques, challenges remain for the wider application of PCD CT imaging. One of the challenges is inconsistent response of detector units of PCD, which leads to stripe artifacts on the sinogram and subsequent concentric ring artifacts on reconstructed CT images [2]. These ring artifacts can significantly degrade the quality of the CT images, hampering quantitative analysis and reducing diagnostic accuracy. With the wide range of application requirements, there is an urgent need to develop an efficient ring artifact removal method.

Hongyang Zhu, Xin Lu, Yanwei Qin, Xinran Yu, Tianjiao Sun and Yunsong Zhao are with School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China, and with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China.

E-mail: zhao_yunsong@cnu.edu.cn.

Hongyang Zhu and Xin Lu contribute equally to the manuscript.

This work is supported in part by the National Natural Science Foundation of China (NSFC) (62271330, 61871275, 61771324, and 61827809) and National Key R&D Program of China (2023YFA1011400).

The authors are grateful to Beijing Higher Institution Engineering Research Center of Testing and Imaging as well as Beijing Advanced Innovation Center for Imaging Technology for funding this research work.

Existing ring artifact removal methods can be roughly categorized into hardware-based methods and software-based methods. Hardware-based methods change the relative positions of the CT system and the scanned object, and spread out the effects of detector unit response inconsistencies on the reconstructed CT images [3]. These methods often increase manufacturing costs and may introduce additional artifacts in the reconstructed CT images, such as noise amplification [4]. Consequently, additional algorithms are often required to further correct these artifacts.

Compared to hardware-based methods, software-based methods are generally easier to implement. Software-based methods can be broadly categorized into flat-field correction methods, pre-processing methods, post-processing methods, and dual-domain processing methods. Commonly used methods like flat-field correction [5] usually cannot completely eliminate ring artifacts as these artifacts are influenced by various factors, such as the x-ray spectrum and the scanned object. Pre-processing methods primarily aim to eliminate ring artifacts by removing vertical stripe artifacts on the sinogram. Various stripe artifact correction methods have been proposed, including filtering-based methods [6], [7] and mean projection-based methods [8], [9]. However, the effectiveness of these methods relies on the accurate localization and compensation of vertical stripe artifacts. Incomplete correction or inaccurate compensation may lead to the loss of important information. As for post-processing methods, since ring artifacts in Cartesian coordinates appear as vertical stripe artifacts in polar coordinates which are generally easier to detect, several methods based on polar coordinates have been developed [10], [11], [12]. However, a common challenge faced by these methods is the need for coordinate transformations, which can make it difficult to completely remove the ring artifacts while simultaneously maintaining the spatial resolution of images.

Dual-domain methods are a series of methods that removing vertical stripe artifacts on the sinogram and the ring artifacts on the reconstructed CT images simultaneously. Most recently developed deep-learning based methods belong to this category [13], [14]. These methods are usually supervised methods and rely on training datasets, posing challenges in maintaining consistency between numerically simulated data and real data. Compared to pre-processing methods and post-processing methods, dual-domain methods show better ring artifacts removal ability, such as the iterative method proposed by Salehjahromi et al [15], which aims to correct ring artifacts by iteratively optimizing the clean reconstructed image and compensation coefficient vector. However, this method uses the same compensation coefficient vector for scan data at

different projection views, and may results in unsatisfactory results when dealing with vertical stripe artifacts that exhibit significant intensity variations in the projection view direction.

To address the limitations of the existing methods, we present a novel dual-domain regularization model designed for ring artifact removal. The proposed model addresses the correction of vertical stripe artifacts in the sinogram by updating the compensation coefficients for detector response inconsistency, which is accomplished by leveraging a group sparse prior for the vertical stripe artifact image and a sparse prior along the projection views. Simultaneously, the ring artifacts in the image domain are further rectified through the imposition of sparse constraints on the reconstructed CT image. A key advantage of the proposed method lies in considering the relationship between the response inconsistency compensation coefficient of the detector and the projection views, which allows for a more accurate correction of the response of detector. The effectiveness of the method is evaluated through numerical experiments conducted on real data obtained by a PCD. The experimental results demonstrate that our method is capable of effectively removing ring artifacts in the reconstructed image while preserving the structure and details of the image when compared to other methods used for comparison.

II. Materials and methods

A. Formulation for Ring Artifacts Removal

Consider the fan-beam CT scan. Let $\boldsymbol{I}_i \in R^{V\times 1}(i = 1,\cdots,D)$ be the scan data of the ith detector unit, and V is the number of projection views and D is the number of detector units, and let I_f and I_d be the flat-field data and dark-field data, respectively; then the normalized scan data $\boldsymbol{q}_i \in R^{V\times 1}$ can be calculated with formula:

$$\boldsymbol{q}_i = \frac{\boldsymbol{I}_i - I_d}{I_f - I_d} \tag{1}$$

Because of inconsistent response of the detector units, the normalized scan data composed of $\boldsymbol{q}_i$ contains strip artifacts. Let $\widetilde{\boldsymbol{q}}_i$ represent the ideal normalized scan data, and $\boldsymbol{s}_i \in R^{V\times 1}$ be the response inconsistency compensation coefficients for the ith detector unit, we have:

$$\widetilde{\boldsymbol{q}}_i = \boldsymbol{s}_i \odot \boldsymbol{q}_i \tag{2}$$

Taking the negative logarithm on both sides of (2), we can obtain the clean sinogram $\boldsymbol{p}_i \in R^{V\times 1}$:

$$\boldsymbol{p}_i \triangleq -\log \widetilde{\boldsymbol{q}}_i = -\log \boldsymbol{s}_i - \log \boldsymbol{q}_i \triangleq \boldsymbol{S}_i + \boldsymbol{y}_i \tag{3}$$

where $\boldsymbol{y}_i \in R^{V\times 1}$ is the real sinogram and $\boldsymbol{S}_i \in R^{V\times 1}$ is the response inconsistency compensation coefficients in projection domain. It is obvious that if we can accurately estimate the compensation coefficients $\boldsymbol{s}_i$ or $\boldsymbol{S}_i$ for every detector unit, we can obtain the clean sinogram with simple mathematical operations.

The corrected reconstruction image $\boldsymbol{x} \in R^{N\times 1}$ (N denote the number of image pixels) as well as the estimation $\boldsymbol{S}_i$ can be obtained by solving the following optimisation model:

$$\underset{\boldsymbol{x},\boldsymbol{S}_i}{\operatorname{argmin}} \left\{ \begin{array}{c} \frac{1}{2}\sum_{i=1}^{D} \|A_i\boldsymbol{x} - \boldsymbol{p}_i + \boldsymbol{S}_i\| + \lambda_1 R(\boldsymbol{x}) \\ +\lambda_2 \sum_{i=1}^{D} R(\boldsymbol{S}_i) \end{array} \right\} \tag{4}$$

where, $A_i \in R^{V\times N}$ represents the system matrix corresponding to the ith detector unit; λ_1 and λ_2 are regularization parameters; The design of $R(\boldsymbol{x})$ and $R(\boldsymbol{S}_i)$ plays a crucial role in the ring artifact removal effect.

B. Related methods

In practice, various factors would affect the correction coefficients $\boldsymbol{s}_i$ or $\boldsymbol{S}_i$, such as scanning voltage, current, filter, and the scanned object. So the correction coefficients $\boldsymbol{s}_i$ or $\boldsymbol{S}_i$ are generally projection view dependent. But many ring artifact removal methods simplify the model by assuming that all elements of $\boldsymbol{s}_i$ or $\boldsymbol{S}_i$ are the same, such as [9], [15].

[9] proposes a mean projection-based method to estimate the correction coefficients $\boldsymbol{s}_i$. Suppose that $\boldsymbol{m} \in R^{D\times 1}$ is the real mean projection consisting of the mean value m_i of scanned data $\boldsymbol{q}_i$, and $\widetilde{\boldsymbol{m}} \in R^{D\times 1}$ is the ideal mean projection, which can be approximately estimated by operating TV-L^1 filtering and Gaussian filtering on $\boldsymbol{m}$. Therefore, the compensation coefficients for all projection views at ith detector unit are calculated by the following equation:

$$[\boldsymbol{s}_i]_j = \frac{\widetilde{m}_i}{m_i}, j = 1,\cdots,V$$

Here $[\boldsymbol{s}_i]_j$ denotes the compensation coefficient for the ith detector unit and jth projection view.

[15] proposes a dual-domain regularisation model to estimate the compensation coefficients $\boldsymbol{S}_i$ in the projection domain. In this model, a novel ring total variation regularization $R(\boldsymbol{x}) = RTV_{\alpha,\theta}(\boldsymbol{x})$ is developed to correct the ring artifacts in the image domain. Moreover, to correct the sinogram, a projection domain regularization $R(\boldsymbol{S}_i) = \sum_{i=1}^{D} \|\boldsymbol{S}_i\|$ is proposed to estimate the compensation coefficients of each detector unit. Here, all elements of $\boldsymbol{S}_i$ are the same i.e. $\boldsymbol{S}_i = S_i\mathbf{1}$, and $\mathbf{1} \in R^{V\times 1}$ is a vector whose elements are also "1".

However, the basic assumption of these methods leads to the fact that those vertical stripe artifacts that show significant differences in different projection views cannot be completely removed.

C. The proposed optimization model

In this paper, we aim to address the limitations of existing methods and propose a novel dual-domain regularized ring artifact removal model. In the proposed model, we incorporate the correlation between the response inconsistency of the detector units with the projection views within the optimization

framework. The model in this paper can be summarized as follows:

$$\underset{\boldsymbol{x},\boldsymbol{S}_i}{\operatorname{argmin}} \left\{ \begin{array}{c} \frac{1}{2}\sum_{i=1}^{D} \|A_i\boldsymbol{x} - \boldsymbol{p}_i + \boldsymbol{S}_i\| \\ +\lambda_1 \left(\|D_H\boldsymbol{x}\|_1 + \|D_V\boldsymbol{x}\|_1\right) \\ +\lambda_2 \sum_{i=1}^{D} \|D_V\boldsymbol{S}_i\|_1 + \lambda_3 \sum_{i=1}^{D} \|\boldsymbol{S}_i\|_2 \end{array} \right\} \quad (5)$$

where λ_1, λ_2 and λ_3 are regularisation parameters. $D_H \in R^{N\times N}$ and $D_V \in R^{N\times N}$ are the difference matrices corresponding to the horizontal and vertical directions, respectively. Noting that the projection domain regularization $\sum_{i=1}^{D} \|D_V\boldsymbol{S}_i\|_1$ is aimed to provide a sparse solution for $\boldsymbol{S}_i$ along the projection views, which considers the correlation between correction coefficients and the projection views, and $\sum_{i=1}^{D} \|\boldsymbol{S}_i\|_2$ is aimed to provide a column sparse solution for $\boldsymbol{S} = [\boldsymbol{S}_1, \cdots, \boldsymbol{S}_i, \cdots, \boldsymbol{S}_D] \in R^{V\times D}$. The image domain regularization $\|D_H\boldsymbol{x}\|_1 + \|D_V\boldsymbol{x}\|_1$ is considered to further rectify the ring artifacts on the reconstructed CT image.

D. Algorithm for solving the model

The model (5) is non-convex, so that we can only expect to get local minimizer. In our practice, we split the model into two sub-problems which are solved in alternating iterations:

Sub-problem 1 :

$$\boldsymbol{x}^{(k+1)} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ \begin{array}{c} \frac{1}{2}\sum_{i=1}^{D} \left\|A_i\boldsymbol{x} - \boldsymbol{p}_i + \boldsymbol{S}_i^{(k)}\right\|_2^2 \\ +\lambda_1 \left(\|D_H\boldsymbol{x}\|_1 + \|D_V\boldsymbol{x}\|_1\right) \end{array} \right\} \quad (6)$$

Sub-problem 2:

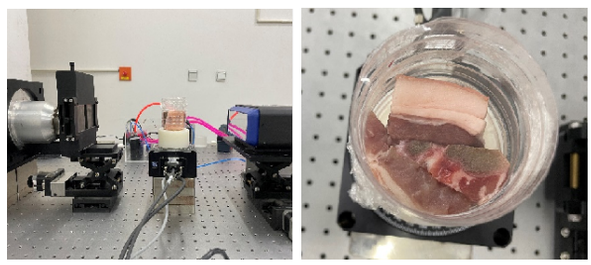
$$\boldsymbol{S}_i^{(k+1)} = \underset{\boldsymbol{S}_i}{\operatorname{argmin}} \left\{ \begin{array}{c} \frac{1}{2}\sum_{i=1}^{D} \left\|A_i\boldsymbol{x}^{(k+1)} - \boldsymbol{p}_i + \boldsymbol{S}_i\right\|_2^2 \\ +\lambda_2 \sum_{i=1}^{D} \|D_V\boldsymbol{S}_i\|_1 + \lambda_3 \sum_{i=1}^{D} \|\boldsymbol{S}_i\| \end{array} \right\} \quad (7)$$

A similar alternating minimization method is used to solve sub-problem 1, and the alternating direction method of multipliers (ADMM) is used to solve sub-problem 2.

E. Experiments

To assess the effectiveness of the proposed method, we perform experiments using real data. The real data is acquired with an experimental platform developed by our laboratory (Fig. 1(a)), which is equipped with a HAMAMATSU-L10101 x-ray source and an EIGER2 1M-W R-DECTRIS PCD detector. The scanned object is a small piece of pork with bones, and its photograph is shown in Fig. 1(b). The scanning parameters are listed in Table I. For the experiment, only the projection data corresponding to the central slice was extracted to simulate a fan-beam CT scan.

To demonstrate the benefits of the proposed method, we compared it with four state-of-the-art methods: the filtering-based method proposed by Münch et al [6] and Podgorsak et al [8], the mean projection-based method proposed by Wu et al [9], and the iterative optimization method proposed by Salehjahromi et al [15]. For convenience, we will refer to them as the WF method, Podgorsak method, Wu method, and Salehjahromi method, respectively. we carefully adjusted their parameters to ensure the optimal performance of these comparative methods.



(a) (b)

Figure 1. The experimental CT platform (a) and the photograph of scanned object (b).

Table I
GEOMETRICAL PARAMETERS OF THE REAL SCANNING DEVICE.

Parameter	Value
Voltage	100 kV
Current	0.3 mA
Number of views	720
Energy bin thresholds	50, 60 keV
Width of detector unit	0.075 mm
Number of detector units	2068×512
Size of reconstruction image	2068 × 2068
Distance of X-ray source to detector	433.507 mm
Distance of X-ray source to rotation center	205 mm

III. RESULTS

Fig. 2 shows the reconstruction results, including the uncorrected image for comparison (Fig. 2(a)), the corrected result of the WF method (Fig. 2(b)), the Podgorsak method (Fig. 2(c)), the Wu method (Fig. 2(d)), the Salehjahromi method (Fig. 2(e)), and the proposed method (Fig. 2(f)). Magnified images of regions of interest (ROIs) are shown at the bottom of each image. The locations of the ROIs are labeled with blue and red rectangles in Fig. 2(a), respectively.

As can be seen from Fig. 2(a), there are severe ring artifacts on the reconstructed CT image, which seriously interfere with the interpretation of the image. The WF method effectively removes most of the ring artifacts, but it introduces new low-frequency artifacts (as shown by the solid yellow arrow). Both the Eldib method and the Wu method can correct most of the ring artifacts and preserve image details, but they encounter challenges in correcting wider ring artifacts. The Salehjahromi method successfully corrects the wider ring artifacts, but some ring artifacts persist. Moreover, these methods struggle with ring artifacts that pass through different substances simultaneously, leading to even more pronounced ring artifacts at these locations (as shown by the dotted yellow arrow). This phenomenon mainly arises from their failure to consider the correlation between the response inconsistency of detector units and the scanned object. In contrast, the proposed method

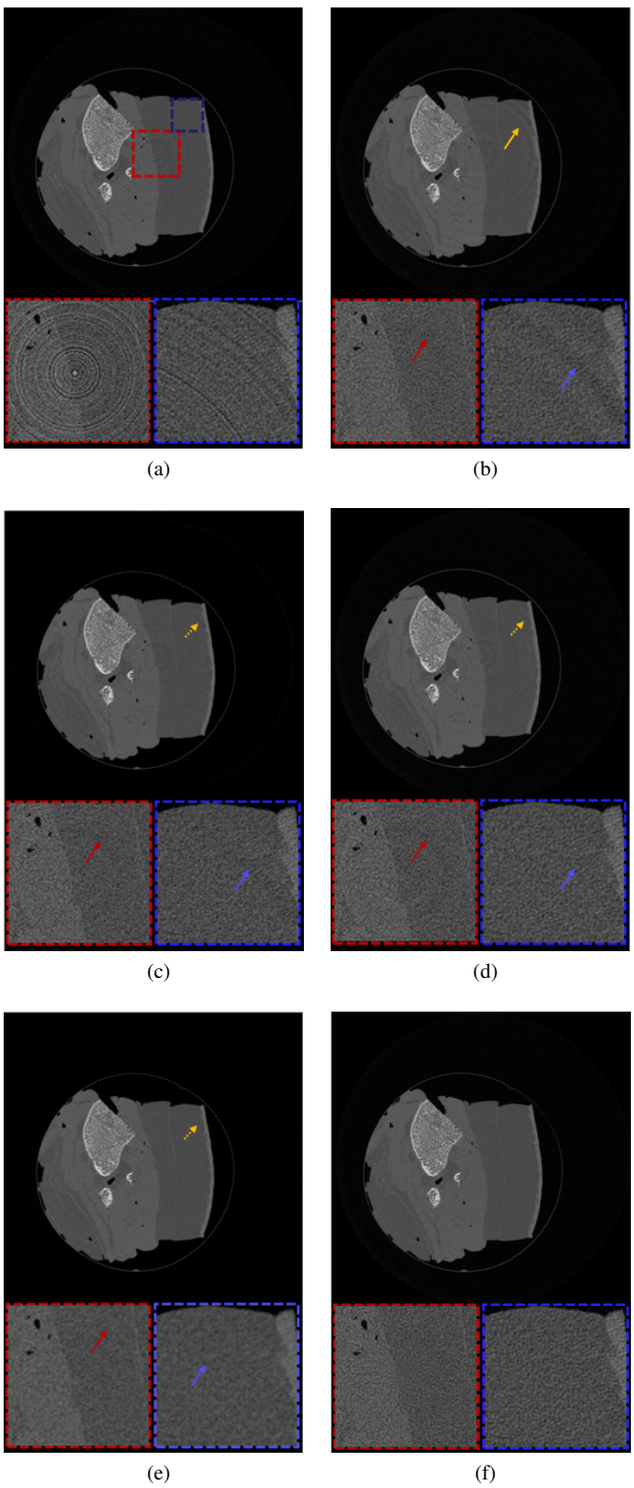


Figure 2. The reconstruction results. (a) Uncorrected image; (b) WF method; (c) Podgorsak method; (d) Wu method; (e) Salehjahromi method; (f) Our method. The display window is set to [0, 0.08].

can remove almost all ring artifacts, particularly excelling in the correction of wider ring artifacts and those traversing different substances.

IV. Conclusions

A novel dual-domain regularization model is proposed for ring artifact removal, which corrects stripe artifacts in the projection domain by updating the compensation coefficients of response inconsistency of detector units. Simultaneously, it further rectified the ring artifacts by the imposition of sparse constraints on the reconstructed CT image. The effectiveness of the proposed method is validated through experiments conducted on real data. The results demonstrate that the proposed method can remove both high frequency and low frequency ring artifacts effectively, and outperforms the competition methods in terms of ring artifact removal.

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.

[2] 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.

[3] Y. Zhu, M. Zhao, H. Li, and P. Zhang, "Micro-ct artifacts reduction based on detector random shifting and fast data inpainting," *Medical physics*, vol. 40, no. 3, p. 031114, 2013.

[4] Y. Liu, C. Wei, and Q. Xu, "Detector shifting and deep learning based ring artifact correction method for low-dose ct," *Medical Physics*, 2023.

[5] V. Van Nieuwenhove, J. De Beenhouwer, F. De Carlo, L. Mancini, F. Marone, and J. Sijbers, "Dynamic intensity normalization using eigen flat fields in x-ray imaging," *Optics express*, vol. 23, no. 21, pp. 27 975–27 989, 2015.

[6] B. Münch, P. Trtik, F. Marone, and M. Stampanoni, "Stripe and ring artifact removal with combined waveletfourier filtering," *Optics express*, vol. 17, no. 10, pp. 8567–8591, 2009.

[7] C. Raven, "Numerical removal of ring artifacts in microtomography," *Review of scientific instruments*, vol. 69, no. 8, pp. 2978–2980, 1998.

[8] A. R. Podgorsak, S. S. Nagesh, D. Bednarek, S. Rudin, and C. N. Ionita, "Use of a cmos-based micro-ct system to validate a ring artifact correction algorithm on low-dose image data," in *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573. SPIE, 2018, pp. 903–914.

[9] K. An, J. Wang, R. Zhou, F. Liu, and W. Wu, "Ring-artifacts removal for photon-counting ct," *Optics Express*, vol. 28, no. 17, pp. 25 180–25 193, 2020.

[10] X. Liang, Z. Zhang, T. Niu, S. Yu, S. Wu, Z. Li, H. Zhang, and Y. Xie, "Iterative image-domain ring artifact removal in cone-beam ct," *Physics in Medicine & Biology*, vol. 62, no. 13, p. 5276, 2017.

[11] Z. Wang, J. Li, and M. Enoh, "Removing ring artifacts in cbct images via generative adversarial networks with unidirectional relative total variation loss," *Neural computing and applications*, vol. 31, pp. 5147–5158, 2019.

[12] L. Yuan, Q. Xu, B. Liu, Z. Wang, S. Liu, C. Wei, and L. Wei, "A deep learning-based ring artifact correction method for x-ray ct," *Radiation Detection Technology and Methods*, pp. 1–11, 2021.

[13] W. Fang, L. Li, and Z. Chen, "Removing ring artefacts for photon-counting detectors using neural networks in different domains," *IEEE access*, vol. 8, pp. 42 447–42 457, 2020.

[14] W. Fang and L. Li, "Comparison of ring artifacts removal by using neural network in different domains," in *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, 2019, pp. 1–3.

[15] 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," *Medical physics*, vol. 46, no. 11, pp. 4803–4815, 2019.

[16] T. Fu, Y. Wang, K. Zhang, J. Zhang, S. Wang, W. Huang, C. Yao, C. Zhou, and Q. Yuan, "Deep-learning-based ring artifact correction for tomographic reconstruction," *Journal of Synchrotron Radiation*, vol. 30, no. 3, 2023.

Image-to-Image Translation for Spatial Alignment of Sequential Dual-Energy CT Acquisitions

Maximilian Hammermann, Elias Eulig, Joscha Maier, and Marc Kachelrieß

***Abstract*—While dual-energy computed tomography (DECT) is an established technique in clinical radiology, the development of DECT scanners nowadays requires extensive hardware design and technical efforts. Even though the sequential acquisition of two CT scans per slice is theoretically possible using a single energy device and comes with only little requirements regarding hardware adjustments, this approach typically suffers from temporal incoherence due to patient motion. To correct for this, deformable image registration algorithms are usually applied during image post-processing, which may, however, still not achieve a perfect alignment of two subsequent scans. To achieve this and thereby improve subsequent DECT applications like material decomposition in the image domain, we propose a deep learning-based approach that intends to improve alignment of two subsequently acquired DECT images. We therefore simulate imperfect registrations by applying slicewise random deformation vector fields to one scan of a dual-source CT (DSCT) image pair. A deep neural network is then trained to generate an undeformed image from the imperfectly registered image pair in the setting of an image-to-image translation task. In our experiments, the proposed approach produces well-aligned images with the ideal ground truth yielding an average deviation of less than 15 HU inside the patient volume. Furthermore, we show that the approach is able to recover information from the moving images and therefore able to make use of the provided multispectral information.**

***Index Terms*—Dual-energy CT, Convolutional neural network, Deep learning, Patient motion correction**

I. Introduction

DUAL energy computed tomography (DECT) has become an important part of medical diagnostic methods, as it is able to acquire CT data using two different X-ray spectra. This results in a variety of clinical applications mostly being related to the generation of material-specific images (*e.g.* iodine or virtual non-contrast images) [1]. Furthermore, DECT can help to differentiate distinct materials that are represented by the same CT value in a single energy scan. This is for example of high clinical relevance regarding the classification of renal stones, for which their material composition plays a crucial role in choosing a suitable therapeutic strategy [2] or in the context of gout diagnosis [3].

However, the development of DECT scanners requires extensive work referring to hardware considerations in tube-wise (*e.g.* via dual-source CT (DSCT) [4]) or detector-wise approaches (*e.g.* dual-layer detectors [5]). Therefore and due to the diagnostic benefits arising from DECT, performing these scans with standard single energy CT devices and making the possibility therefore available in a wider clinical environment, is a topic of ongoing research.

It should be noted that there have been considerations in the past about generating dual-energy CT or material images from a single energy acquisition alone using deep learning techniques, *e.g.* in [6], [7]. While these approaches illustrate that grayscale correlations in DECT scans of different energies can be learned very effectively making use of the respective pixel-wise anatomical context in the image, they are likely to fail when multispectral CT acquisitions provide additional information for morphological ambiguous contexts, that is not included in a single energy CT image [8].

While the subsequent acquisition of two full CT scans at different energies is typically avoided due to patient motion, using two consecutive gantry rotations with different tube voltages before table incrementation is able to solve this problem to some extent due to increased temporal coherence [9]. Nevertheless, such an approach still requires the application of a deformable registration algorithm during image post-processing [10], which might not be able to produce aligned scans on the pixel level [11].

We therefore present a deep learning-based approach, that is trained to predict spatially aligned DECT scans from coarsely registered sequentially acquired image pairs and ensures the use of the provided multispectral CT image data.

II. Methods

A. Approach

When performing a DECT acquisition in a sequential setup, two CT images of a patient slice are acquired using different tube voltages and potentially prefilters within two subsequent gantry rotations. After reconstruction, images of the lower and the higher tube voltage rotation are obtained, which we denote as f_{low} and f_{high} in the following. In an ideal case without patient motion, f_{low} and f_{high} were well-aligned images, to which image post-processing techniques, such as material decomposition, could be applied. However, in the presence of inter-scan patient motion, a slightly different pair of images, denoted as $(f_{\mathrm{low}}, f'_{\mathrm{high}})$ would be acquired, where f'_{high} denotes a deformed version of the correctly aligned high energy image f_{high} and which has already been post-processed using a deformable image registration algorithm. In order to derive a prediction $\hat{f}_{\mathrm{high}} \approx f_{\mathrm{high}}$ for the desired aligned high energy image, we train a deep neural network to learn the mapping $\mathcal{M}$ given by

$$\mathcal{M} : (f_{\mathrm{low}}, f'_{\mathrm{high}}) \rightarrow f_{\mathrm{high}}. \tag{1}$$

Maximilian Hammermann, Elias Eulig, Dr. Joscha Maier and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls University, Heidelberg, Germany.

Corresponding author: maximilian.hammermann@dkfz.de

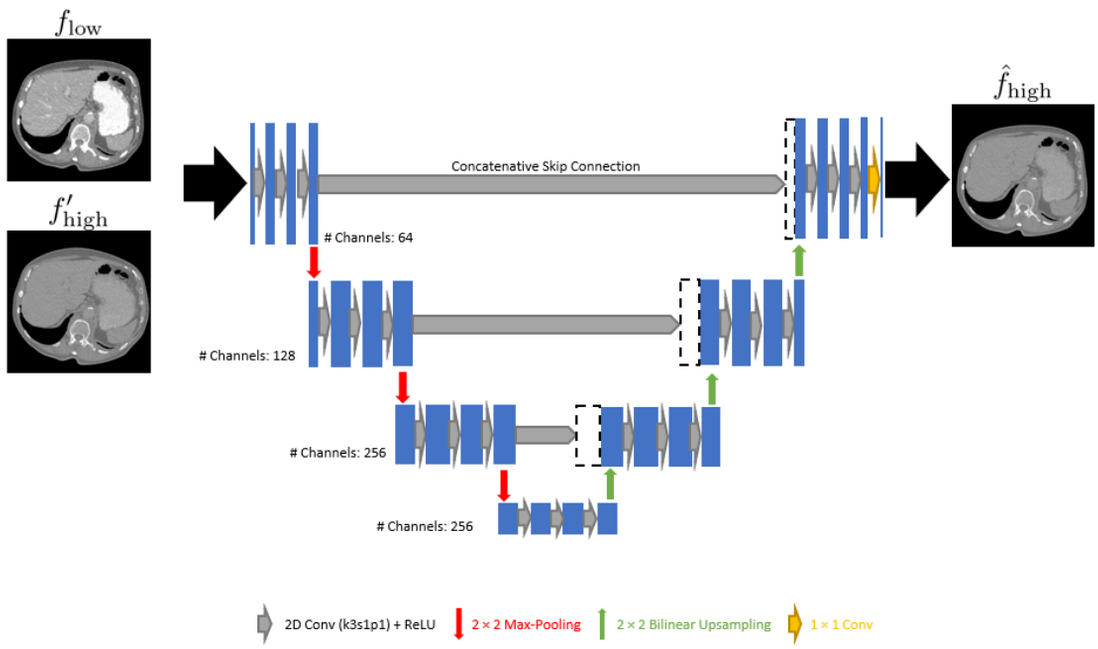


Fig. 1. Illustration of the proposed approach. We employ a 2D U-net architecture in order to infer a high energy CT image, which is aligned with the low energy image given as input.

For this, we are using a U-net like network structure [12] that uses three 2D convolutional layers per stage and takes a low energy scan as well as a deformed high energy scan as a dual-channel input, as depicted in Fig. 1.

B. Data generation

In order to generate artificial data for a supervised training setup, we make use of 42 dual-source CT patient data sets (thorax, abdomen and pelvis scans), which were acquired with a Siemens Somatom Definition Force scanner using a tube voltage combination of 70 kV/150 kV Sn and reconstructed using the Siemens Qr32d reconstruction kernel. Due to the design of the scanner, we restrict our considerations to a part of the field of measurement (FOM) of about 35 cm in diameter, where full dual-energy information is available and preselect patient slices that are fully contained in this part of the FOM. All patients included in the data set received an iodine contrast agent before their examination.

In order to simulate image data of a single-source dual-energy CT acquisition, the scans are assumed to be already coarsely aligned by a preliminary registration algorithm. Potential small remaining misalignment errors between the low and the high energy scans are then modelled by applying two-dimensional random deformation vectors field to the corresponding high energy image slices. As a preliminary registration algorithm may lead to overall unpredictable individual results for a certain image pair, the use of random deformation vector fields seemed to be a justified approach for our considerations.

C. Model and training

We simulate training data as described in Sec. II-B and split our data set in 32 patients for training, while we use five patients for validation and five patients for testing purposes. Since a naive training strategy was shown to lead to shortcuts [13] learned by the network making only little use of the deformed high energy image f'_{high}, we included two probabilistic online data augmentation strategies in the training: First, we dynamically replace a part of the low and high energy images in the training set via convex combinations in order to impede the network to learn an implicit grayscale mapping, *i.e.* $\mathcal{M}' : f_{\text{low}} \rightarrow f_{\text{high}}$ and thereby not making use of the deformed high energy image at all. Specifically, we derive new training samples $\tilde{f}_{\text{low}}$ and $\tilde{f}_{\text{high}}$ from the existing ones by calculating

$$\tilde{f}_{\text{low}} = w_1 f_{\text{low}} + (1 - w_1) f_{\text{high}} \tag{2}$$

$$\tilde{f}_{\text{high}} = w_2 f_{\text{low}} + (1 - w_2) f_{\text{high}} \tag{3}$$

with randomly sampled weights $w_1, w_2 \in [0, 1]$. Moreover, we introduce a random number of spatially independent artificial contrasts in both input images via creating randomly shaped patch masks that are multiplied with the images inside the patient volume.

The optimization then consisted of minimizing an L_1-loss function between the high energy ground truth f_{high} and the respective high energy estimate $\hat{f}_{\text{high}}$ as the output of the network.

The network itself was implemented with PyTorch and the training was performed on a NVIDIA GeForce RTX 3090 using an Adam optimizer with a learning rate of $1 \cdot 10^{-4}$ and a batch size of 8. In that way, we trained the network for 250 epochs and evaluated its generalization capabilities by monitoring the loss evolution for the separate validation patients after every epoch.

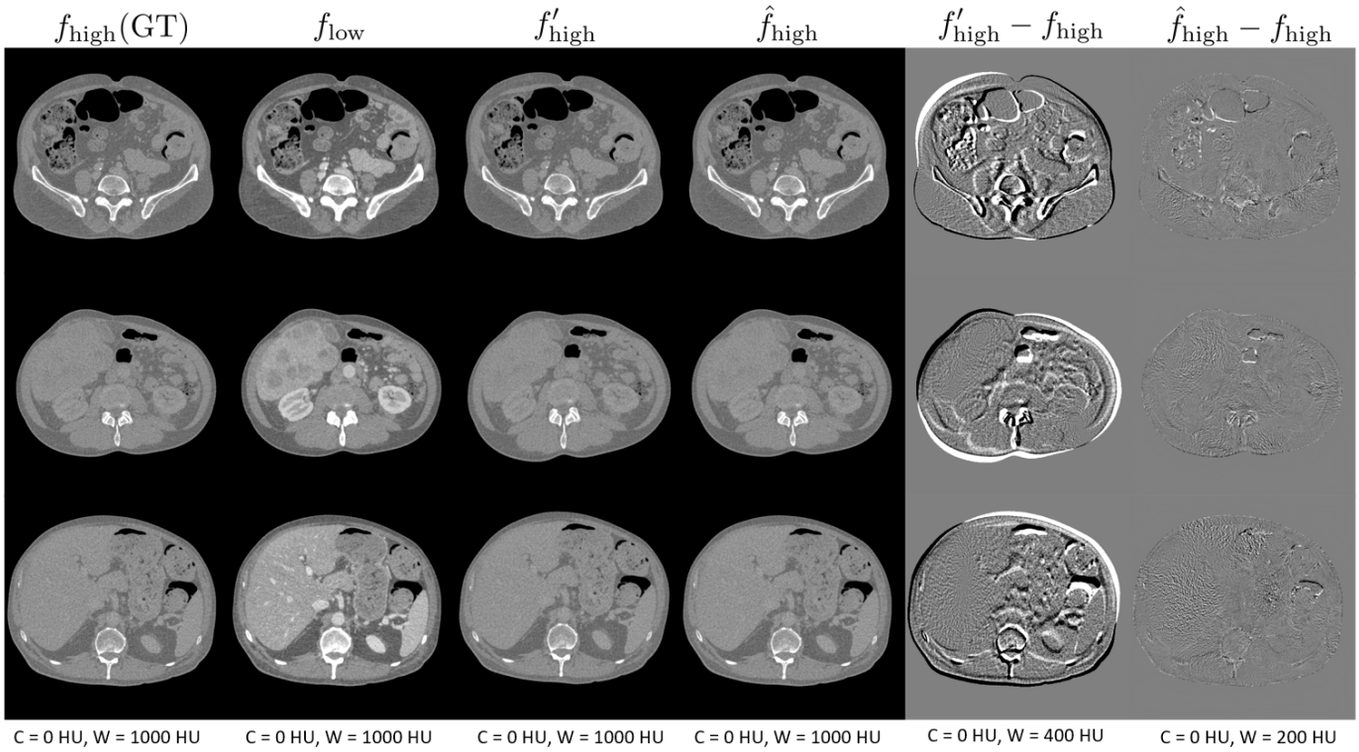


Fig. 2. Three example slices from the test data set. The first column shows the high tube voltage ground truth, whereas the second and third column illustrate the corresponding inputs provided to our network. The fourth column shows the corresponding network prediction, while the fifth and the last column contain difference images of the deformed images as well as the prediction to the ground truth.

III. Results

The performance of our approach was evaluated qualitatively and quantitatively on an independent test data set (see Sec. II-C) using the network configuration with the lowest validation loss during training. In Fig. 2, we show exemplary slices of the respective test patients and the corresponding predictions of our method.

It can be seen, that the resulting network outputs are well aligned with the ground truth and show only minor deviations in the region of interfaces where large differences in CT values occur (*e.g.*, air cavities or patient borders). Moreover, the network produces overall realistically looking image textures with only little deviations compared to the ground truth. However, these deviations are known problems in the context of image-to-image translation tasks and are likely explained by the pixel-wise loss function that is used for network training [14].

For quantitative evaluation, we calculate the mean absolute error (MAE) between the network predictions $\hat{f}_{\mathrm{high}}$ and the corresponding ground truth images f_{high} for our test patient data set using the definition

$$\mathrm{MAE}(f,g) = \frac{1}{\sum_k w_{f\cup g,k}} \sum_k w_{f\cup g,k} |f_k - g_k| \qquad (4)$$

for two images f and g. Here, $w_{f\cup g,k}$ is a weight function that is equal to one if a pixel k is inside the patient in image f or g and zero elsewhere. This yields an overall deviation of (14.7 ± 3.7) HU for the entire test set, whereas a value of (133.7 ± 35.7) HU is obtained without applying further corrections. In terms of the structural similarity index measure (SSIM), we obtain a value of (0.991 ± 0.005) on the complete test set, which is noticeably higher than a value of (0.845 ± 0.040) for the uncorrected scans and indicates high perceptual similarity of the network outputs to the underlying ground truth images.

In order to demonstrate that our network makes use of the provided moving high energy scan f'_{high} and does not learn a simple grayscale mapping from the low to the high tube voltage scan, we generate additional test samples from the existing test set by inserting circular contrasts, which are embedded with equal CT values of 400 HU in the low energy scan, but with different randomly chosen CT values from -200 to 200 HU in the high energy scan. Exemplary results of these experiments are shown in Fig. 3. While this does not represent a real-world clinical situation, it allows to visually evaluate whether the proposed approach makes use of the deformed high energy information or learns a shortcut towards a single-energy mapping. From this illustration, it can be seen that the network is indeed able to resolve artificial ambiguities in the low energy scan and does not simply translate gray scales from the low energy image into a virtual high energy scan. Furthermore, it can be seen that the corresponding circular shapes are nearly perfectly aligned after the application of the network suggesting that it is indeed able to fulfill the downstream image alignment task.

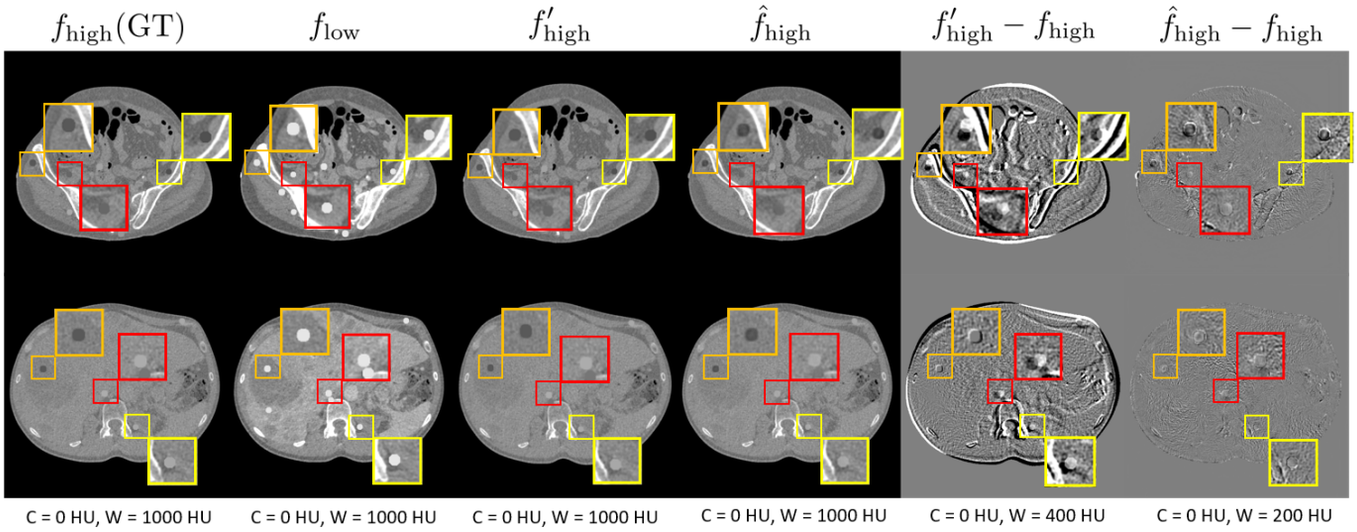


Fig. 3. Two patient slices from the test data set, in which we randomly inserted circular contrasts. All the contrasts in the low energy image were embedded to have a CT value of 400 HU, but their equivalent in the high energy image was chosen as a random value between −200 and 200 HU. The image alignment order follows the one explained below Fig. 2.

IV. Discussion and Conclusion

In this work, we analyzed the use of a deep image-to-image translation approach to provide spatial alignment in coarsely registered sequential DECT acquisitions and illustrate, that the proposed strategy can achieve convincing results: We obtain good alignment of the simulated subsequent scans and are furthermore able to make use of the spectral information of both sequential scans by training our network appropriately.

Despite these promising results of our initial experiments, further research questions arise from the presented findings. This involves an investigation of the robustness of the network to larger deformations as well as the development of further improvements for the training strategy and network structure. Since patient motion typically occurs in three dimensions, it will be required to extend the current implementation in the future to provide alignment in three-dimensional patient scans. However, as we simulated incomplete registrations as random deformations, the presented results indicate that our chosen approach might be very flexible and subsequent extensions might be able to generalize well.

Acknowledgements

Parts of the reconstruction software were provided by RayConStruct® GmbH, Nürnberg, Germany.

References

[1] 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.

[2] A. N. Primak, J. G. Fletcher, and T. J. e. a. Vrtiska, "Noninvasive Differentiation of Uric Acid versus Non–Uric Acid Kidney Stones Using Dual-Energy CT," *Academic Radiology*, vol. 14, no. 12, pp. 1441–1447, Dec. 2007.

[3] A. Y. Tashakkor, J. T. Wang, D. Tso, H. K. Choi, and S. Nicolaou, "Dual-energy computed tomography: a valid tool in the assessment of gout?" *International Journal of Clinical Rheumatology*, vol. 7, no. 1, pp. 73–79, Feb. 2012.

[4] T. G. Flohr, C. H. McCollough, H. Bruder, M. Petersilka, K. Gruber, C. Süss, M. Grasruck, K. Stierstorfer, B. Krauss, R. Raupach, A. N. Primak, A. Küttner, S. Achenbach, C. Becker, A. Kopp, and B. M. Ohnesorge, "First performance evaluation of a dual-source CT (DSCT) system," *European Radiology*, vol. 16, no. 2, pp. 256–268, Feb. 2006.

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

[6] W. Zhao, T. Lv, P. Gao, L. Shen, X. Dai, K. Cheng, M. Jia, Y. Chen, and L. Xing, "A deep learning approach for dual-energy CT imaging using a single-energy CT data," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, S. Matej and S. D. Metzler, Eds. Philadelphia, United States: SPIE, May 2019, p. 27.

[7] C.-K. Liu, C.-C. Liu, C.-H. Yang, and H.-M. Huang, "Generation of Brain Dual-Energy CT from Single-Energy CT Using Deep Learning," *Journal of Digital Imaging*, vol. 34, no. 1, pp. 149–161, Feb. 2021.

[8] J. Maier, J. Erath, S. Sawall, E. Fournié, K. Stierstorfer, and M. Kachelrieß, "Raw data consistent deep learning-based field of view extension for dual-source dual-energy CT," *Medical Physics*, Aug. 2023.

[9] R. Forghani, B. de Man, and R. Gupta, "Dual-Energy Computed Tomography," *Neuroimaging Clinics of North America*, vol. 27, no. 3, pp. 371–384, Aug. 2017.

[10] J. Y. Huang, M. J. Lawless, C. K. Matrosic, L. D. Di Maso, and J. R. Miller, "Evaluation of a commercial deformable image registration algorithm for dual-energy CT processing," *Journal of Applied Clinical Medical Physics*, vol. 21, no. 9, pp. 227–234, 2020.

[11] P. Wang and G. Quan, "From coordinate system mapping to deep feature mapping: A generative deep learning approach for pixel–level alignment of images in rotation-to-rotation DECT," in *Proceedings of the 17th International Meeting on Fully 3D Image Reconstruction*. Fully3D, Jul. 2023, p. D3–P2–14.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," May 2015, arXiv:1505.04597 [cs].

[13] R. Geirhos, J.-H. Jacobsen, C. Michaelis, R. Zemel, W. Brendel, M. Bethge, and F. A. Wichmann, "Shortcut Learning in Deep Neural Networks," *Nature Machine Intelligence*, vol. 2, no. 11, pp. 665–673, Nov. 2020, arXiv:2004.07780.

[14] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss Functions for Image Restoration With Neural Networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

Improving Automated Hemorrhage Detection in Sparse-view Computed Tomography via U-Net based Artifact Reduction

Johannes Thalhammer, Manuel Schultheiß, Tina Dorosti, Tobias Lasser, Franz Pfeiffer, Daniela Pfeiffer, Florian Schaff

Abstract*—*Purpose: **Sparse-view computed tomography (CT) is an effective way to reduce dose by lowering the total number of views acquired, albeit at the expense of image quality, which, in turn, can impact the ability to detect diseases. We explore deep learning-based artifact reduction in sparse-view cranial CT scans and its impact on automated hemorrhage detection.**
Methods: **We trained a U-Net for artifact reduction on simulated sparse-view cranial CT scans from 3,000 patients obtained from a public dataset and reconstructed with varying sparse-view levels. Additionally, we trained the EfficientNetB2 on full-view CT data from 17,545 patients for automated hemorrhage detection. We evaluated the detection performance using the area under the receiver operator characteristic curves (AUC-ROCs) with corresponding 95% confidence intervals (CIs) and the DeLong test, along with confusion matrices. A total variation (TV) approach, commonly applied to sparse-view, served as comparison.**
Results: **The U-Net performed superior compared to unprocessed and TV-processed images with respect to image quality and automated hemorrhage detection. With U-Net post-processing, the number of views can be reduced from 4096 (AUC-ROC: 0.974; 95% CI: 0.972-0.976) views to 512 views (0.973; 0.971-0.975) with minimal decrease in hemorrhage detection (Bonferroni P<.00017) and to 256 views (0.967; 0.964-0.969) with a slight performance decrease (Bonferroni P<.00017).**
Conclusion: **U-Net based artifact reduction substantially enhances automated hemorrhage detection in sparse-view cranial CTs. Appropriate post-processing is crucial for optimal image quality and diagnostic accuracy while minimizing radiation dose.**

***Index Terms*—Artifact reduction, computed tomography (CT), deep learning, hemorrhage detection, sparse-view CT, U-Net**

I. Introduction

Intracranial hemorrhage (ICH) is a potentially life-threatening disease with a 30-day fatality rate of 40.4%, accounting for 10% to 30% of annual strokes [1][2]. Prompt and accurate diagnosis is vital for optimal treatment, typically achieved through cranial Computed Tomography (CT) scans [3]. However, the increased use of medical CT scans, including cranial scans, raises concerns about radiation-related health risks [4]. Besides lowering the X-ray tube current, sparse-view CT is a promising approach to reduce dose by decreasing the number of views. However, this reduction causes artifacts in the filtered back projection (FBP) reconstructed images. To restore image quality for accurate diagnosis, suitable processing methods are needed. In the past, compressed sensing (CS) and iterative reconstruction approaches have been widely investigated, which generally minimize a CS-based regularization term as well as a data-fidelity term that ensures data consistency [5]. These approaches have been proven to yield good results in terms of reducing image noise and artifacts. However, they are also computationally demanding due to the repeated update steps during iterative optimization. Additionally, these methods require parameter optimization and can alter image texture. Recently, machine learning approaches using deep neural networks have gathered vast scientific attention. In the context of artifact reduction in sparse-view CT, it has been shown that convolutional neural networks are able to achieve excellent results with comparably low computational effort during inference [6] [7]. In parallel, extensive research has been conducted on the application of deep learning techniques for automated detection and classification of pathological features in CT images and many systems are already FDA-approved and CE-certified [8]. Typically, these algorithms are trained on standard-dose data, which restricts their applicability to dose-reduced or sparse-view data. This is because the additional noise and artifacts in sparse-view data are expected to negatively impact the reliability of these systems.
Therefore, in this study, our aim is twofold. Firstly, we aim to explore the potential benefits of deep learning-based artifact reduction in sparse-view cranial CTs. Secondly, we assess whether this approach can enhance the performance of automated hemorrhage detection.

This work was funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder, the German Research Foundation (GRK2274), as well as by the Technical University of Munich–Institute for Advanced Study.

Johannes Thalhammer, Franz Pfeiffer and Daniela Pfeiffer are with the Chair of Biomedical Physics, Department of Physics, School of Natural Sciences, Munich Institute of Biomedical Engineering, Institute for Advanced Study, all at the Technical University of Munich, 85748 Garching, Germany; Department of Diagnostic and Interventional Radiology, School of Medicine and Klinikum Rechts der Isar, Technical University of Munich, 81675 Munich, Germany (e-mail: johannes.thalhammer@tum.de). Manuel Schultheiß and Tina Dorosti are with the Chair of Biomedical Physics, Department of Physics, School of Natural Sciences, Munich Institute of Biomedical Engineering, all at the Technical University of Munich, 85748 Garching, Germany; Department of Diagnostic and Interventional Radiology, Klinikum rechts der Isar, Technical University of Munich, 81675 Munich, Germany. Florian Schaff is with the Chair of Biomedical Physics, Department of Physics, School of Natural Sciences; Munich Institute of Biomedical Engineering, all at the Technical University of Munich, 85748 Garching, Germany. Tobias Lasser is with the Computational Imaging and Inverse Problems, Department of Computer Science, School of Computation, Information, and Technology; Munich Institute of Biomedical Engineering, all at the Technical University of Munich, 85748 Garching, Germany.

II. Materials and Methods

Network Architectures - The artifact reduction was performed using a U-Net architecture with five levels of depth and an additional skip connection between network input and output [9]. Pooling is performed by strided convolutions with stride 2x2, followed by a 1x1 convolution. Unpooling is implemented with transposed strided convolutions. For hemorrhage detection, the EfficientNetB2 was used [10].

Datasets - This retrospective study used the RSNA 2019 Brain CT Hemorrhage Challenge dataset, with each CT image annotated for hemorrhage presence and subtype (subarachnoid, intraventricular, subdural, epidural, and intraparenchymal) [11]. The dataset consists of CT scans from 18,938 different patients. After filtering out images that either did not have a resolution of 512x512 pixels, or led to various errors during data generation, the dataset was narrowed down to 18,545 patients. We randomly selected 1,000 patients for testing and excluded them from all further training. For U-Net training, 3,000 patients were randomly selected. 2,400 patients (80%) served as training data, the remaining 600 patients (20%) as validation data. Sinograms with 4,096 views were created under parallel beam geometry from the CT dataset. Ground truth images were reconstructed from 4,096 views using FBP. The number of views was selected to prevent the introduction of new artifacts in the ground truth data compared to the source images. Six sparse-view subsets were generated using FBP with only 64, 128, 256, 512, 1,024 and 2,048 views, respectively. Due to overfitting concerns, the remaining dataset was used for training the EfficientNetB2 (17,545 patients). Individual images were scaled to Hounsfield Units (HUs) and clipped to the diagnostically relevant brain window [0 HU-80 HU]. To meet the requirements of the pre-trained ImageNet, they were transformed to three channel images by concatenating three neighboring CT slices. Figure 1 depicts the data selection process.

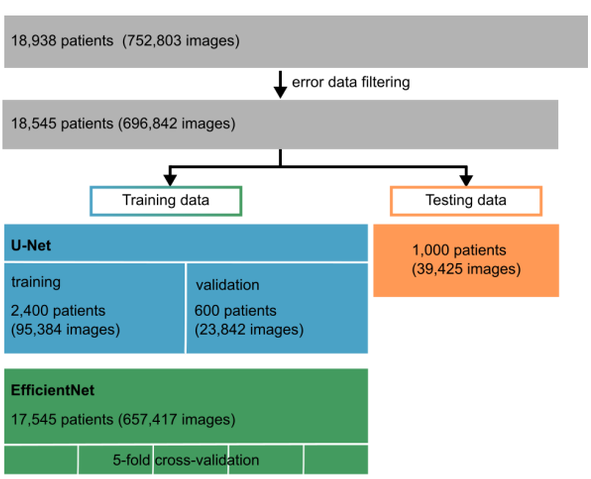


Fig. 1. Flowchart of the data selection process.

Total Variation - In this work, we used the isotropic total variation (TV) method by Chambolle for artifact reduction [12]. The optimal weight for each sparse subset was determined by randomly sampling 1,000 images from the U-Net training set and iterating through weights ranging from 0.001 to 1.000 in 0.001 increments. We identified a global maximum within this weight range for each subset, suggesting the range was reasonable. The weights that yielded the best score on the structural similarity index measure (SSIM) were selected to calculate the metrics on the test set.

Saliency Maps - The saliency maps were obtained by computing the gradient of the class score with respect to the input image, as described in [13]. To quantify the saliency maps, we selected 100 images from the test set containing intraparenchymal hemorrhages. The hemorrhage regions were manually segmented and verified by a board-certified radiologist. For the selected images, we then calculated the saliency map for each sparse-view level with and without U-Net post-processing, respectively. The segmented masks were used to compute the ratio of the sum of saliency map values within the mask to the sum of values outside the mask.

Statistical Analysis - Image quality of the different post-processing methods were compared using SSIMs. To ensure a focus on diagnostically relevant pixels while excluding potentially distorting background pixels, the metric was applied within a mask of the intracranial region, which was extracted using the CNN developed and trained by Cai et al. [14]. To quantify hemorrhage detection performance, empirical area under the receiver operator characteristic curves (AUC-ROCs) including 95% CIs were estimated as described by DeLong et al. [15]. Statistical differences in AUC-ROCs between sparse-view and full-view, as well as between different post-processing methods, were evaluated using the DeLong two-sided test. Bonferroni correction was applied, leading to a significant level of P=0.00017 (0.0010/6) for comparing the six sparse-view subsets with the full-view data, and to a significant level of P=0.00033 (0.0010/3) for comparing raw FBP, TV post-processing, and U-Net post-processing. The saliency map ratios with and without U-Net post-processing were compared by using the Wilcoxon signed-rank test.

III. Results

Artifact Reduction - Figure 3 shows a CT image from the test set, reconstructed with varying number of views. While in (a) no artifacts are visible, the image quality deteriorates if the number of views is further reduced (b) - e)). In images f) - i) the results of TV post-processing are depicted, which can reduce the artifacts well down to 256 views sparse sampling. In images k) - n) the U-Net predictions of the sparse-view images are depicted, revealing a clear reduction in streak artifacts compared to the respective input images. Table I presents the mean SSIM values of the reconstructed images calculated on the test set. Both, U-Net and the TV post-processing quantitatively increased the SSIM by reducing streak artifacts compared to raw FBP reconstructions. Direct comparison revealed U-Net's stronger performance in all cases.

Detection of Hemorrhage Subtypes -

The impact of artifact reduction on automated hemorrhage detection is shown in figure 3, in which the AUC-ROC values of the EfficientNetB2 hemorrhage detection network are depicted. Overall, detection performance declines as the number

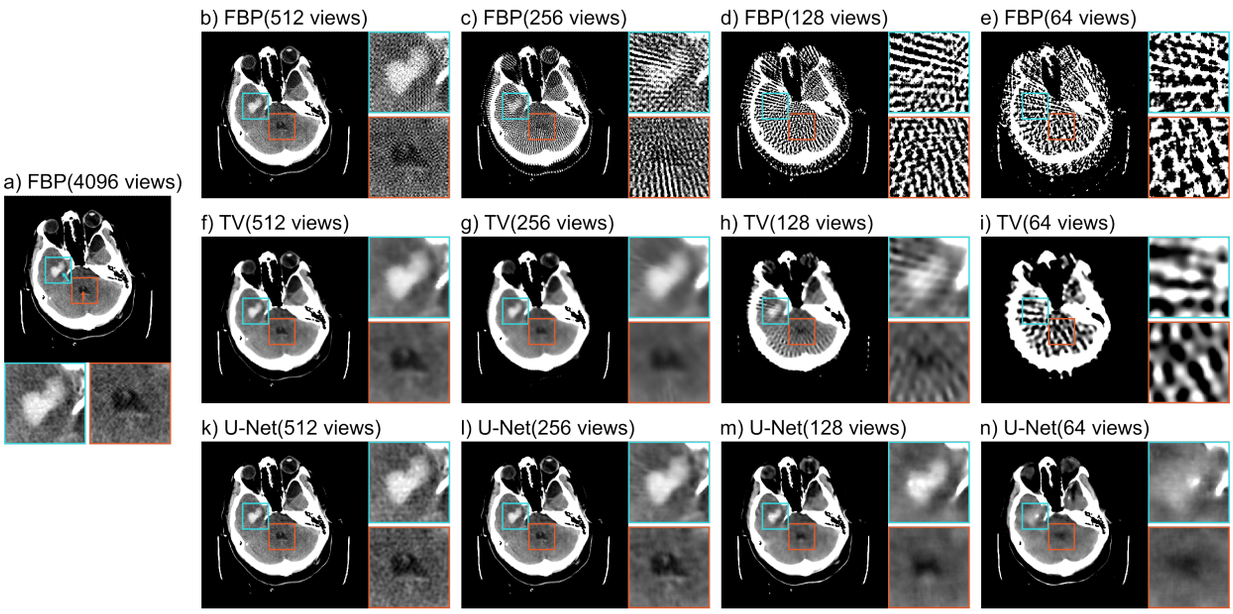


Fig. 2. CT image from the test set labeled with intraparenchymal (cyan arrow) and intraventricular (orange arrow) hemorrhage. Image a) shows the image reconstructed from 4096 views using FBP. Images b) - e) show the same image reconstructed from 512, 256, 128, and 64 views, respectively. Images f)- i) show the result of the TV post-processing, images k) - n) the U-Net predictions of the corresponding sparse-view images in the upper row. All images are presented in the brain window ranging from 0 HU to 80 HU. Both inserts are 80x80 pixels.

#views	**2048**	**1024**	**512**	**256**	**128**	**64**
FBP	1.000	0.994	0.979	0.943	0.815	0.655
U-Net	1.000	**1.000**	**1.000**	**0.999**	**0.998**	**0.997**
TV	1.000	0.999	0.999	0.998	0.991	0.975

TABLE I
STRUCTURAL SIMILARITY INDEX MEASUREMENT (SSIM) VALUES OF THE REGION OF INTEREST OF THE SPARSE-VIEW CT IMAGES OF THE TEST SPLIT.

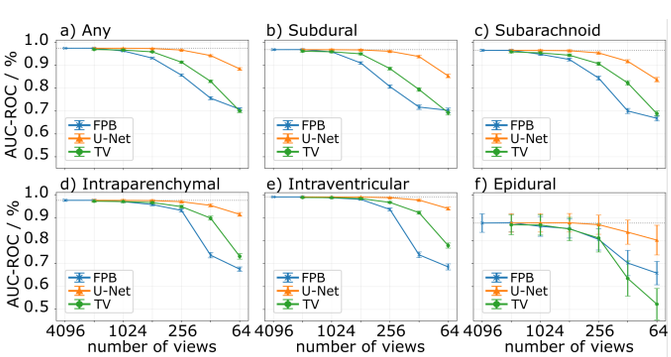


Fig. 3. Results of the EfficientNetB2 classification network. Images a) - f) depict the mean AUC values (95% CI) of the ROC curves for varying number of views for each subtype, respectively. The images were either not post-processed (blue), post-processed by the U-Net (orange) or post-processed via TV (green).

of views used for the reconstruction decreases. When reducing the number of views to 1024, there is a slight decrease in performance (P=0.012 for epidural, P<.00017 for other subtypes), which further drops with fewer views. In TV post-processed images, the AUC-ROC values slightly decrease down to 512 views (P=0.00033 for epidural, P<.00017 for other subtypes) and decline substantially for fewer views. With the U-Net, the number of views can be reduced from 4096 views down to 512 views with a minimal decrease (P=0.62 for epidural, P<.00017 for other subtypes) in detection performance, and down to 256 views with only a slight performance decrease. Below 256 views, a noticeable decline is visible. For all the cases reconstructed with 1024 views or fewer, except the epidural subtype, the AUC-ROC values obtained from U-Net post-processed images surpass those of TV-processed and raw sparse-view images significantly (P<.00033). For the epidural subtype, a significant difference is only observable for 512 views or fewer. Notably, the detection of epidural hemorrhages is generally poorer compared to other subtypes.

Saliency Maps - Figure 4 shows the saliency maps for the images of figure 3 with regards to the 'any' class. The rectangles are positioned identically to those in figure 2, indicating the location of the hemorrhages. For images reconstructed down to and including 256-view sparse-sampling (subfigures a) - c)), the network primarily focused on the area of the intraparenchymal hemorrhage for its prediction. However, for reconstructions with fewer views (subfigures d) and f)), such a focused area is no longer discernible. Conversely, for the U-Net post-processed images (subfigures f) - i)) all the saliency maps exhibit a clear emphasis on the subarachnoid hemorrhage area. The results of the quantitative saliency map analysis are displayed in II. When using 128 views or fewer, a significant difference is discernible between raw FBP reconstructions and U-Net post-processed ones (P<0.0010).

#views	**2048**	**1024**	**512**	**256**	**128**	**64**
FBP	0.094	0.092	**0.100**	0.086	0.052	0.016
U-Net	0.094	**0.094**	0.094	**0.098**	**0.104**	**0.107**
P-value	0.470	0.176	0.967	0.020	<0.001	<0.001

TABLE II
AVERAGE RATIO OF THE SALIENCY MAP VALUES INSIDE THE INTRAPARENCHYMAL MASK TO THOSE OUTSIDE THE MASK WITHOUT (FBP) AND WITH U-NET POST-PROCESSING OF 100 TEST SAMPLES. THE LAST ROW DEPICTS THE P-VALUES ASSOCIATED WITH THE WILCOXON TWO-SIDED STATISTICAL TEST, COMPARING THE FBP AND U-NET VALUES FOR STATISTICAL DIFFERENCES.

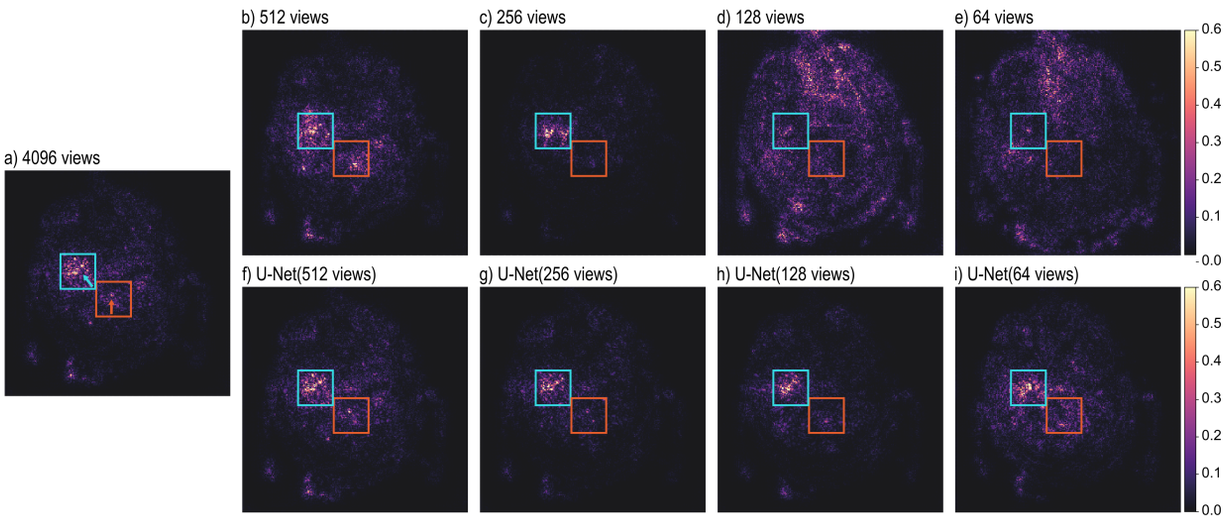


Fig. 4. Saliency maps by the EfficientNetB2 model of the CT images of figure 2 with regards to the 'any' class.

IV. Conclusion

In this work we investigated convolutional neural network-based artifact reduction in sparse-view cranial CTs and the subsequent impact on automated hemorrhage detection. We demonstrated that a deep convolutional neural network, specifically a U-Net architecture, leads to substantial improvements in the visual quality of sparse-view cranial CTs. Furthermore, we trained a hemorrhage detection network on full-view CT images and applied it to sparse-view images with and without post-processing by the U-Net. The results showed that the U-Net enabled a reduction in views from 4096 to 512 with minimal impact on detection performance, and to 256 views with only a slight performance decrease. The U-Net was found to perform superior to TV post-processing with respect to image quality parameters and automated hemorrhage detection. Saliency maps provided additional support to these findings by revealing that pronounced artifacts in sparse-view images disrupted the detection process, leading to increased false positives and false negatives. By reducing these artifacts through post-processing, these effects can be mitigated, and the detection performance can be maintained, without the need for specific training on sub-sampled data. It is important to acknowledge some limitations of this study. The sparse-view data used in our study was retrospectively generated under simplified conditions from CT volumes, which may not fully capture the complexity of real-world scenarios. Additionally, even without post processing, a meaningful decline in detection performance was only observed for subsets using fewer than 2048 views. This suggests that using fewer views for the full-view image would also be sufficient. Our findings highlight the importance of employing appropriate post-processing techniques to achieve optimal image quality and diagnostic accuracy while minimizing radiation dose. Furthermore, our study demonstrates that leveraging deep learning methods for artifact reduction can lead to significant improvements in the detection of sparse-view cranial CTs. This has promising implications for automated hemorrhage detection on sparse-view cranial CT data to assist radiologists in clinical practice.

References

[1] C. J. van Asch, M. J. Luitse, G. J. Rinkel, I. van der Tweel, A. Algra, C. J. Klijn, *The Lancet Neurology* **2010**, *9*, 167–176.

[2] V. L. Feigin, et al., *The Lancet Neurology* **2009**, *8*, 355–369.

[3] S. Currie, N. Saleem, J. A. Straiton, J. Macmullen-Price, D. J. Warren, I. J. Craven, *Postgraduate Medical Journal* **2016**, *92*, 41–50.

[4] D. J. Brenner, E. J. Hall, *New England Journal of Medicine* **2007**, *357*, DOI 10.1056/nejmra072149.

[5] Z. Hu, J. Gao, N. Zhang, Y. Yang, X. Liu, H. Zheng, D. Liang, *Scientific Reports* **2017**, *7*, 10747.

[6] Y. Han, J. C. Ye, *IEEE Transactions on Medical Imaging* **2018**, *37*, DOI 10.1109/TMI.2018.2823768.

[7] T. Cheslerean-Boghiu, et al., *IEEE Transactions on Computational Imaging* **2023**, *9*, DOI 10.1109/TCI.2023.3240078.

[8] X. Wang, T. Shen, S. Yang, J. Lan, Y. Xu, M. Wang, J. Zhang, X. Han, *NeuroImage: Clinical* **2021**, *32*, 102785.

[9] O. Ronneberger, P. Fischer, T. Brox in Lecture Notes in Computer Science, *Vol. 9351*, **2015**.

[10] M. Tan, Q. V. Le in 36th International Conference on Machine Learning, ICML 2019, *Vol. 2019-June*, **2019**.

[11] A. E. Flanders, L. M. Prevedello, e. a. Shih, *Radiology: Artificial Intelligence* **2020**, *2*, e190211.

[12] A. Chambolle in Journal of Mathematical Imaging and Vision, *Vol. 20*, **2004**.

[13] K. Simonyan, A. Vedaldi, A. Zisserman in 2nd International Conference on Learning Representations, ICLR 2014 - Workshop Track Proceedings, **2014**.

[14] J. C. Cai, et al., *Radiology: Artificial Intelligence* **2020**, *2*, DOI 10.1148/ryai.2020190183.

[15] E. R. DeLong, D. M. DeLong, D. L. Clarke-Pearson, *Biometrics* **1988**, *44*, DOI 10.2307/2531595.

Deep learning-based tumour segmentation on spectral photon-counting CT images

Margarete Kattau, Oliver Pickford Scienti, Ben Glocker, Dimitra Darambara

***Abstract*—The improved soft tissue contrast and quantitative imaging capabilities of photon-counting CT are potentially of great use in radiotherapy treatment planning. Therefore, we propose photon-counting CT as a single modality approach in combination with deep learning-based tumour segmentation, with the potential of improvements in planning time, radiation dose, and accuracy.**

We performed a range of simulation studies using a combination of Monte Carlo and finite element models to simulate patient CT acquisitions and to train a deep learning segmentation model on these. The performance of a model trained on photon-counting detector (PCD) images is higher than the performance of one trained on energy-integrating detector (EID) images (DSC: 87.1% PCD vs. 81.9% EID), indicating that the PCD images may contain more information relevant to the segmentation task than EID images. This was also observed when comparing low energy PCD images with a bin of [0, 75] keV to EID images (DSC: 88.7% low energy PCD vs. 80.3% EID).

Ongoing work involves a more detailed assessment of how segmentation performance varies with energy bin placement. Additionally, the robustness of the segmentation network to electronic noise is being investigated. We expect that PCD trained models will be more robust to noise than their EID counterparts. The results of this part are currently being processed and will be presented at the meeting.

***Index Terms*—photon-counting CT, deep learning, tumour segmentation, radiotherapy planning**

I. Introduction

With a 5 year overall survival below 35%, malignant primary brain tumours are among the most challenging cancers to treat [1]. Additionally, about 8% of cancer patients will develop brain metastases [2]. Due to the location of these tumours within the brain, highly efficient and accurate neuro-oncological treatment is necessary to ensure tumour control while minimising radiation-induced side effects. Common treatment approaches are stereotactic radiosurgery (SRS) and proton therapy.

Currently, SRS treatment planning involves a multi-modal approach, where the MRI scan is used to manually delineate the tumour, whereas a CT scan is needed to calculate the patient dose based on the electron density. Besides a reduction in dose compared to conventional CT [3], photon-counting CT has been shown to improve contrast to noise ratio and differentiation between white matter and gray matter within the brain [4]. Differentiation of such closely related soft tissues raises the prospect that this technology also enhances the contrast between brain tumours and healthy tissue. Photon-counting CT could therefore be employed in a single modality approach, alleviating dosimetric uncertainties caused by the MRI-CT image registration. For proton therapy treatment planning, the proton stopping power - which determines the dose to the patient - is also calculated based on the electron density. Previous studies have demonstrated the improvement in stopping power calculations using dual-energy CT compared to conventional single-energy CT [5] and the potential of photon-counting CT to improve both planning and treatment accuracy in these highly specialised radiotherapy applications [6].

In addition to the integration of photon-counting CT into the SRS and proton therapy planning workflow, further improvement can be achieved by using a deep learning model to automatically delineate the tumour, saving valuable time for clinicians. Previous experiments we conducted have shown a low performance of automatic segmentation models on conventional CT, which is why we are investigating the improvement offered by the combination of photon-counting CT and deep learning-based segmentation.

In this work, we present an *in silico* imaging pipeline for the simulation and reconstruction of photon-counting CT brain scans. A deep learning-based model is then used to automatically segment tumours on these scans as a proof-of-concept. We investigate the robustness of the deep learning models to electronic noise. Additionally, we will evaluate how energy threshold placement affects segmentation performance to establish optimal windowing for this task.

II. Methods

A. Image acquisition and reconstruction

To develop the *in silico* imaging system, the in-house software tool CoGI [7] is used, which combines Monte Carlo simulations (GATE [8]) with finite element methods (COMSOL [9]) to simulate the image acquisition process. The resulting geometry is shown in Figure 1 and consists of a 2 mm thick CdTe line detector and a 120 kVp X-ray source. Conventional CT scans from the open-source data set GLIS-RT [10] are used to build a voxelised patient volume and images of a single brain slice are reconstructed for each patient using the reconstruction toolkit CIL [11][12], both as acquired with a photon-counting detector (PCD) and with an energy-integrating detector (EID). This enables a direct comparison and allows us to assess the potential improvement achievable

M. Kattau, O. Pickford Scienti, and D. Darambara are with the Joint Department of Physics at the Institute of Cancer Research and the Royal Marsden NHS Foundation Trust, London. M. Kattau and B. Glocker are with the Department of Computing at Imperial College London. This project is a collaboration between both departments. The authors would like to acknowledge NHS funding to the NHIR Biomedical Research Centre at the Royal Marsden NHS Foundation Trust and the Institute of Cancer Research.

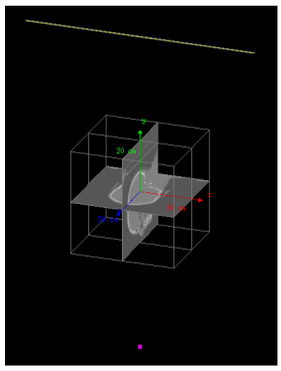


Fig. 1: The imaging system developed within GATE. The voxelized phantom based on a patient CT scan is shown in the center. The CdTe detector can be found at the top of the image as a yellow line with the X-ray source positioned opposite at the bottom as a magenta volume.

using a PCD. In total, three different voxelised phantoms are imported.

B. Deep-learning based tumour segmentation

Once the images are reconstructed, a generative algorithm is introduced, applying random perturbances to the EID and PCD images in a pairwise manner. The same modifications are applied to the corresponding ground truth segmentations provided in the GLIS-RT data set. This type of data augmentation enables the training of a deep learning model without the need to simulate a large number of images. A two-dimensional nnU-Net [13] is used for tumour segmentation due to its applicability to a variety of image segmentation tasks. A schematic depiction of the segmentation pipeline is provided in Figure 2.

As a first experiment, the EID images are compared to PCD images with a 0 keV threshold. In both cases, the training set consists of 200 random variations of the images of patient 1 and 2. Within the nnU-Net pipeline, each training set is randomly split into 160 training and 40 validation images. The trained model is then applied to a testing set consisting of 100 random variations of patient 3. The performance is assessed by calculating the Dice similarity coefficient (DSC), which measures the overlap between the predicted segmentations and the corresponding ground truth, and the average symmetric surface distance (ASSD), which describes the mean spatial distance of the contours of the predicted segmentation and the ground truth.

C. Deep-learning based tumour segmentation on spectral images

To assess the performance of automatic tumour segmentation on spectral images, further experiments are conducted using a low energy bin ([0, 75] keV) and a high energy bin ([75, 120] keV) PCD image instead of an image with a 0 keV threshold. The parameters of the segmentation pipeline and model are kept the same as in the previous section.

D. Deep-learning based tumour segmentation on different noise levels

To assess the robustness of the approach to electronic noise, the imaging pipeline is extended to incorporate electronic noise. The way in which electronic noise manifests in EID- and PCD-based systems differs fundamentally from each other. While the electronic noise is integrated into the signal measured by an EID, it can be excluded from the PCD image by setting a suitable threshold. Therefore, the PCD image can be considered free from electronic noise with the exception of a shift in the detected energy. To implement this into our pipeline, the following correction is done for each event measured in the PCD:

$$x_{new} = x + a, \quad a \sim \mathcal{N}(\mu, \sigma^2), \tag{1}$$

where x corresponds to the energy of the detected event, a is a randomly sampled value of electronic noise, and x_{new} describes the new energy corrected by this noise value. $\mathcal{N}(\mu, \sigma^2)$ represents a Gaussian distribution with mean μ and variance σ^2. For an image with realistic electronic noise, μ is set to 5 keV and σ is set to $\mu/3$, based on a prototype system previously used in the group. By increasing the value of μ, the level of noise in the image is raised. To exclude the resulting false counts, the images are reconstructed with a threshold of $\mu + 3\sigma$, which is an approach commonly used in PCD-based systems.

The implementation of electronic noise into the EID in our system is conducted in the image domain based on the following criterion for high image quality in a clinical CT system: the variation in the pixel value for air in a region of interest should be within $\pm$ 10 HU [14], corresponding to a 1% deviation. Higher noise levels are modelled by increasing this tolerance value.

III. Results and Discussion

A. Image acquisition and reconstruction

Figure 3 shows the reconstructed PCD and EID images of two different patients 1 and 2 [15]. To demonstrate the presence of structures within the simulated brains and improve their visibility to the human eye, additional images are shown with an adjusted gray scale. For both patients, the ventricles can be identified, as well as tissue abnormalities caused by the tumour. This demonstrates our ability to simulate and reconstruct brain images based on patient CT scans.

B. Deep-learning based tumour segmentation

The results of the deep learning experiments are presented in Table I. A higher performance is observed for the PCD 0 keV images as compared to the EID images. We propose that this is due to a greater soft tissue contrast provided by PCD-based systems and could indicate that the PCD images contain more diagnostically relevant data than the EID images. Additionally, we see a higher performance when applying the model trained on PCD images to EID images as compared to vice versa. This suggests a higher generalisibility of the PCD model compared to the EID one.

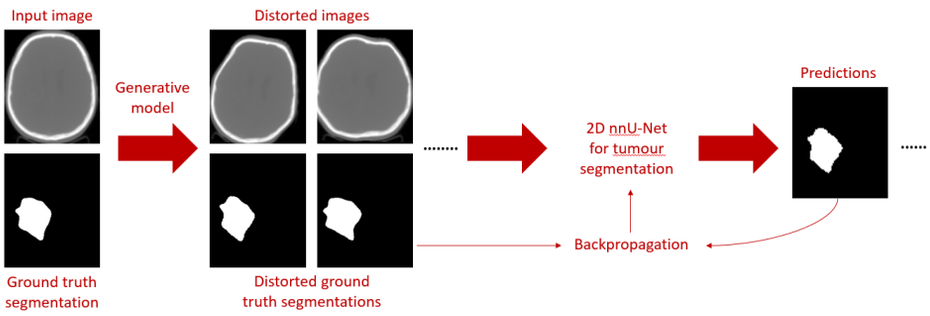


Fig. 2: Pipeline for the automatic tumour segmentation based on a generative algorithm and a 2D nnU-Net.

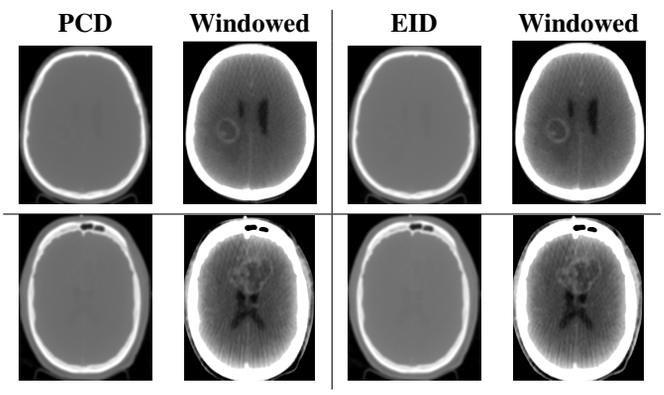


Fig. 3: Images simulated of patients 1 (top row) and 2 (bottom row). The images in the first column are reconstructed using a PCD with a 0 keV threshold. The images in the third column are reconstructed using an EID. To visualise structures within the brain, the second and fourth columns show corresponding windowed images with the gray scale adjusted.

Detector type	DSC [%]	ASSD [mm]
PCD (0 keV threshold)	87.1 ± 1.9	5.7 ± 0.8
EID	81.9 ± 4.2	9.3 ± 3.0
Cross-validation		
PCD (0 keV threshold)	82.4 ± 3.1	8.6 ± 1.8
EID	61.7 ± 6.9	25.5 ± 5.5

TABLE I: The first half of the table shows the performance of the 2D nnU-Net models trained and tested on the reconstructed PCD and EID images. The second half shows the results of the cross-validation where the PCD-trained model was applied to EID images and vice versa. Values are shown as mean $\pm$ standard deviation.

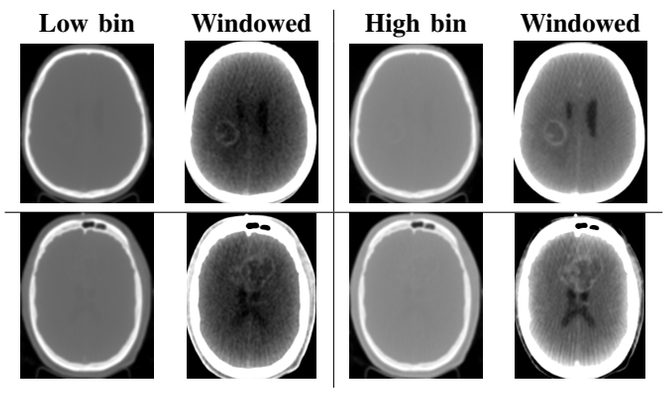


Fig. 4: Reconstructed spectral low and high energy images with bins of [0, 75] keV and [75, 120] keV, respectively. The top row shows the images of patient 1, the bottom row those of patient 2. The first column shows the images reconstructed with the low energy bin and the third column those reconstructed with the high energy bin. The second and fourth columns show the corresponding images with the gray scale adjusted to visualise brain structures.

Detector type	DSC [%]	ASSD [mm]
PCD ([0, 75] keV bin)	88.7 ± 3.0	5.5 ± 1.9
PCD ([75, 120] keV bin)	71.1 ± 8.2	15.2 ± 4.3
EID	80.3 ± 5.6	10.5 ± 3.4

TABLE II: Performance of the 2D nnU-Net models trained and tested on the reconstructed spectral PCD and EID images. Values are shown as mean $\pm$ standard deviation.

C. Deep-learning based tumour segmentation on spectral images

Figure 4 shows the high and low energy PCD images reconstructed based on patient 1 and 2. Upon visual inspection, the low energy images feature an increased contrast between the tumour structures and healthy brain tissue. This is confirmed by the quantitative results of the nnU-Net experiments presented in Table II. The model trained and tested on the low energy PCD images achieves the highest performance when compared to high energy PCD images and EID images. Example predictions are shown in Figure 5. The low energy PCD model has a closer overlap with the ground truth prediction compared to the high energy and EID models. The performance of the low energy PCD model is similar to the model trained on non-spectral PCD images, indicating that most of the diagnostically relevant information is carried by photons with an energy below 75 keV for our particular system and task.

Ongoing work involves a more detailed investigation by assessing multiple bin placements with the aim of improving the segmentation performance and determining the optimal bin placement for this task.

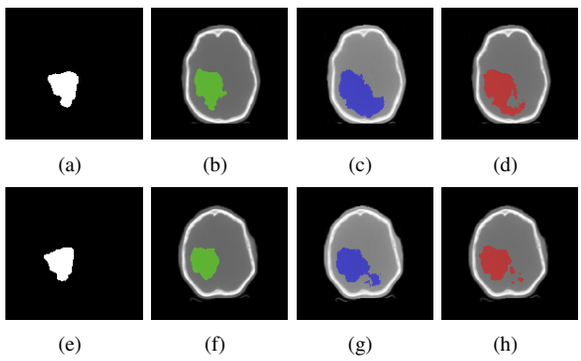


Fig. 5: Example predictions shown for two different images: (a) and (e) show the corresponding ground truth maps, (b) and (f) the predictions in green of the model trained on low energy PCD images, (c) and (g) the predictions in blue of the model trained on high energy PCD images and (d) and (h) the predictions in red of the model trained on EID images.

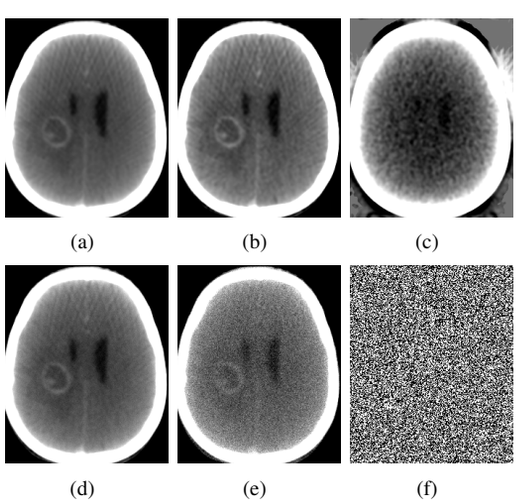


Fig. 6: Example PCD (top row) and EID (bottom row) images of patient 1 reconstructed with low noise (a and d) and high noise (b and e). (c) and (f) show the difference between the high and low noise images.

D. Deep-learning based tumour segmentation on different noise levels

In Figure 6, the left column shows PCD (top row) and EID images (bottom row) of patient 1 reconstructed with a low noise level. The middle column shows the same image reconstructed with a high noise level. For the PCD images, the low noise value for μ was set to 5 keV and the high noise value was set to 25 keV. For the EID images, the tolerance value for low noise was set to 1% and the tolerance value for high noise was set to 5%. A direct comparison between the noisy PCD and EID images is not appropriate, as the modelled level of noise in a PCD system is based on a prototype, whereas the modelled noise in the EID system is set according to guidelines for clinical systems. However, we can investigate the effect of the noise on the image features when increasing the noise level from low to medium. For the PCD images, there is only a small decrease in image quality and all relevant features of the tumour and the healthy brain remain clearly visible. On the other hand, the degradation in image quality for the EID images is significant when raising the noise level, blurring features of the tumour which are likely important for its correct segmentation.

The right column in Figure 6 is produced by subtracting the high noise images from the low noise images. The noise in the PCD-based system still shows the structures within the brain, whilst the noise in the EID-based system manifests all over the image. This explains why the image quality is preserved better in the PCD-based system when increasing the noise level, as the noise is structured on the underlying anatomy rather than being random, as is the case in the EID-based system.

Ongoing work involves modelling a higher number of noise levels and training one nnU-Net model per noise level and detector type in order to assess nnU-Net's robustness against a degradation in image quality caused by electronic noise.

IV. Conclusion

The CT images simulated and reconstructed with the proposed *in silico* imaging pipeline feature brain anatomy and pathological changes caused by cancerous tissue, demonstrating the feasibility of the approach. Deep learning-based tumour segmentation achieves a higher performance when applied to PCD images as compared to EID images. The same observation holds when the approach is applied to low energy PCD images and EID images. The EID and PCD images are simulated and reconstructed with the same pipeline and the only difference between the two image types is the way the detected events are processed and counted in the detector. Additionally, the same nnU-Net configuration was used for both. We can therefore conclude that the presented improvement in segmentation performance is solely due to an enhanced image quality provided by the PCD images.

Furthermore, the imaging pipeline is capable of simulating and reconstructing images of different electronic noise levels. Preliminary results show that the PCD-based system obtains qualitatively better images when the noise level is increased.

These observations demonstrate the potential of integrating photon-counting CT into the SRS and proton therapy treatment planning workflow. Ongoing work involves a more detailed assessment of the benefit of including spectral information and a quantitative investigation of the impact of electronic noise on segmentation performance.

References

[1] S. Lapointe et al. *Lancet*, 392(10145):432–446, 2018.
[2] B. Lippitz et al. *Cancer Treat. Rev.*, 40(1):48–59, 2014.
[3] D. Graafen et al. *J. Thorac. Imaging*, 37(5):315–322, 2022.
[4] A. Pourmorteza et al. *Am. J. Neuroradiol.*, 38(12):2257–2263, 2017.
[5] J. Zhu and S. Penfold. *Med. Phys.*, 43(6Part1):2845–2854, 2016.
[6] V. Taasti et al. *Med. Phys.*, 45(11):5186–5196, 2018.
[7] O. Pickford Scienti et al. *IEEE TRPMS*, 5(4):485–492, 2021.
[8] S. Jan et al. *Physics in Medicine and Biology*, 49(19), 2004.
[9] COMSOL Multiphysics®. www.comsol.com.
[10] N. Shusharina and T. Bortfeld. *TCIA*, 2021.
[11] J. Jørgensen et al. *Philos. Trans. Royal Soc. A*, 379(2204), 2021.
[12] E. Papoutsellis et al. *Philos. Trans. Royal Soc. A*, 379(2204), 2021.
[13] F. Isensee et al. *Nat. Methods*, 18(2):203–211, 2020.
[14] R. Cropp et al. *J. Appl. Clin. Med. Phys.*, 14(6):338–349, 2013.
[15] M. Kattau et al. Presentation at *IEEE NSS MIC RTSD 2023*, 2023.

Towards Efficient Truncation Correction in Arbitrary Geometry: Residual Projections for Region Of Interest Algebraic Reconstruction

Alexander Jung, Andreas Maier and Gabriel Herl

***Abstract*— In the last years, there have been more and more flexible image acquisition systems like twin robot CT systems supporting arbitrary scan geometries and trajectories. Practical use cases are large objects with only small regions of interest. To reduce scan time, memory consumption and computation time, it is appropriate to scan and reconstruct not the complete object but only the region of interest. However, at the state of the art, arbitrary trajectories are challenging for analytic reconstruction and region of interests are challenging for iterative reconstruction as the complete equation system must be considered. Multigrid methods reduce the resolution outside the region of interest and thus both memory consumption and computation time, but they suffer from some disadvantages and require a minimum resolution to keep the equation system correct and to avoid artefacts. The presented approach avoids multiple grids by moving the residuals outside the region of interest from volume to projection space. These residual projections cover equation system inconsistencies and thus may help not only with region of interest reconstruction but also in other applications and use cases like artefact reduction.**

***Index Terms*— CT, Twin Robots, Algebraic Reconstruction, Region Of Interest (ROI), Multigrid, Residual Projections**

I. Introduction

After the approximation of physics and the discretization, the relationship between an object and its projections can be written as a linear equation system (1):

$$M\vec{X} = \vec{P} \tag{1}$$

The vector $\vec{P}$ contains the ray sum between the X-ray source and the X-ray detector pixel for each projection and for each detector pixel. The vector $\vec{X}$ contains the unknown volume voxels. The matrix M as system matrix contains the known weights or ray lengths for each volume voxel to project $\vec{X}$ on $\vec{P}$. "Reconstructing" is the task of solving (1) for $\vec{X}$. Typically, a solution is not exact but approximated e.g. for the least squares.

A. Jung is with Deggendorf Institute of Technology, Germany, and Friedrich Alexander University Erlangen-Nuremberg, Germany, (e-mail: alexander.jung@th-deg.de, alexander.jung@fau.de).

A. Maier is with Friedrich Alexander University Erlangen-Nuremberg, Germany, (e-mail: andreas.maier@fau.de).

G. Herl is with Deggendorf Institute of Technology, Germany, (e-mail: gabriel.herl@th-deg.de).

In 1917 Johann Radon formulated the forward and inverse Radon transformation [1]. This analytical approach to solve (1) can be implemented as a Filtered Back Projection (FBP) but is, at the state of the art, limited to specific acquisition geometries like circular or helical trajectories.

In 1937 Kaczmarz [2] discovered an iterative solver for linear equation systems that was rediscovered and used in 1970 by Gordon et al. for the Algebraic Reconstruction Technique (ART) [3], in 1971 by Gilbert for the Simultaneous Iterative Reconstruction Technique (SIRT) [4] and in 1984 by Andersen and Kak for the Simultaneous Algebraic Reconstruction Technique (SART) [5]. This can be implemented in three steps as a forward projection operation to simulate a projection, a difference operation of the measured and simulated projection and a back projection operation to update the volume voxels. As these steps are independent of the overall trajectory, algebraic reconstruction methods are suitable not only for circular or helical, but arbitrary trajectories. But, in contrast to the FBP and the requirements of twin robot CT systems [6][7], algebraic reconstruction methods are not suitable for the direct reconstruction of a region of interest (ROI) as shown in Fig. 1:

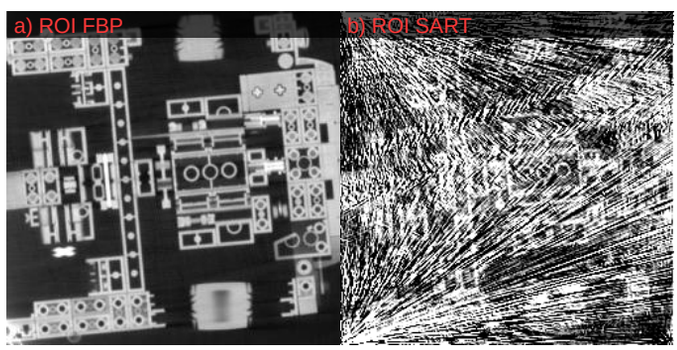


Fig. 1. Out-of-the-box reconstruction with FBP (left) and SART (right) of a region of interest. The left image is clear. The right image contains artefacts. All images show the attenuation range from -0.02/cm (black) to 0.2/cm (white).

The next Section II defines the problem of the region of interest algebraic reconstruction and the root cause of these artefacts. Section III outlines the state of the art and solutions how to do region of interest reconstructions without or at least reduced artefacts. In Section IV and V, the residual projection approach is introduced and first results are shown. Finally, Section VI gives a summary of the first results, the current and future work.

II. Problem Definition

For this and the next sections, M and $\vec{X}$ of (1) are decomposed into two submatrices and two subvectors: M_x and $\vec{x}$ for the weights and coefficients inside the ROI. M_r and $\vec{r}$ for the weights and coefficients outside the ROI. This results in (2):

$$M_x\vec{x} + M_r\vec{r} = \vec{P} \qquad (2)$$

In (1) and (2) the vector $\vec{P}$ contains the measured ray sum between the X-ray source and the X-ray detector pixel. Without a priori information about the object dimension and position, the worst-case scenario of an object covering the complete region between the source and the detector must be assumed. Thus, the reconstruction volume ($\vec{x}$ and $\vec{r}$) must cover the region covered by all source positions and detector corners (Fig. 2a).

If the reconstruction volume covers the region of interest only ($\vec{r}$ is assumed to be zero and omitted), there are two cases: If there is no part of the object exceeding the region of interest, then the assumption and the equation system is correct (Fig 2b). If there are parts of the object exceeding the region of interest, then the assumption and the equation system is wrong (Fig 2c): objects outside the reconstruction volume $\vec{x}$ are "reconstructed" into the reconstruction volume $\vec{x}$ causing the artefacts in Fig. 1.

With information about the object dimension and position, all coefficients outside the object can be assumed to be zero and omitted. Geometrically: The reconstruction volume ($\vec{x}$ and $\vec{r}$) can be minimized to the intersection of the object volume and the volume covering all source positions and detector corners. This case is outlined in Fig. 3. Compared to Fig. 2a, both the memory consumption and computation time can be reduced. Mostly, the object dimensions and positions are roughly known resulting in the use cases outlined in Fig. 3 and not Fig. 2a.

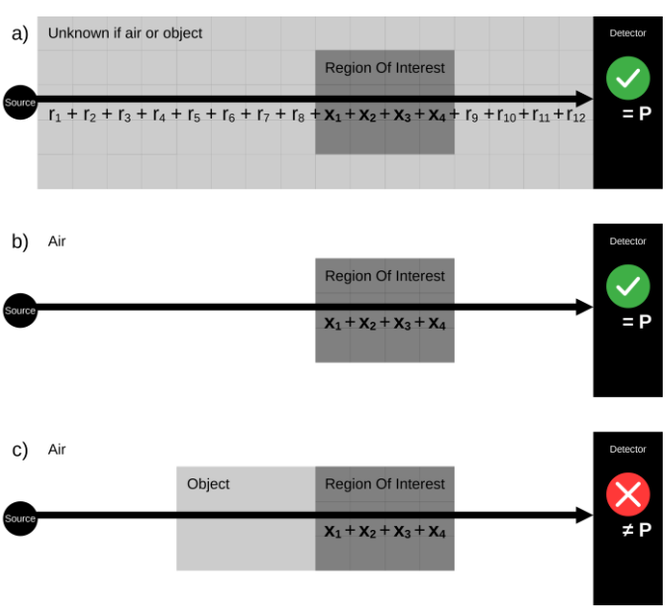


Fig. 2. Outline of one ray to one detector pixel with ***P*** as measured ray sum: For the cases a and b the outlined equation is correct. For the case c it is wrong.

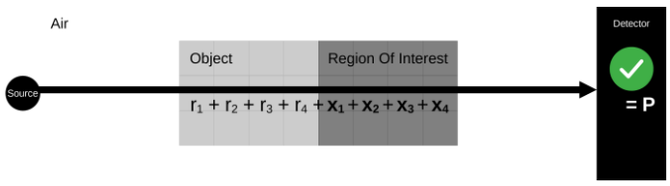


Fig. 3. Outline of one ray to one detector pixel with ***P*** as measured ray sum: With a priori information about the dimension and position of the object, all coefficients outside the region of interest and outside the object can be omitted.

III. State Of The Art

Fig. 3 directs to the basic approach for the region of interest reconstruction: The high-resolution reconstruction is first computed for the complete object at high memory and runtime costs and then cropped to the desired region of interest. More advanced algorithms reduce the costs by reducing the resolution outside the region of interest by using adaptive or multiple grids ("multigrid") outside the region of interest, as shown in Fig. 4.

Initially, the motivation for these methods were the unknown dimensions and positions of the object of interest, especially with fluids and gases measured with optical systems [10]. Later, these methods were adopted and refined for X-ray systems [11]. One of the main challenges is the choice of the grid resolutions as higher resolutions have more costs and lower resolutions show more artefacts. This resolution depends on the object: Regions with blurry objects like fluids and gases can be reconstructed with a lower resolution to reduce the costs. Regions with sharp objects like industrial parts must be reconstructed with a higher resolution to reduce the artefacts. Other challenges are the switching between resolutions itself, interpolation artefacts and moiré and aliasing artefacts caused by the different resolutions in different regions of the volume.

In [12], the reconstruction volume grid was replaced with sample points and radial basis functions allowing a smooth transition between different resolutions and different regions.

In [13], the FBP of the complete object was computed, the region of interest was zeroed out and the forward projections were subtracted from the measured projections resulting in projections of only the region of interest.

In [14], the Neural Adaptive Tomography uses an adaptive multigrid/octree-based approach, and one may expect steps towards efficient truncation correction also in deep-learning.

In [15] the idea was proposed to extend each projection pixel with a residual pixel to cover the inconsistencies between the measured projections like noise, beam hardening and scattering. Now, this idea is taken up and improved for the challenge of the region of interest reconstruction with algebraic reconstruction.

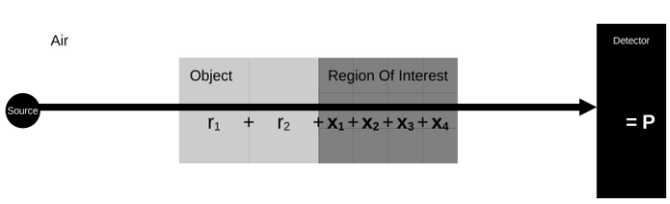


Fig. 4. Outline of one ray to one detector pixel with ***P*** as measured ray sum: The different object regions are sampled with different resolutions (multigrid).

IV. Residual Projections

As defined in Section II, the vectors $\vec{x}$ and $\vec{r}$ contain the attenuation coefficients inside and outside the region of interest respectively. (2) can be rearranged to (3):

$$M_r\vec{r} = \vec{P} - M_x\vec{x} \tag{3}$$

With (3), $M_r\vec{r}$ can be seen as the residuum of $M_x\vec{x} = \vec{P}$ with $\vec{x}$ as the region of interest volume and $\vec{r}$ as the residual volume.

The state-of-the-art approaches for the computation of a region of interest try to reduce the memory and computation costs by reducing the spatial resolution of the residual volume. The residual projection approach projects the residual volume from the volume space to the projection space and minimizes the number of residuals for each line of the equation system to exactly one. This is the minimum to keep the system correct:

$$M_r\vec{r} = \vec{p}_r \tag{4}$$

In (4), $\vec{p}_r$ is the residual projection of the residual volume $M_r\vec{r}$. In (3), $M_r\vec{r}$ can be replaced by $\vec{p}_r$, resulting in Fig. 5 and (5):

$$M_x\vec{x} + \vec{p}_r = \vec{P} \tag{5}$$

To transfer this approach from theory to practice a SART [5] implementation was adapted with residual projections giving the following pseudo code (SART in black and adaption in red):

$set\ \vec{x}\ to\ zero$
$set\ \vec{p}_r\ to\ zero$
$for\ each\ iteration$:
  $for\ each\ projection\ i$:
    // forward projection
    $\vec{f}_i = forwardproj_i(\vec{x}) + s_i(\vec{p}_r)$
    // difference
    $\vec{d}_i = s_i(\vec{P}) - \vec{f}_i$
    // update volume
    $\vec{x} = \vec{x} + \alpha \cdot (\beta) \cdot backproj_i(\vec{d}_i)$
    // update residual projection
    $s_i(\vec{p}_r) = s_i(\vec{p}_r) + \alpha \cdot (1 - \beta) \cdot \vec{d}_i$

$s_i(\vec{p}_r)$ and $s_i(\vec{P})$ are subvectors of $\vec{p}_r$ and $\vec{P}$ containing the pixels for projection i, respectively. $\vec{d}_i$ is the difference between the measured projection $s_i(\vec{P})$ and the simulated projection $\vec{f}_i$. α is the relaxation factor. β defines the distribution of $\vec{d}_i$ to the region of interest volume $\vec{x}$ and the residual projection $s_i(\vec{p}_r)$. As a starting point it was assumed that ray sums are distributed equally on the ray between the source and the detector pixel (6):

$$\beta = \frac{ray\ length\ in\ volume}{ray\ length\ from\ source\ to\ detector\ pixel} \tag{6}$$

This assumption was derived from the case Section II Fig. 2a without information about the object dimension and position.

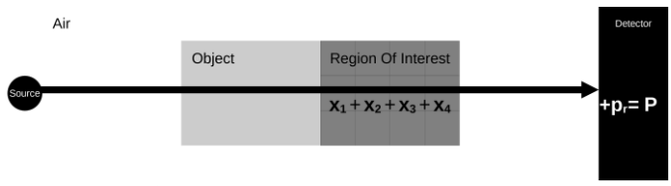


Fig. 5. Outline of one ray to one detector pixel with $\boldsymbol{P}$ as measured ray sum. $\boldsymbol{p}_r$ is a scalar element of $\vec{\boldsymbol{p}}_r$ and $\boldsymbol{P}$ is a scalar element of $\vec{\boldsymbol{P}}$. In the former images the residual was expressed as a volume. In Fig. 5 it is expressed as a projection.

V. First Results

Fig. 6 shows a LEGO 1967th Ford Mustang as an example object. 1600 circular projections were measured with the Deggendorf Institute of Technology twin robot CT system [8][9] containing a COMET XRS-225VF X-ray source and a VAREX IMAGING PaxScan 4343DX-I X-ray detector with 3072^2 pixels à 139^2 µm². The source was set to a focal point size of 300 µm, a voltage of 155 kV and a current of 2400 µA. The detector was set to an exposure time of 3 x 250 ms. The distance from the rotation axis to both the source and the detector was 1000 mm. The projections were down sampled from 3072^2 to 384^2 pixels à 1112^2 µm² for faster data handling. Fig. 7 shows a slice of a SART-reconstruction of the object after two iterations with a relaxation factor $\alpha = 0.5$ and a voxel size of ~0.55 mm.



Fig. 6. LEGO 1967th Ford Mustang. The red rectangle outlines the position of the slice containing the region of interest shown in Fig. 1, Fig. 7 and Fig. 8.

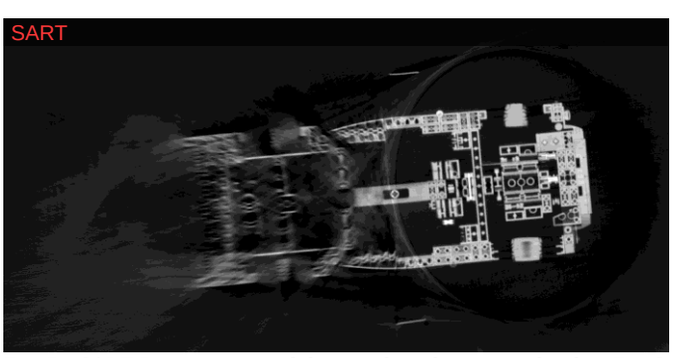


Fig. 7. SART-Reconstruction of the complete object as outlined in Fig. 3. On the left side the data is incomplete and shows limited angle artefacts. On the right side inside the circle the data is complete. The data was measured with the Deggendorf Institute of Technology twin robot CT system with two half circles. As the half circles were measured not parallel to the reconstructed slice but rotated by 25°, the shape of the region with complete data is rather an ellipse.

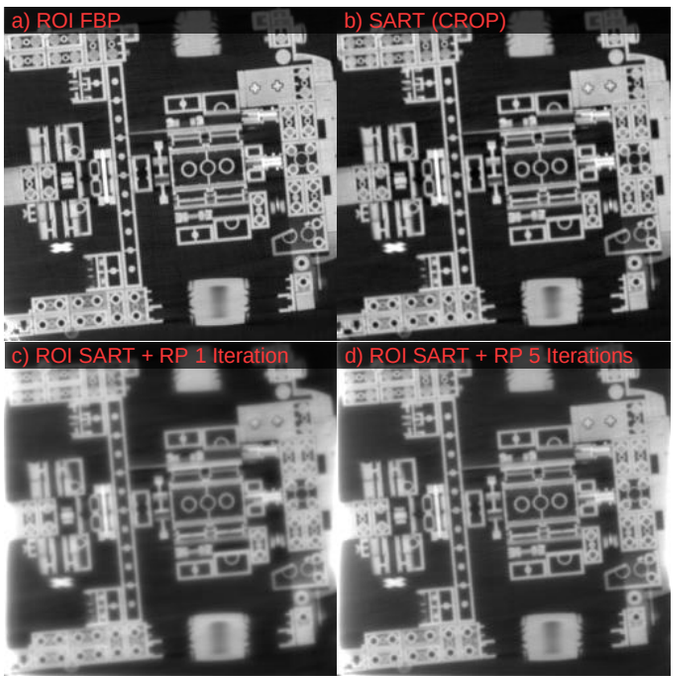


Fig. 8. a) FBP b) Crop of Fig. 7 c+d) Reconstruction of the region of interest with residual projections as outlined in Fig. 5 after c) one and d) five iterations. All images show the attenuation range from -0.02/cm (black) to 0.2/cm (white).

Fig. 8 shows a crop of the FBP- and SART-reconstruction (Fig. 7) for comparison and the SART-reconstruction with residual projections after one and five iterations (with $\alpha = 0.5$). The reconstruction quality in Fig. 7d is far better than in Fig. 1a (out-of-the-box SART reconstruction of the region of interest). Nevertheless, some blurriness and bright regions on the edges are visible and subject of future work. With more iterations the sharpness increases marginally but the bright edges still remain.

VI. Summary And Future Work

It was shown that the presented residual projection approach is working and converging. In contrast to the multigrid method as state of the art the residual projections are "gridless": They reuse the existing grid and do not introduce new resampling issues to the equation system. As both $\vec{p}_r$ and $\vec{P}$ have the same dimensions, also the memory consumption is predictable. Further evaluations are required regarding memory and runtime costs and quality. Finally, there are multiple discussion points:

For each detector pixel a residual pixel is introduced. Thus, the equation system is very underdetermined. Possible solutions are $M_r\vec{p}_r = \vec{P}$ with $\vec{x}$ only zeroes, $\vec{p}_r = 0$ with $\vec{x}$ only artefacts, the desired solution or something in between. More effort must be invested to understand and control the convergency to the desired solution. One option is to introduce a regularization or penalty term to control the distribution of the difference to the volume and the residual projection to "motivate" ray sums to be moved into the volume or the residual projection. Another option is to consider a priori information about the object, for example the object spatial extent or the object mean attenuation. Finally, more realistic estimates can be expected for β than the equal distribution that was assumed as the starting point in (6).

The residual projection approach moves inconsistencies not fitting to the reconstruction volume into residual projections. It's imaginable that this will work not only for truncated parts of the object but also for physical effects like noise, beam hardening and scattering. Thus, residual projections could help to reduce artefacts and to improve image quality in general.

The ART forward projection and back projection operators transform between volume and projection space. The residuals can be transformed between volume and projection space, too. This allows the combination with existing other reconstruction algorithms. For example, a reconstruction could start with the multigrid method and switch to the residual projection approach after some iterations by a forward projection of the residuals. This way it should be possible to get the best of the two worlds.

References

[1] J. Radon, "Über die Bestimmung von Funktionen längs gewisser Mannigfaltigkeiten", Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse, 69, 1917, pp. 262-277

[2] S. Kaczmarz, "Angenäherte Auflösung von Systemen linearer Gleichungen", Bulletin International de l'Académie Polonaise des Sciences et des Lettres, 35, 1937

[3] R. Gordon, R. Bender, G. T. Herman, "Algebraic Reconstruction Techniques (ART) for Three-dimensional Electron Microscopy and X-ray Photography", Journal of Theoretical Biology, 29, 1970, pp. 471-481

[4] P. Gilbert, "Iterative Methods for the Three-dimensional Reconstruction of an Object from Projections", Journal of Theoretical Biology, 36, 1972, pp. 105-117

[5] A. H. Andersen, A. C. Kak, "Simultaneous Algebraic Reconstruction Technique (SART): A superior implementation of the ART algorithm", Ultrasonic Imaging, 6, 1984, pp. 81-94

[6] W. Holub, U. Haßler, C. Schorr, M. Maisl, P. Janello, P. Jahnke, "XXL-Micro-CT Comparative Evaluation of Microscopic Computed Tomography for Macroscopic Objects", Digital Industrial Radiology and Computed Tomography (DIR2015), 2015

[7] A. Ziertmann, P. Jahnke, S. Kerscher, M. Koch, W. Holub, "Robot guided Computed Tomography", Journal of the Japan Society for Precision Engineering, vol. 86, no. 5. Japan Society for Precision Engineering, pp. 316-322, May 05, 2020. doi: 10.2493/jjspe.86.316.

[8] A. Weiss, S. Wittl, G. Herl, S. Zabler, "Simulated and experimental evaluation of the accuracy of twin robotic CT systems", Proceedings of the 12th International Conference on Industrial Computed Tomography (iCT2023), 2023

[9] S. Wittl, A. Weiss, G. Herl, S. Zabler, "Keep Attention to the Mapping: Application of AI for Geometric X-Ray CT Scan Calibration", Proceedings of the 12th International Conference on Industrial Computed Tomography (iCT2023), 2023

[10] I. Ihrke, M. Magnor, "Adaptive grid optical tomography", Graphical Models, vol. 68, no. 5-6. Elsevier BV, pp. 484-495, Sep. 2006. doi: 10.1016/j.gmod.2006.08.001.

[11] D. Marlevi, H. Kohr, J.-W. Buurlage, B. Gao, K. J. Batenburg, M. Colarieti-Tosti, "Multigrid Reconstruction in Tomographic Imaging", IEEE Transactions on Radiation and Plasma Medical Sciences, vol. 4, no. 3. Institute of Electrical and Electronics Engineers (IEEE), pp. 300-310, May 2020. doi: 10.1109/trpms.2019.2942186.

[12] J. Gregson, M. Krimerman, M. B. Hullin, W. Heidrich, "Stochastic tomography and its applications in 3D imaging of mixing fluids", ACM Transactions on Graphics, vol. 31, no. 4. Association for Computing Machinery (ACM), pp. 1-10, Jul. 2012. doi: 10.1145/2185520.2185548.

[13] F. K. Kopp et al., "Region of interest processing for iterative reconstruction in x-ray computed tomography", SPIE Proceedings. SPIE, Mar. 18, 2015. doi: 10.1117/12.2081911.

[14] D. Rückert, Y. Wang, R. Li, R. Idoughi, W. Heidrich, "NeAT", ACM Transactions on Graphics, vol. 41, no. 4. Association for Computing Machinery (ACM), pp. 1-13, Jul. 2022. doi: 10.1145/3528223.3530121.

[15] A. Jung, "Algebraische Rekonstruktionstechnik für die Industrielle Computertomographie", M.S. thesis, Dept. Computer Science, Univ. Erlangen-Nuremberg, Germany, 2013

PNIR: A Physics-Constrained and Noise-Controlled Iterative Reconstruction Method for Photon Counting Computed Tomography

Ruifeng Chen, Yang Chen, Dong Liang, Yinsheng Li

***Abstract*—Recently, there has been significant interest in using photon counting detectors for computed tomography (PCCT). Enhancing the quality of material basis images and the precision of elemental composition quantification are important in PCCT, especially when PCCT data is acquired at lower radiation dose levels. In this work, we introduce a physics-constrained and noise-controlled iterative reconstruction (PNIR in short) method to address challenges such as compromised image quality of material basis images and imprecise quantification of elemental composition. These challenges primarily arise from inadequate noise modeling and the reliance on hand-crafted regularization techniques, such as local smoothness or sparsity employed in current direct material basis image reconstruction methods. Contrastingly, PNIR adopts a distribution-level regularization approach, learning to characterize ideal material basis images through training a noise-controlled diffusion model. This method enables the sampling of optimal material basis images for each subject, guided by the constraints of the physical model specific to a given PCCT and the measured data acquired from the subject.**

***Index Terms*—Spectral Computed Tomography, Photon Counting Detector, Deep Learning, Diffusion Model.**

I. Introduction

SPECTRAL computed tomography (CT) emerged as a solution to a significant limitation in traditional CT technology: the dependence of the linear attenuation coefficient on both the intrinsic characteristic quantities of the scanned subject (i.e., elemental composition) and the extrinsic data acquisition conditions (i.e., the energy spectrum existed in a specific CT scanning protocol). Among various strategies for implementing spectral CT, the utilization of energy-resolving photon counting detectors in computed tomography (PCCT) has attracted extensive attention [1]. However, accurate quantification of elemental composition and acquisition of high-quality material basis images pose challenges when obtaining PCCT data at lower radiation dose levels.

Ruifeng Chen (e-mail: rf.chen1@siat.ac.cn) is with Research Center for Medical Artificial Intelligence, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, China and with Laboratory of Image Science and Technology, Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing, Jiangsu, China.

Yang Chen (e-mail: chenyang.list@seu.edu.cn) is with Laboratory of Image Science and Technology, Key Laboratory of Computer Network and Information Integration, Southeast University, Nanjing, Jiangsu, China, and with Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, China.

Dong Liang (e-mail: dong.liang@siat.ac.cn) and Yinsheng Li (e-mail: ys.li2@siat.ac.cn) are with Research Center for Medical Artificial Intelligence, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China and Key Laboratory of Biomedical Imaging Science and System, Chinese Academy of Sciences, Shenzhen 518055, China.

The existing approaches for quantifying elemental composition through PCCT can be widely classified into two categories. (i) Indirect model-based material basis image reconstruction approaches. One sub-category of these approaches involves performing data-domain material decomposition [2], followed by analytical or iterative material basis image reconstruction [3]. The other sub-category focuses on iterative image reconstruction [4], followed by the image-domain material decomposition [5]. These methods aim to alleviate errors caused by nonlinear and nonuniform energy response of PCCT through modeling its data acquisition process. (ii) Direct model-based material basis image reconstruction approaches [6]. These approaches aim to directly reconstruct material basis images from the measured PCCT data by exploiting the physical model of PCCT data acquisition and regularization of material basis images, resulting in potentially superior performance. However, direct reconstruction of material basis images is technically challenging due to ill-conditioned inversion and the nonlinear polyenergetic PCCT forward model. Moreover, it poses intellectual challenges in designing appropriate regularization methods that accurately capture the features of ideal material basis images.

In this work, we propose a physics-constrained and noise-controlled iterative reconstruction (PNIR in short) method to address the degraded quality of material basis images and inaccurate quantification of elemental composition. PNIR leverages the physical model of a given PCCT and the measured data obtained from the subject, and learns distribution-level regularization to describe the ideal feature of material basis images through training a noise-controlled diffusion model. Beside, by incorporating specifically designed covariances during training, the noise-controlled diffusion model can accurately predict realistic statistical characteristics across different interaction channels along the reverse diffusion process, thereby enabling more effective noise removal.

II. Methods

A. Forward Modeling of PCCT Data Acquisition Process

According to Polyenergetic Lambert-Beer's Law, the forward model describing the modeled number of photons $\bar{N}_{i,k}$ at the i-th X-ray interaction path and k-th energy bin, can be represented in the following form:

$$\bar{N}_{i,k}(\mathbf{C}) = \sum_{\varepsilon} w_{i,k,\varepsilon} \exp\left(-\sum_{k'} b_{k'\varepsilon} \sum_{j} a_{ij} c_{jk'}\right), \quad (1)$$

where ε denotes the digitized index to refer photon energies, $w_{i,k,\varepsilon}$ denotes the photon number at i-th X-ray interaction

path, k-th energy bin and ε-th spectral energy, a_{ij} denotes the matrix element of the PCCT system matrix, $\mathbf{A}_{M\times N}$, M denotes the total number of measurements for a given energy bin, N denotes the total number of image voxels, and $\bar{N}_{i,k}$ denotes a function of the basis-material-equivalent concentration matrix to be reconstructed, $\mathbf{C} = \{c_{jk'}\} \in \mathbb{R}^{N\times K}$, K denotes the number of energy bins.

B. Material Basis Image Reconstruction Using Maximum A Posterior Estimation

Due to the intrinsic nature of photon statistics, the measured number of photons $\bar{N}_{i,k}$ can be regarded as a random variable that follows the Poisson distribution that takes $\bar{N}_{i,k}$ as the mean value. And the probability of $N_{i,k}$ can be represented as the conditional probability value:

$$P(N_{i,k}|\bar{N}_{i,k}) = \frac{\bar{N}_{i,k}^{N_{i,k}} \exp^{-\bar{N}_{i,k}}}{N_{i,k}!}. \tag{2}$$

Note that the above conditional probability value is also a function of $\mathbf{C}$ according to Eq. (1).

The elements in $\{\bar{N}_{i,k}\}$ can be considered to be independent and identically distributed random variables. Consequently, the joint conditional probability of the random variable representing the measured number of photons is then expressed as:

$$P(\mathbf{N}|\bar{\mathbf{N}}) = \prod_{i=1}^{M}\prod_{k=1}^{K} P(N_{i,k}|\bar{N}_{i,k}). \tag{3}$$

Using Bayesian's formula, the posterior probability value $P(\bar{\mathbf{N}}|\mathbf{N})$ is converted to the conditional probability value $P(\mathbf{N}|\bar{\mathbf{N}})$ and the prior probability value $P(\bar{\mathbf{N}})$. Note that, both $P(\bar{\mathbf{N}}|\mathbf{N})$ and $P(\bar{\mathbf{N}})$ are functions of $\mathbf{C}$, hence, we formulate the following maximum a posterior problem:

$$\begin{aligned}\mathbf{C}^* &= \arg\max_{\mathbf{C}} \log P(\bar{\mathbf{N}}|\mathbf{N}) \\ &= \arg\max_{\mathbf{C}} \left\{\log P(\mathbf{N}|\bar{\mathbf{N}}) + \log P(\bar{\mathbf{N}})\right\} \\ &= \arg\min_{\mathbf{C}} \left\{ \left[\sum_i\sum_k \bar{N}_{i,k} - N_{i,k}\log \bar{N}_{i,k}\right] - \log P(\bar{\mathbf{N}})\right\} \\ &:= \arg\min_{\mathbf{C}} \left\{L(\mathbf{C}) - \log P(\mathbf{C})\right\}.\end{aligned} \tag{4}$$

The generation of high-quality and accurate material basis images via solving the optimization problem defined in Eq. (4) can be challenging due to the following two reasons. Firstly, it is technically challenging to solve the above problem because $L(\mathbf{C})$ in Eq. (4) is nonlinear and non-convex with respect to $\mathbf{C}$. Secondly, it is intellectually challenging to hand-craft the prior probability function $P(\mathbf{C})$ to properly describe the feature of ideal material basis images.

C. Separable Paraboloidal Surrogate of $L(\mathbf{C})$ via Optimization Transfer Principle

To tackle the first technical challenge, we apply the optimization transfer principle [7] and derive a separable paraboloidal surrogate (SPS) [8] of $L(\mathbf{C})$, such that, the minimization of the surrogate $Q(\mathbf{C};\mathbf{C}_t)$ asymptotically approximates the minimization of $L(\mathbf{C})$ defined in Eq. (4).

Based on the optimization transfer principle, $Q(\mathbf{C};\mathbf{C}_t)$ can be represented as:

$$\begin{aligned}&Q(\mathbf{C};\mathbf{C}_t) \\ &= \sum_i\sum_k\sum_\varepsilon \frac{w_{ik\varepsilon}}{d_{ik\varepsilon,t}} \\ &\sum_j \gamma_{ij} q_{ik\varepsilon,t}\left(\left\{\frac{b_{k'\varepsilon}a_{ij}}{\gamma_{ij}}(c_{jk'} - c_{jk',t}) + b_{k'\varepsilon}a_{ij}c_{jk',t}\right\}_{k'=1}^{K}\right),\end{aligned} \tag{5}$$

where $d_{ik\varepsilon,t}$, γ_{ij}, and the real-valued quadratic function $q_{ik\varepsilon,t}(\cdot)$ are defined in Eq. 13, Eq. 21 and Eq. 18 respectively in the paper [9]. $c_{jk',t}$ denotes the basis-material-equivalent concentration corresponding to the j-th voxel, k'-th basis material evaluated at the t-th iteration.

D. Learning $\log P(\mathbf{C})$ via Training a Noise-Controlled Diffusion Model Using Sampled Noise Labels with Specifically Designed Covariances

To tackle the second intellectual challenge, PNIR learns distribution-level prior knowledge, $\log P(\mathbf{C})$, to describe the feature of ideal material basis images. The deep diffusion models [10] are employed to learn the needed prior distribution.

To better describe the noise anti-correlation property [11] among material basis images, we generalize the traditional diffusion process to learn prior joint distribution of random variables in basis material dimension, which can be described as:

$$\mathbf{C}_{j,t} = \sqrt{\alpha_t}\mathbf{C}_{j,t-1} + \sqrt{1-\alpha_t}\Delta_j, \tag{6}$$

where $\mathbf{C}_{j,t} \in \mathbb{R}^{1\times K}$ denotes the basis-material-equivalent concentration corresponding to the j-th image voxel at t-th time step, α_t denotes the scaling coefficients controlling the amplitude of added noise at t-th step [12], and a Gaussian random variable, $\Delta_j \sim \mathcal{N}(\mathbf{0}_{1\times K}, \mathbf{S}_j)$, denotes the sampled noise corresponding to the j-th image voxel and $\mathbf{S}_j \in \mathbb{R}^{K\times K}$ denotes its covariance, which is defined in Eq. 8. Subsequently, defining the diffusion model $\mathcal{G}_{\mathbf{\Theta}}(\mathbf{C}_t, t)$ to approximate the scaled score $(\hat{\alpha}_t - 1)\left[\nabla_{\mathbf{C}}\log P_t(\mathbf{C})\right]_j \mathbf{S}_j$, the model parameter $\mathbf{\Theta}^*$ is optimized by score matching:

$$\mathbf{\Theta}^* = \arg\min_{\mathbf{\Theta}} \sum_{s,j,t}\left\{\lambda_{tj}\left|\left|\left[\mathcal{G}_{\mathbf{\Theta}}(\mathbf{C}_{s,t},t)\right]_j - \sqrt{1-\hat{\alpha}_t}\Delta_{s,j}\right|\right|_2^2\right\} \tag{7}$$

where $\nabla_{\mathbf{C}}\log P_t(\mathbf{C}) \in \mathbb{R}^{N\times K}$ denotes the gradient of the prior distribution to be approximated at t-th time step along the reverse diffusion process, $\left[\nabla_{\mathbf{C}}\log P_t(\mathbf{C})\right]_j \in \mathbb{R}^{1\times K}$ denotes the corresponding quantity at j-th image voxel, $\mathcal{G}_{\mathbf{\Theta}}(\mathbf{C}_{s,t},t) \in \mathbb{R}^{N\times K}$ denotes the output of the generalized diffusion model at t-th time step parameterized by $\mathbf{\Theta}$, $\left[\mathcal{G}_{\mathbf{\Theta}}(\mathbf{C}_{s,t},t)\right]_j \in \mathbb{R}^{1\times K}$ denotes the corresponding basis-material-equivalent concentration values at j-th image voxel, $||\cdot||_2$ denotes the L_2 norm, $\hat{\alpha}_t = \prod_{\tau=1}^{t}\alpha_\tau$, and s denotes the index of a finite set of ideal material basis images.

Besides, the diffusion model $\mathcal{G}_{\mathbf{\Theta}}(\cdot,\cdot)$ is designed as a U-shaped convolutional network architecture showed in Fig. 1.

As for PCCT material basis image reconstruction problem, the noise variance within each image voxel and the noise

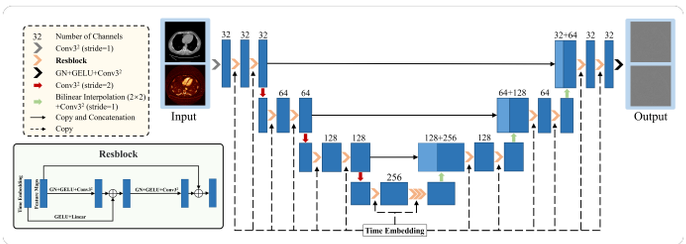


Fig. 1. The detailed architecture of the deep diffusion model.

covariance between any pair of material basis images can be specifically determined for each individual subject scanned with given data acquisition conditions, which can be formulated as:

$$\mathbf{S}_j = \lambda \begin{bmatrix} \sum_k \tilde{b}_{k1}^2 \sigma_{jk}^2 & \sum_k \tilde{b}_{k1}\tilde{b}_{k2}\sigma_{jk}^2 & \cdots & \sum_k \tilde{b}_{k1}\tilde{b}_{kK}\sigma_{jk}^2 \\ \sum_k \tilde{b}_{k2}\tilde{b}_{k1}\sigma_{jk}^2 & \sum_k \tilde{b}_{k2}^2 \sigma_{jk}^2 & \cdots & \sum_k \tilde{b}_{k2}\tilde{b}_{kK}\sigma_{jk}^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sum_k \tilde{b}_{kK}\tilde{b}_{k1}\sigma_{jk}^2 & \sum_k \tilde{b}_{kK}\tilde{b}_{k2}\sigma_{jk}^2 & \cdots & \sum_k \tilde{b}_{kK}^2 \sigma_{jk}^2 \end{bmatrix}, \tag{8}$$

where $\lambda = 1/\max(\{\tilde{\mathbf{S}}_j\})$, $\tilde{\mathbf{S}}_j$ denotes the covariance of the basis-material-equivalent concentration corresponding to the j-th image voxel, $\{\tilde{b}_{kk'}\}$ denotes the k, k'-th element in the inverse of the matrix $\{b_{k'}(\bar{E}_k)\}$, $b_{k'}(\bar{E}_k)$ denotes the mass attenuation value of k'-th basis material at the effective energy of k-th energy bin, σ_{jk}^2 denotes the noise variance measured on FBP image corresponding to the j-th image voxel and k-th energy bin.

The proposed noise-controlled diffusion model effectively captures the statistical characteristics of noise across diverse interaction channels. When incorporated as the regularization term $\log P(\mathbf{C})$ in Eq. 4, it enables significantly enhanced noise removal capabilities.

E. Numerical Optimization Algorithm

The optimization problem Eq. 4 can be effectively solved via the following two-step alternating minimization strategy [13].

$$\mathbf{C}_{t-\frac{1}{2}} = \arg\min_{\mathbf{C}} Q(\mathbf{C}; \mathbf{C}_t), \tag{9}$$

$$\mathbf{C}_{t-1} = \arg\min_{\mathbf{C}} \left\{ \frac{1}{2} \left|\left| \mathbf{C} - \mathbf{C}_{t-\frac{1}{2}} \right|\right|_F^2 - \log P(\mathbf{C}) \right\}, \tag{10}$$

where and $||\cdot||_F$ denotes the Frobenius norm. The first sub-problem Eq. (9) can be solved using Newton's Method:

$$\begin{aligned} \mathbf{C}_{j,(t-\frac{1}{2})} &= \mathbf{C}_{j,t} - \frac{\left[\frac{\partial Q(\mathbf{C};\mathbf{C}_t)}{\partial \mathbf{C}_j}\right]_{\mathbf{C}=\mathbf{C}_t}}{\left[\frac{\partial^2 Q(\mathbf{C};\mathbf{C}_t)}{\partial \mathbf{C}_j^2}\right]_{\mathbf{C}=\mathbf{C}_t}} \\ &:= \mathbf{C}_{j,t} - \left[\mathcal{H}(\mathbf{C}_t)\right]_j, \end{aligned} \tag{11}$$

where $\frac{\partial Q(\mathbf{C};\mathbf{C}_t)}{\partial \mathbf{C}_j}$ and $\frac{\partial^2 Q(\mathbf{C};\mathbf{C}_t)}{\partial \mathbf{C}_j^2}$ are defined in Eq. 26 and Eq. 28 in the paper [9]. The second sub-problem can be solved by the following updating sequence:

$$\mathbf{C}_{t-1} = \mathbf{C}_{t-\frac{1}{2}} - \eta_t \mathcal{G}_{\mathbf{\Theta}^*}\left(\sqrt{\hat{\alpha}_t}\mathbf{C}_{t-\frac{1}{2}}, t\right), \tag{12}$$

$$\eta_t = \frac{1}{\sqrt{\hat{\alpha}_t}} - \sqrt{\frac{1-\hat{\alpha}_{t-1}}{(1-\hat{\alpha}_t)\hat{\alpha}_{t-1}}}. \tag{13}$$

And the whole iterative reconstruction process is shown in Algorithm 1.

Algorithm 1 PNIR

Require: $\mathbf{N}, n_1, n_2, \text{RSP}, \{\hat{\alpha}_t\}_{t=1}^T, \mathcal{F}(\cdot), \mathcal{H}(\cdot), \mathcal{G}_{\mathbf{\Theta}^*}(\cdot,\cdot)$

$\mathbf{C} \leftarrow \mathcal{F}(\mathbf{N})$
for $i = 1, 2, \cdots, n_1$ **do**
  $\mathbf{C} \leftarrow \mathbf{C} - \mathcal{H}(\mathbf{C})$
end for
for $t = \text{RSP}, \text{RSP}-1, \cdots, 1$ **do**
  for $i = 1, 2, \cdots, n_2$ **do**
    $\mathbf{C} \leftarrow \mathbf{C} - \mathcal{H}(\mathbf{C})$
  end for
  $\mathbf{C} \leftarrow \mathbf{C} - \eta_t \mathcal{G}_{\mathbf{\Theta}^*}(\sqrt{\hat{\alpha}_t}\mathbf{C}, t)$
end for
return $\mathbf{C}$

F. Experimental Setup

The numerical simulation studies are performed to verify the superiority of PNIR. A numerically simulated PCCT data set was generated by collecting clinical water-iodine material basis image pairs. The simulated high-energy and low-energy sinogram data were then generated according to the polyenergetic Beer-Lambert formula with a pair of simulated both high and low energy spectra. The entire data set was randomly split into a training set (2627 image samples) and a validation set (469 image samples). The overall reconstruction accuracy is quantified with relative root mean square error (RRMSE) and the overall image quality is assessed with structural similarity index (SSIM).

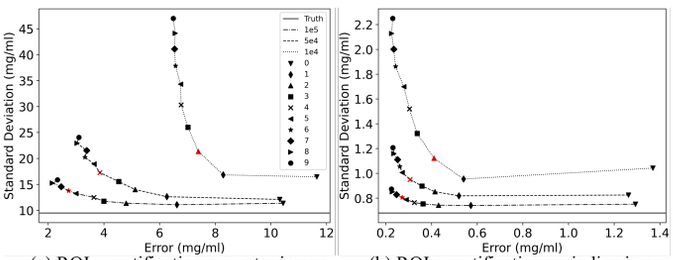
III. Results

A. Performance Dependence on Regularization Strength (n_2)

The regularization strength manipulating the trade-off between the noise amplitude measured in the reconstructed material basis images and the reconstruction accuracy is controlled by the hyperparameter n_2 in the Algorithm 1. Fig. 2 shows the trade-off characteristic curves measured on the PNIR material basis images using different n_2 values and hence facilitates the optimization of n_2 for the exposure levels with $I_0 = 1 \times 10^5$, 5×10^4 and 1×10^4.

B. Qualitative Image Quality Assessments and Comparisons

As shown in Fig. 3, the ablation studies are performed on exposure level $I_0 = 5 \times 10^4$ including image-domain material decomposition (ID), image-domain material decomposition followed by a noise-controlled diffusion model based



(a) ROI quantification on water image (b) ROI quantification on iodine image

Fig. 2. Trade-off characteristic curves measured on the PNIR material basis images using different n_2 values. For the exposure levels with $I_0 = 1 \times 10^5$, 5×10^4 and 1×10^4, the optimal n_2 value is selected, such that, the needed reconstruction accuracy is achieved while the noise amplitude is minimized, and labeled in red.

image denoising (ID+Diffusion), model-based polyenergetic image reconstruction using the SPS solver with total-variation regularization (polyMBIR+TV), and PNIR. The PNIR images demonstrate significantly reduced noise amplitude, a natural noise texture that closely resembles the ground truth image, and nearly accurate quantification of elemental composition.

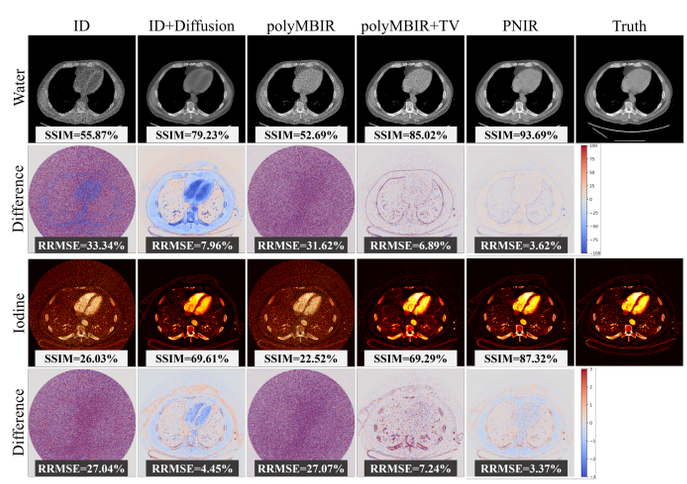


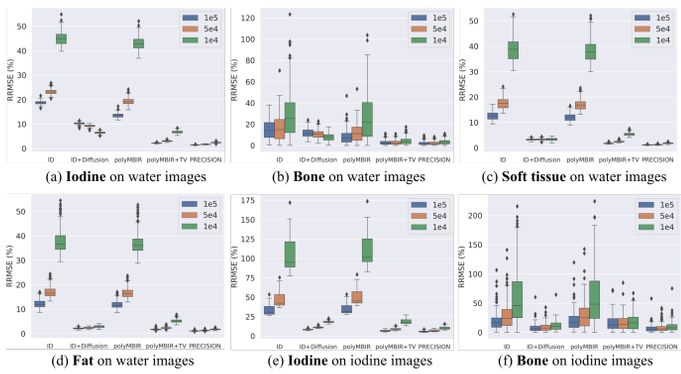
Fig. 3. Numerical phantom images generated from ground truth, ID, ID+Diffusion, polyMBIR+TV and PNIR. The SSIM and RRMSE for these images are calculated.

C. Quantitative Image Quality Assessments and Comparisons

The comparison of reconstruction error quantified by RRMSE using those five methods is shown in Fig. 4. A total of 469 testing samples are included in the statistics. The median [25th, 75th percentiles] of RRMSE are measured for different materials in the material basis images for the exposure levels with $I_0 = 1 \times 10^5$, 5×10^4 and 1×10^4 respectively.

IV. Conclusion

In this work, it has been demonstrated that the utilization of PNIR can effectively address inaccurate quantification of elemental composition and degraded quality of material basis images caused by imperfect noise modeling and hand-crafted regularization. The numerical simulation studies indicate that PNIR has the potential to enhance the quantitative accuracy of



(a) **Iodine** on water images (b) **Bone** on water images (c) **Soft tissue** on water images

(d) **Fat** on water images (e) **Iodine** on iodine images (f) **Bone** on iodine images

Fig. 4. The comparison of reconstruction error quantified by RRMSE using ID, ID+Diffusion, polyMBIR+TV and PNIR.

elemental composition, reduce noise amplitude, and improve image quality for PCCT systems.

References

[1] S. Leng, W. Zhou, Z. Yu, A. Halaweish, B. Krauss, B. Schmidt, L. Yu, S. Kappler, and C. McCollough, "Spectral performance of a whole-body research photon counting detector ct: quantitative accuracy in derived image sets," *Physics in Medicine & Biology*, vol. 62, no. 17, p. 7216, 2017.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Physics in Medicine & Biology*, vol. 21, no. 5, p. 733, 1976.

[3] J. Schlomka, E. Roessl, R. Dorscheid, S. Dill, G. Martens, T. Istel, C. Bäumer, C. Herrmann, R. Steadman, G. Zeitler *et al.*, "Experimental feasibility of multi-energy photon-counting k-edge imaging in pre-clinical computed tomography," *Physics in Medicine & Biology*, vol. 53, no. 15, p. 4031, 2008.

[4] H. Ding, H. Gao, B. Zhao, H.-M. Cho, and S. Molloi, "A high-resolution photon-counting breast ct system with tensor-framelet based iterative image reconstruction for radiation dose reduction," *Physics in Medicine & Biology*, vol. 59, no. 20, p. 6005, 2014.

[5] C. Maaß, M. Baer, and M. Kachelrieß, "Image-based dual energy ct using optimized precorrection functions: A practical new approach of material decomposition in image domain," *Medical physics*, vol. 36, no. 8, pp. 3818–3829, 2009.

[6] B. Zhao, H. Ding, Y. Lu, G. Wang, J. Zhao, and S. Molloi, "Dual-dictionary learning-based iterative image reconstruction for spectral computed tomography application," *Physics in Medicine & Biology*, vol. 57, no. 24, p. 8217, 2012.

[7] A. R. De Pierro, "On the relation between the isra and the em algorithm for positron emission tomography," *IEEE transactions on Medical Imaging*, vol. 12, no. 2, pp. 328–333, 1993.

[8] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic x-ray computed tomography," *IEEE transactions on medical imaging*, vol. 21, no. 2, pp. 89–99, 2002.

[9] T. Weidinger, T. M. Buzug, T. Flohr, S. Kappler, K. Stierstorfer *et al.*, "Polychromatic iterative statistical material image reconstruction for photon-counting computed tomography," *International journal of biomedical imaging*, vol. 2016, 2016.

[10] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.

[11] W. A. Kalender, E. Klotz, and L. Kostaridou, "An algorithm for noise suppression in dual energy ct material density images," *IEEE transactions on medical imaging*, vol. 7, no. 3, pp. 218–224, 1988.

[12] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[13] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale modeling & simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.

Data-Driven Filter Design in FBP: Transforming CT Reconstruction with Trainable Fourier Series

Yipeng Sun, Linda-Sophie Schneider, Fuxin Fan, Mareike Thies, Mingxuan Gu, Siyuan Mei, Yuzhong Zhou, Siming Bayer, Andreas Maier, *Member, IEEE*

***Abstract*—In this study, we introduce a Fourier series-based trainable filter for computed tomography (CT) reconstruction within the filtered backprojection (FBP) framework. This method overcomes the limitation in noise reduction by optimizing Fourier series coefficients to construct the filter, maintaining computational efficiency with minimal increment for the trainable parameters compared to other deep learning frameworks. Additionally, we propose Gaussian edge-enhanced (GEE) loss function that prioritizes the L_1 norm of high-frequency magnitudes, effectively countering the blurring problems prevalent in mean squared error (MSE) approaches. The model's foundation in the FBP algorithm ensures excellent interpretability, as it relies on a data-driven filter with all other parameters derived through rigorous mathematical procedures. Designed as a plug-and-play solution, our Fourier series-based filter can be easily integrated into existing CT reconstruction models, making it an adaptable tool for a wide range of practical applications. Code and data are available at https://github.com/sypsyp97/Trainable-Fourier-Series.**

***Index Terms*—Computed Tomography, Deep Learning, Filtered Backprojection, Loss Function**

I. Introduction

X-RAY computed tomography (CT) plays a crucial role in internal structure visualization, finding widespread use in fields such as medical imaging analysis and materials science. Traditional filtered backprojection (FBP) methods, which employ fixed analytical filters like Ram-Lak [1] and Shepp-Logan [2], are renowned for their computational efficiency and straightforward CT reconstruction. However, their effectiveness decreases when dealing with noise and artifacts, a shortcoming arising from the assumption of noise-free projections [3].

To overcome the limitations of traditional FBP, recent work proposed data-driven filter kernels. Syben et al. introduced the precision learning-based Ramp filter [4], which represents an improvement in this direction. By optimizing discrete filter kernels via back propagation, it reduces the dependence on manual filter design. However, the robustness of this approach to noise is constrained, due to its training on relatively simplistic phantom datasets.

To further advance the field, Xu et al. developed resFBP [5], integrating ResNet blocks [6] into the filtering process to improve noise resilience. While this approach marks a progress, resFBP faces challenges in terms of scalability, mainly because of its dependence on fixed input sizes, which complicates the handling of higher-resolution sinograms.

All authors are with the Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany (e-mail: yipeng.sun@fau.de)

Our work addresses these issues by introducing a Fourier series-based trainable filter within the FBP framework. Operating in the Fourier domain, this method constructs the FBP filter through the Fourier series with trainable series coefficients. This approach ensures robust performance across different resolution scales while maintaining computational efficiency and a fixed number of parameters.

Moreover, our methodology tackles the issue of image blurring commonly seen in deep learning approach with mean squared error (MSE) loss functions, particularly in noisy scenarios. We propose a novel loss function that is based on a Gaussian high-pass filter in the frequency domain. Specifically, we focus on the L_1 distance between the high-frequency amplitudes of the target and the output. This approach is particularly effective in preserving high-frequency details in reconstructed images.

Our trainable filter, employing Fourier series analysis in conjunction with an advanced loss function, effectively improves the sharpness of reconstructed images. Additionally, this approach improves the quality of the reconstructed images in the presence of noise. This makes it highly suitable for a wide range of practical CT imaging applications. Our method is intentionally designed as a plug-and-play solution. This design facilitates straightforward integration into more complex models that take sinograms as inputs and reconstructed phantoms as outputs. This adaptability allows for the effortless enhancement and refinement of existing CT reconstruction frameworks.

II. Method & Materials

First, we explain the theoretical basis of filtered backprojection in parallel beam geometry. Subsequently, we will explore the theoretical framework for constructing a discrete filter kernel with trainable Fourier series coefficients. Finally, we will introduce the gradient variance (GV) loss [7] and our approach, which incorporates the Gaussian edge-enhanced (GEE) loss based on a Gaussian high-pass filter.

A. Filtered Backprojection

1) Projection Filtering: The filtering step in the FBP algorithm is essential for reducing low-frequency components in the projection, as these components can accumulate and cause blurring in the reconstructed image. This process is formulated as follows

$$p_f(\theta, s) = p(\theta, s) * h(s). \tag{1}$$

where $p(\theta, s)$ represents the projection data at angle θ and detector position s, within the real number space $\mathbb{R}^{1\times N}$, where N represents the width of detector. The filter kernel $h(s)$, such as the Ram-Lak filter, also exists in $\mathbb{R}^{1\times N}$. Operator $*$ denotes convolution.

The filtering process can also be described in the Fourier domain

$$p_f(\theta, s) = \mathcal{F}^{-1}\left\{\mathcal{F}\{p(\theta, s)\} \cdot H(s)\right\}. \tag{2}$$

Here, $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the 1D Fourier and inverse Fourier transforms, respectively. The projection data $p(\theta, s)$ and the Fourier transform of the filter kernel $H(s)$, both belong to $\mathbb{R}^{1\times N}$.

The filtered projections $p_f(\theta, s)$ are obtained by convolving the original projections $p(\theta, s)$ with the filter kernel $h(s)$ or by multiplying the Fourier transform of $p(\theta, s)$ with $H(s)$ in the Fourier domain.

2) Backprojection: Following filtering, the backprojection step reconstructs the image from the filtered projections in a discrete manner

$$f(x, y) = \sum_{i=0}^{M-1} p_f(\theta_i, s_{xy})\Delta\theta. \tag{3}$$

In this equation, M represents the total number of discrete angles, $s_{xy} = x\cos(\theta_i) + y\sin(\theta_i)$ is the discrete detector coordinate for image coordinates x, y and projection angle θ_i. This step involves summing up the filtered projection data p_f for each discrete angle θ_i to create the reconstructed image $f(x, y)$. Here, $p_f(\theta_i, s_{xy})$ refers to the value of the filtered projection at angle θ_i and the detector position s_{xy} corresponding to the image coordinates x, y. $\Delta\theta$ is the angular increment.

B. Trainable Fourier Series Filter

1) Sinogram Filtering: For efficiency, in practical applications, the filtering process is performed on the entire sinogram

$$P_f = \mathcal{F}^{-1}\left\{\mathcal{F}\{P\} \odot H\right\}. \tag{4}$$

In this context, $\mathcal{F}$ and $\mathcal{F}^{-1}$ remain 1D, performing only along s direction. P is the sinogram within $\mathbb{R}^{M\times N}$, and H is the filtering matrix, also in $\mathbb{R}^{M\times N}$. The matrix H is a row-repeated matrix, where each row corresponds to the filter kernel $H(s)$ as defined in Equation (2). The element-wise multiplication in the Fourier domain is indicated by $\odot$. Each row of the sinogram P corresponds to $p(\theta, s)$ for a fixed θ, and the filtered sinogram P_f contains rows that correspond to $p_f(\theta, s)$ for each θ. Thus, P_f is the collection of all $p_f(\theta, s)$ over θ.

2) Deep Learning Computed Tomography: Expanding on the theoretical foundation, according to prior work, the FBP algorithm can be interpreted as a neural network with fixed weights [8]. This analogy allows us to express FBP in matrix multiplication form

$$I_{rec} = ReLU(A^{-1}P_f) = ReLU(A^{-1}F^{-1}HFP). \tag{5}$$

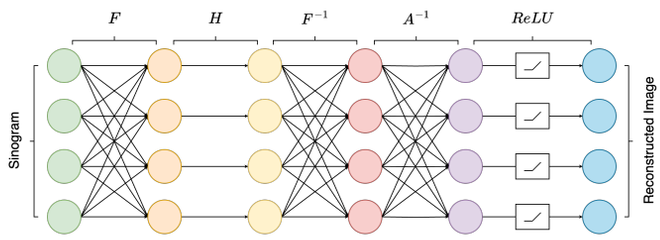


Fig. 1. Deep learning computed tomography: Reconstruction network for $I_{rec} = ReLU(A^{-1}F^{-1}HFP)$ from sinogram P to image I_{rec}.

In this formulation, A^{-1} is the backprojection, implemented by the backprojection operator from PYRO-NN [9]. The terms F^{-1} and F denote the inverse Fourier transform and Fourier transform, respectively, executed using PyTorch's Fourier operators, and $ReLU$ suppresses negative values in the final reconstruction [8]. The analytic differentiable implementation prevents the need for explicit storage of matrices A^{-1}, F^{-1}, and F, thus conserving memory. Furthermore, aligning with the concept of known operator learning, Equation (5) can be translated into a neural network structure, as it only consists of a diagonal and fully connected matrix, which are shown in Figure 1.

3) Fourier Series Filter Construction: Given that H is a row-repeated matrix in $\mathbb{R}^{M\times N}$, it can be simplified to a diagonal matrix K in $\mathbb{R}^{N\times N}$. This reduction is achieved through the transformation $K = TH$, where T is a known transformation matrix within $\mathbb{R}^{N\times M}$ designed to condense H into its diagonal form K. This process leverages the repetitive structure of H for more efficient computation and representation. We can then construct this filter using the Fourier series

$$k_{diag} = [k_1, k_2, \ldots, k_N]. \tag{6}$$

Here, k_{diag} represents the array of diagonal elements from the K matrix, encapsulating the essential components in a simplified form

$$k_{fourier} = a_0 + \sum_{l=1}^{L} a_l\cos(2\pi l\omega) + b_l\sin(2\pi l\omega). \tag{7}$$

The equation for $k_{fourier}$ represents the Fourier series reconstruction of the diagonal elements in k_{diag}. Here, a_0 is a constant term, and a_l and b_l are the Fourier coefficients for the cosine and sine components, respectively. The variable ω denotes the frequency component. This approach effectively reconstructs the original signal by summing the constant term and the infinite series of cosine and sine terms, each multiplied by their corresponding Fourier coefficients.

We can easily construct a new diagonal matrix K_{rec} from $k_{fourier}$, such that $K_{rec} \in \mathbb{R}^{N\times N}$. Therefore, Equation (5) can be reformulated as follows

$$I_{rec} = ReLU(A^{-1}F^{-1}T^{-1}K_{rec}FP). \tag{8}$$

In this configuration, T^{-1} is computed as the inverse of the matrix T. The Fourier series coefficients in Equation (7)

are designed as trainable parameters. The gradients of these parameters can be computed automatically, enabling dynamic adjustments to the reconstructed diagonal matrix K_{rec}. We restrict the Fourier series to its initial 50 orders to form K_{rec}, with $L = 50$. This constraint guarantees that the model maintains a fixed set of 101 trainable parameters, regardless of the resolution of the input sinogram.

C. Gradient Variance Loss

GV loss is designed to overcome the limitations of MSE in deep learning-based image reconstruction. Traditional MSE-supervised methods often leads to image blurring due the averaging pixel intensities. In contrast, GV loss focuses on improving the sharpness of edges in reconstructed images, denoted as I_{rec}, by leveraging the variance of gradient maps [7]. We calculate the gradient maps $G_x^{I_{rec}}$, $G_y^{I_{rec}}$, $G_x^{I_g}$, and $G_y^{I_g}$ for both I_{rec} and the ground truth image I_g, using Sobel operators. These maps are then divided into $n \times n$ non-overlapping patches to form matrices $\tilde{G}_x^{I_{rec}}$, $\tilde{G}_y^{I_{rec}}$, $\tilde{G}_x^{I_g}$, and $\tilde{G}_y^{I_g}$, each of size $n^2 \times \frac{w \cdot h}{n^2}$. The i-th element of the variance maps is calculated as

$$v_i = \frac{\sum_{i=1}^{n^2}(\tilde{G}_{i,j} - \mu_i)^2}{n^2 - 1}, \quad \text{for } j = 1, \ldots, \frac{w \cdot h}{n^2}, \tag{9}$$

where μ_i is the mean value of the i-th patch. The GV loss is formulated as

$$\mathcal{L}_{GV} = \|v_x^{I_{rec}} - v_x^{I_g}\|_2 + \|v_y^{I_{rec}} - v_y^{I_g}\|_2, \tag{10}$$

aiming to minimize the variance disparity between I_{rec} and I_g, thus enhancing edge sharpness in I_{rec} [7].

D. Gaussian Edge-Enhanced Loss

Whereas GV loss is designed to enhance edge sharpness, it tends to overlook high-frequency details by focusing primarily on variance reduction. To complement this, we introduce the GEE loss, targeting both the blurring observed for MSE and the high-frequency detail limitations of GV loss. GEE emphasizes high-frequency components to enhance edges and details, leading to clearer images. This method is inspired by high-pass filtering techniques in image processing [10]. The GEE loss is computed as follows

$$\mathcal{L}_{GEE} = \left\| W \odot \left| \mathcal{F}^2(I_{rec}) \right| - W \odot \left| \mathcal{F}^2(I_g) \right| \right\|_1, \tag{11}$$

where I_{rec} represents the reconstructed image, G the ground truth, and $\mathcal{F}^2$ the 2D Fourier transform. The Gaussian high-pass filter weights, W, are defined as

$$W = 1 - e^{-\frac{(\sqrt{f_x^2 + f_y^2} - \kappa)^2}{2\sigma^2}}. \tag{12}$$

f_x and f_y are the frequency components along the x and y directions, and κ sets the cutoff frequency. The parameter σ adjusts the spread of the Gaussian function.

GEE loss focuses on the L_1 distance between the high-frequency amplitudes of the target and the output, prioritizing high-frequency information crucial for preserving edge details. The Gaussian high-pass filter smoothly transits emphasis to high frequencies, maintaining image integrity while enhancing clarity and sharpness. This mathematical framework allows W to modulate the frequency components, suppressing lower frequencies below κ while accentuating higher frequencies. The parameter σ controls the spread of the Gaussian function, influencing the filter's emphasis on higher frequencies.

III. Experiments and Results

Our study utilized the LoDoPaB-CT dataset [11], acclaimed for its clinical significance in low-dose CT image reconstruction. Comprising $35,802$ training samples, $3,522$ for validation, and $3,553$ for testing, the dataset provides an extensive range of 362×362 phantom images and corresponding 1000×513 sinograms, equivalent to 1,000 projections.

Our experimental framework was constructed using PyTorch 2.0 and integrated with Python 3.10. We selected the Adam optimizer [12], and initialized it with a learning rate of 5×10^{-3}. To further refine the training process and mitigate overfitting, we implemented the OneCycle learning rate policy. This strategy dynamically modulates the learning rate between 5×10^{-3} and 2×10^{-2}, optimizing the convergence speed and training stability. The model underwent training for a total of 20 epochs, utilizing one NVIDIA RTX A4000 GPU.

We employed a hybrid loss function that integrates MSE, our GEE loss and the GV loss to train our model. The loss function is formulated as follows

$$\mathcal{L} = \mathcal{L}_{MSE} + \alpha \mathcal{L}_{GEE} + \beta \mathcal{L}_{GV}. \tag{13}$$

In this formulation, α and β are weighting coefficients for the GEE and GV loss components. We chose $\alpha = 10$ and $\beta = 20$, ensuring that each component of the loss function contributes effectively to the model's overall performance. This setting equalizes the three loss values to the same order of magnitude, resulting in an optimal balance between overall image quality and precise edge enhancement.

Table I presents a quantitative comparison between our method and the traditional FBP approach with a Hann filter. Notably, our method achieved a Structural Similarity Index (SSIM) of 0.9379 ± 0.0718, MSE of 0.0009 ± 0.0005, and Peak Signal-to-Noise Ratio (PSNR) of 31.2365 ± 2.0532 dB. In contrast, FBP recorded lower values across these metrics, highlighting the superior effectiveness of our approach.

TABLE I
Comparison of Image Quality Metrics

Model	Mean ± Std SSIM	Mean ± Std MSE	Mean ± Std PSNR (dB)
Ours	0.9376 ± 0.0718	0.0009 ± 0.0005	31.2365 ± 2.0532
FBP	0.8688 ± 0.1041	0.0402 ± 0.0111	14.1179 ± 1.1711

Figure 2 provides a detailed visual comparison of CT image reconstructions utilizing traditional methods such as FBP, MSE optimization, and our proposed method employing a composite loss function. The effectiveness of the reconstruction techniques is particularly observed in high-contrast regions, which are crucial for identifying and evaluating clinical features. Our

method's reconstruction, distinctly visible in the red boxed insets, reveals a significant enhancement in the preservation of geometrical structures, especially evident in the vertebral architecture within the chest CT scans.

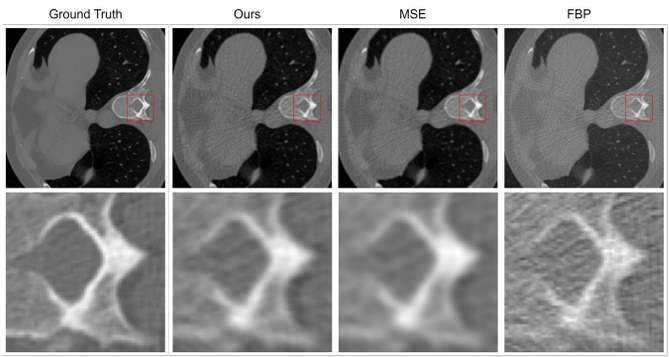


Fig. 2. Comparative analysis of reconstruction results: When compared to MSE, our composite loss function demonstrates superior detail retention and image sharpness. In comparison to the traditional FBP method, our approach excels in high-frequency noise suppression and provides clearer structural definition.

Delving deeper into the qualitative analysis, Figure 3 employs the Laplace map, a second-order gradient representation that accentuates regions of rapid intensity change, often associated with edges and other structural details in medical images. The bright bands in the Laplace map correlate with edges in the image; narrower bands represent sharper edges. The reconstruction of our model demonstrates narrow bands within the zoomed regions, indicating a preservation of edge sharpness.

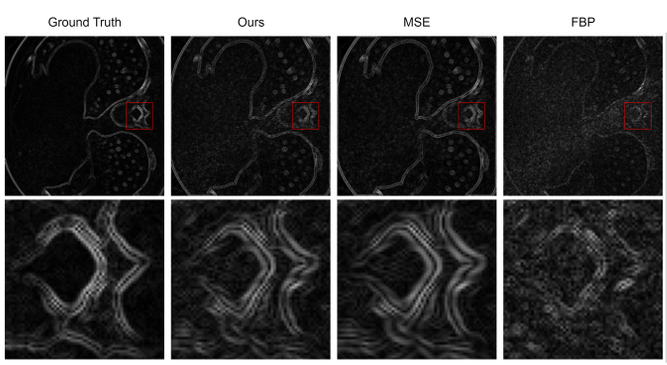


Fig. 3. Comparison of reconstruction in Laplace map: In contrast to supervised training using MSE, our approach demonstrates superior edge sharpness. When compared to the conventional FBP method, our approach showcases a more distinct edge structure and reduced high-frequency noise.

IV. Conclusion and Discussion

Our experiments reveals that our novel approach outperforms the conventional FBP and MSE optimized reconstruction methods in terms of image quality, especially within the context of low-dose CT scans that typically contain high levels of noise. The loss function formulated by us is adept at not only conserving but also enhancing essential diagnostic features within these noisy datasets, which is vital for medical diagnostics, as it aids in clearly visualizing anatomical boundaries and detecting pathology. The high-quality image reconstructions facilitated by our model indicate its potential to improve diagnostic substantially.

In summary, our study introduces a trainable filter for the FBP framework, capitalizing on Fourier series coefficients for minimal parameterization and adaptable resolution handling. With just 101 parameters, our method represents an advancement over traditional FBP, expanding its application scope and enhancing its robustness. The interpretability of our model is maintained, except for the data-driven filter, all parameters can be mathematically derived. This plug-and-play approach, coupled with a novel loss function targeting the L_1 norm of high-frequency magnitudes, ensures the preservation of fine details and optimizes computational efficiency and resource utilization throughout the reconstruction process. Future work will further refine this trainable filter to enhance its support for resolution invariance in FBP. Additionally, we plan to extend our method from 2D to 3D reconstruction.

Acknowledgment

This research was financed by the "Verbundprojekt 05D2022 - KI4D4E: Ein KI-basiertes Framework für die Visualisierung und Auswertung der massiven Datenmengen der 4D-Tomographie für Endanwender von Beamlines. Teilprojekt 5." (Grant number: 05D23WE1).

References

[1] G. N. Ramachandran and A. V. Lakshminarayanan, "Three-dimensional reconstruction from radiographs and electron micrographs: Application of convolutions instead of Fourier transforms," *Proc. Nat. Acad. Sci. U.S.A.*, vol. 68, no. 9, pp. 2236–2240, 1971.

[2] L. A. Shepp and B. F. Logan, "The Fourier Reconstruction of a Head Section," *IEEE Trans. Nucl. Sci.*, vol. 21, no. 3, pp. 21–43, 1974.

[3] J. D. Evans, D. G. Politte, B. R. Whiting, J. A. O'Sullivan, and J. F. Williamson, "Noise-resolution tradeoffs in x-ray CT imaging: A comparison of penalized alternating minimization and filtered back projection algorithms," *Medical Physics*, vol. 38, no. 3, pp. 1444–1458, 2011.

[4] C. Syben, B. Stimpel, K. Breininger, T. Würfl, R. Fahrig, A. Dörfler, and A. Maier, "Precision learning: Reconstruction filter kernel discretization," *arXiv preprint arXiv:1710.06287*, 2017.

[5] J. Xu, C. Sun, Y. Huang, and X. Huang, "Residual Neural Network for Filter Kernel Design in Filtered Back-projection for CT Image Reconstruction," in *Proc. German Workshop on Med. Image Comput.*, pp. 164–169, 2021.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770–778, 2016.

[7] L. Abrahamyan, A. M. Truong, W. Philips, and N. Deligiannis, "Gradient variance loss for structure-enhanced image super-resolution," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, pp. 3219–3223, 2022.

[8] A. K. Maier, C. Syben, B. Stimpel, T. Würfl, M. Hoffmann, F. Schebesch, W. Fu, L. Mill, L. Kling, and S. Christiansen, "Learning with known operators reduces maximum error bounds," *Nat. Mach. Intell.*, vol. 1, no. 8, pp. 373–380, 2019.

[9] C. Syben, M. Michen, B. Stimpel, S. Seitz, S. Ploner, and A. K. Maier, "PYRO-NN: Python reconstruction operators in neural networks," *Med. Phys.*, vol. 46, no. 11, pp. 5110–5115, 2019.

[10] A. Dogra and P. Bhalla, "Image Sharpening by Gaussian and Butterworth High Pass Filter," *Biomed Pharmacol J.*, vol. 7, no. 2, 2014.

[11] J. Leuschner, M. Schmidt, D. Otero Baguer, and P. Maass, "LoDoPaB-CT, a benchmark dataset for low-dose computed tomography reconstruction," *Sci. Data*, vol. 8, no. 1, p. 109, 2021.

[12] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv preprint arXiv:1412.6980*, 2014.

Diffusion Model-based Three-dimensional Image Reconstruction Algorithm from Five-view Projections

Jianpeng Gao, Liang Li, Liang Sheng, Yang Li, Baojun Duan, Yan Song, and Dongwei Hei

***Abstract*—In pulsed x-ray/neutron radiation experiments, such as z-pinch and laser fusion, the diagnosis is of great significance for revealing the properties of radiation sources. Multi-view radiation imaging is an effective diagnostic method to obtain the three-dimensional (3D) distribution of radiation sources. In imaging experiments, photons emitted from radiation sources are measured from different views and used to reconstruct the 3D structure of the source. However, due to the short duration and drastic changes of the pulsed radiation source, only up to five views data are available for reconstruction in the current experiments. 3D reconstruction from such a small amount of data is an extremely ill-posed problem, and many classical algorithms in computed tomography like filtered back projection and algebraic reconstruction technique are inapplicable, while other algorithms such as maximum likelihood expectation maximization and analytical reconstruction algorithms based on basis functions decomposition may introduce artifacts thus degrading the image quality. In this paper, we propose a 3D image reconstruction algorithm based on the diffusion model. We first guide image generation during the sampling process of diffusion through conditional score calculated by imaging statistical model, and then utilize a model-based iterative method to perform posterior correction on the generated images. The simulation results show that our algorithm can obtain high-quality 3D reconstructed images under five-view projections.**

***Index Terms*—Diffusion model, three-dimensional reconstruction, few-view computed tomography**

I. Related Work

THE diffusion model is an algorithmic framework inspired by the diffusion principle in non-equilibrium statistical physics[2]. It is a generative model widely used for image generation. The diffusion model consists of two processes, as shown in Figure 1. In the forward process, the data is progressively perturbed with small noise, and the data distribution is transformed into a Gaussian distribution eventually. In the reverse process, a noise image is sampled from the Gaussian distribution and gradually denoised to generate normal image.

Song et al. proposed a score matching based method to train diffusion models[3]. During the denoising process, the score was used to predict the noise in step t. An improved U-net with time embedding was adopted to approximate the score. The formula for training a score network is as follows[4]:

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_t \left\{ \eta_t \mathbb{E}_{\boldsymbol{x}_0} \mathbb{E}_{\boldsymbol{x}_t|\boldsymbol{x}_0} \left[\left\| \boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \nabla_{\boldsymbol{x}_t} \log p_{0t}(\boldsymbol{x}_t \mid \boldsymbol{x}_0) \right\|_2^2 \right] \right\} \quad (1)$$

where θ is the parameters of score network. η_t is a positive weighting function and $\boldsymbol{s}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)$ denotes the predicted score at step t.

Dhariwal et al. extended the practicality of diffusion models in different applications through classifier guidance[5]. They calculated classifier gradients via pre-trained classifiers, and then performed conditional sampling with modified noise predictions, improving the image quality generated by unconditional models.

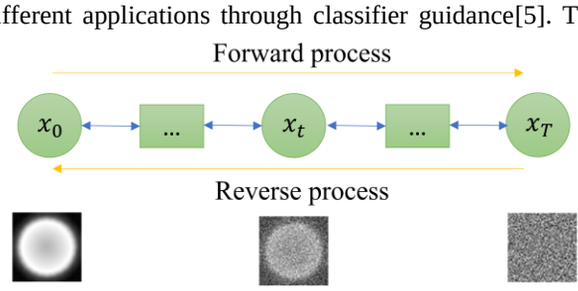


Fig. 1. The diagram of diffusion model.

II. Method

A. Conditional Score

For an unknown image x and the measurement data y, we use y as the condition to guide the generation of diffusion model. In score-based diffusion models, the sore is given by

$$s_\theta(x_t, t) = \nabla_{x_t} \log p(x_t) \quad (2)$$

After introducing the condition y, according to the Bayesian formula:

$$p(x_t \mid y) = \frac{p(y \mid x_t) p(x_t)}{p(y)} \quad (3)$$

Taking the derivative of x_t on both sides, we can get

$$\nabla_{x_t} \log p(x_t \mid y) = \nabla_{x_t} \log p(y \mid x_t) + \nabla_{x_t} \log p(x_t) \quad (4)$$

This work was supported by the grants from NNSFC 12027811.

Jianpeng Gao and Liang Li are with Department of Engineering Physics, Tsinghua University, Beijing, CO 100084 China (e-mail: lliang@tsinghua.edu.cn)

Liang Sheng, Dongwei Hei, Yang Li, Baojun Duan, and Yan Song are with National Key Laboratory of Intense Pulsed Radiation Simulation and Effect, Northwest Institute of Nuclear Technology, Xi'an, CO 710024 China (e-mail: shengliang@tsinghua.org.cn; heidw@163.com).

Equation (4) represents the new score function after introducing the condition, and the second term can be obtained from the pre-trained model $s_{\theta^*}(x_t, t)$.

The conditional probability is

$$p(y \mid x) = p(y \mid \Psi(x)) = \prod_i p(y_i \mid z_i) \tag{5}$$

where Ψ denotes the imaging model and y_i is the number of photons measured by the i-th detector. $z_i = [\Psi(x)]_i$ is the i-th element of vector $\Psi(x)$.

While the photon emission follows Poisson distribution, we can get the following equation:

$$\log p(y \mid x) = \sum_i \left(-z_i + y_i \log z_i\right) + C \tag{6}$$

Thus, we get

$$\nabla_x \log p(y \mid x) = J_\Psi{}^T \left[(y - \Psi(x)) / \Psi(x)\right] \tag{7}$$

where J_Ψ is the Jacobian matrix of Ψ.

The conditional score is

$$s_C = s_{\theta^*}(x_t, t) + \gamma J_\Psi{}^T \left[(y - \Psi(x_t)) / \Psi(x_t)\right] \tag{8}$$

where γ is the parameter to adjust the conditional score in the sampling process.

B. *Posterior Correction*

After one step of denoising generation, we use the model-based iterative method to perform a posterior correction on the generated results. In model-based iterative method, the objective function is

$$L_M = \frac{1}{2}\|y - \Psi(x)\|_2^2 + \lambda_1 \varphi(x) \tag{9}$$

where $\varphi(x)$ is the regularization term.

In the correction step, our objective function is

$$L = \lambda_s L_M + \frac{1-\lambda_s}{2}\left\|x - x\right\|_2^2 \tag{10}$$

where x is the image generated by diffusion model.

Let $\Psi(x) = Ax$ and $\varphi(x) = \|Dx\|_1$, the solution is

$$X^* = \arg\min_x \frac{\lambda_s}{2}\|y - Ax\|_2^2 + \frac{1-\lambda_s}{2}\left\|x - x\right\|_2^2 + \lambda_1\lambda_s \|Dx\|_1 \tag{11}$$

We get the following iterative formulas by the alternating direction method of multipliers[6] (ADMM) with $t = N-1, \ldots, 0$:

$$x_t = \left[(1-\lambda_s)I + \lambda_s(A^TA + \rho_1 D^T D)\right]^{-1} \left[(1-\lambda_s)x + \lambda_s\left(A^T y + \rho_1 D^T(u_{t+1} - w_{t+1})\right)\right] \tag{12}$$

$$u_t = S_{\frac{\lambda_1}{\rho_1}}(Dx_t + w_{t+1}) \tag{13}$$

$$w_t = w_{t+1} + Dx_t - u_t \tag{14}$$

The pseudocode of the algorithm is shown in Algorithm 1. PC-solver is the Predictor-Corrector sampling algorithm in the reverse process of diffusion model.

Algorithm 1: 3D image reconstruction via diffusion model

Inputs: s_{θ^*}, the pre-trained score; y, the measurements; A, system matrix; $\lambda_1, \lambda_s, \rho_1$, the parameters of ADMM algorithm.
1: $x_N \sim p_T(x), u_N \leftarrow 0, w_N \leftarrow 0$.
2: for t= $N-1, \ldots, 0$ do
3: Compute conditional score s_C
4: $x_t \leftarrow$PC-solver(s_c, x_{t+1})
5: $x_t \leftarrow$ ADMM(x_t)
6: end for
Output: x_0

III. Simulation Experiment

In the simulation experiment, we generated a dataset for training and testing. We use single wire z-pinch simulation data to generate thirteen 3D volumes, each with a size of 64×64×64. Ten of these 3D volumes are used to train the diffusion model. The model is trained with 2000 diffusion steps and the learning rate is 2×10^{-4}.

During reconstruction, we use parallel beam imaging to simplify calculations. The five views are distributed on a horizontal plane, as shown in Figure 2. The angles are 1-(90°, 180°), 2-(90°, 135°), 3-(90°, 90°), 4-(90°, 45°) and 5-(90°, 22.5°). We reconstruct the three volumes from test set and one volume from the training set. We compare our method with the 3D TV constrained reconstruction algorithm (ADMM-TV) and the algorithm based on cylindrical harmonic function decomposition and deep image prior[1] (CHD-DIP).

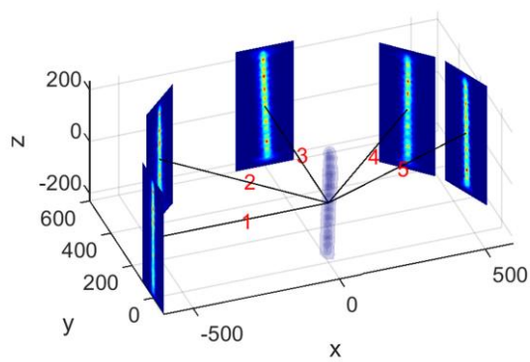


Fig. 2. The projection views distributed on horizontal plane are 1-(90°, 180°), 2-(90°, 135°), 3-(90°, 90°), 4-(90°, 45°) and 5-(90°, 22.5°)[1].

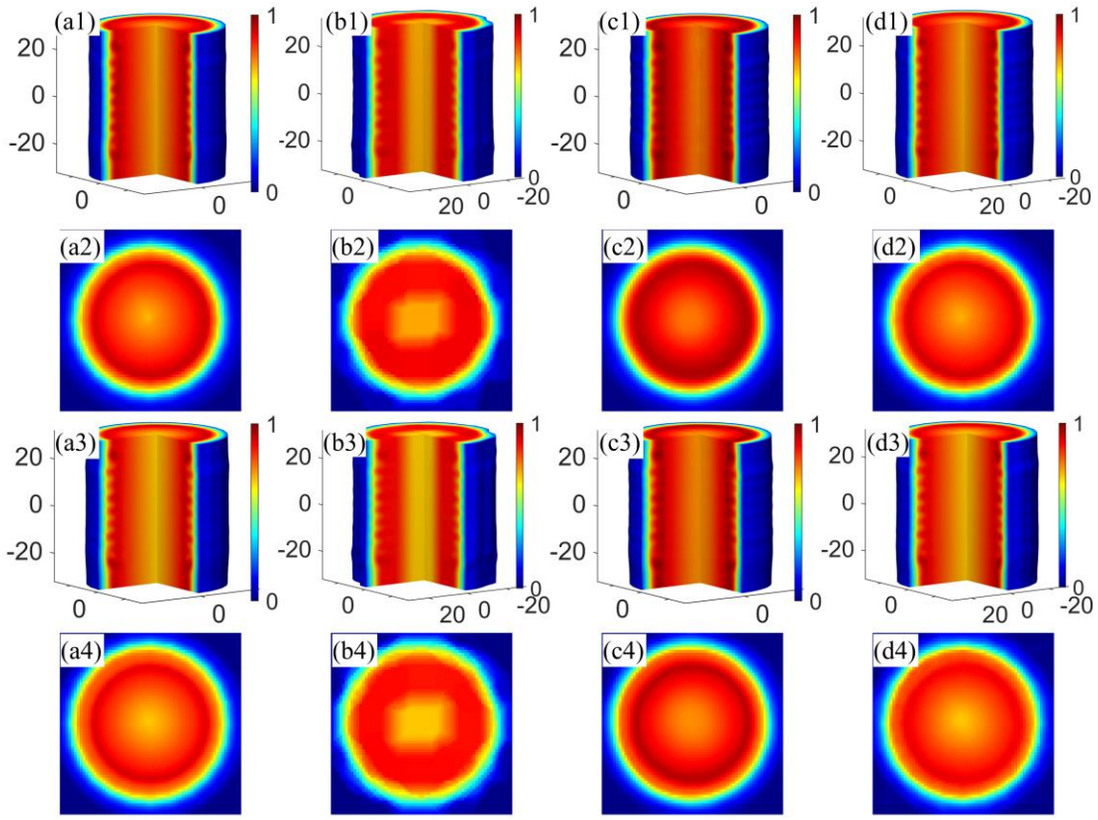


Fig. 3. The reconstructed results of five-view data. Fig. (a1)-(a4) are ground truth 3D volumes from test set. Fig. (a1) and (a3) are the 3D view of the volumes. Fig. (a2) and (a4) are the slice image at z=0. Fig. (b1)-(b4) are the reconstructed images of ADMM-TV. Fig. (c1)-(c4) are the results of CHD-DIP. Fig. (d1)-(d4) are the results of proposed method.

The results are shown in Figure 3. The PSNR and SSIM of the reconstructed 3D images are listed in Table I. It can be seen that there are blocky artifacts in the ADMM-TV reconstruction results, which cannot reconstruct the circular structure. The CHD-DIP can reconstruct the 3D distribution roughly, but there is a certain difference between the results and the ground truth in the center and circular parts. The results of our method show small difference both in the center and circular parts. The numerical evaluation results show that for data in the training set, although the diffusion model only learns 2D distributions separately, it can still reconstruct the 3D structure. As for the data in test set, the diffusion model can also achieve high-quality reconstruction results in five views.

IV. Conclusion

Diffusion model, as a class of generative models, introduces random factors in the image generation process, resulting in uncontrollable images. Through conditional score and posterior correction, diffusion model can handle 3D reconstruction tasks and obtain good reconstruction results from few-view data. In the future, we will expand the algorithm for large image reconstruction and improve the quality of images reconstructed with fewer views to enhance the applicability of the algorithm.

Acknowledgment

The authors would like to thank the research group of Professor Wu Jian from Xi'an Jiaotong University for providing simulation support.

TABLE I
IMAGE ASSESSMENT INDICES OF VOLUMES RECONSTRUCTED BY ADMM-TV, CHD-DIP, AND THE PROPOSED METHOD.

volumes	trian10		test11		test12		test13	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ADMM-TV	30.432	0.89967	30.056	0.89130	29.530	0.89760	28.851	0.88955
CHD-DIP	35.043	0.97949	33.249	0.98143	34.600	**0.97985**	34.608	**0.97943**
Ours	**52.029**	**0.99835**	**43.361**	**0.99192**	**38.601**	0.97830	**35.186**	0.95565

Boldface denotes that the image assessment index of corresponding algorithm is better than that of other algorithms.

REFERENCES

[1] J. Gao *et al.*, "Five-view three-dimensional reconstruction for ultrafast dynamic imaging of pulsed radiation sources," *Matter and Radiation at Extremes*, vol. 9, no. 2, 2024, doi: 10.1063/5.0177342.

[2] J. N. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep Unsupervised Learning using Nonequilibrium Thermodynamics," *ArXiv*, vol. abs/1503.03585, 2015.

[3] Y. Song and S. Ermon, "Generative Modeling by Estimating Gradients of the Data Distribution," in *33rd Conference on Neural Information Processing Systems (NeurIPS)*, Vancouver, CANADA, Dec 08-14 2019, vol. 32, in Advances in Neural Information Processing Systems, 2019.

[4] Y. Song, J. SohlDickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-Based Generative Modeling through Stochastic Differential Equations," *CoRR*, vol. abs/2011.13456, 2020.

[5] P. Dhariwal and A. Nichol, "Diffusion Models Beat GANs on Image Synthesis," *ArXiv*, vol. abs/2105.05233, 2021.

[6] S. Boyd, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1-122, 2010, doi: 10.1561/2200000016.

Simultaneous Radiation Dose and Scatter Reduction in Breast CT Imaging using a Nearly Half-Sized Detector with a Self-Supervised Deep-Learning Algorithm

Hsin Wu Tseng, Zhiyang Fu, and Srinivasan Vedantham
Department of Medical Imaging, The University of Arizona, Tucson, AZ 85724, USA

***Abstract*—The goal of this study was to investigate the feasibility of using a self-supervised deep learning algorithm to reconstruct images from a nearly half-sized detector with narrow beams to reduce both radiation dose and scatter at the same time without sacrificing image quality. Projection datasets were retrospectively truncated into different sizes along the fan-angle direction. Feldkamp-Davis-Kress (FDK) reconstructed images using the complete projection data ($m \times n$ = 1024 × 768) served as the reference. FDK using weighting schemes and a recently developed self-supervised deep-learning (SSDL) algorithm, Attenuation Field Network (AFN), were employed to reconstruct the truncated sinograms. The variance increases when m (along fan angle) decreases for FDK images using weighting schemes, and artifacts can be observed when m is less than 596. The AFN demonstrated the ability to maintain the image quality with m larger than 536. The reference FDK reconstructions were segmented to create numerical phantoms and used with Monte Carlo simulations emulating the truncated imaging geometries. The radiation dose and scatter-to-primary ratio (SPR) were reduced by 37.9% and 33.5% for m = 536, respectively. Quantitative image quality evaluations show that the feasibility of incorporating AFN and using narrower detectors to reduce the dose and SPR simultaneously is promising. The technique developed here can be applied to other cone-beam CT systems using flat-panel detectors.**

***Index Terms*—breast CT, dose reduction, SPR, self-supervised, deep learning, truncated detector**

I. Introduction

X-ray cone-beam computed tomography (CBCT) provides isotropic three-dimensional (3D) visualization of internal anatomical structures over a large area in a single scan at a fast speed. Recent technological progress in detectors and image reconstruction algorithms has made diagnostic imaging of low-contrast lesions using flat-panel detectors (FPDs) feasible. Dedicated breast CT imaging using cone-beam systems with FPDs to distinguish fibroglandular tissue and soft tissue lesions is one example. Cone-beam breast CT (CBBCT) improves over digital mammography (DM) and digital breast tomosynthesis (DBT) in terms of compression-free exams and mitigating the masking effects due to tissue overlap. The masking effect is the main source of lesions missed during interpretation, especially in dense breasts. Also, the sensitivity of CBBCT in diagnostic imaging has been reported to be better than that of DM in a reader study [1]. In addition, CBBCT avoids various artifacts observed in DBT [2]. The CBBCT system using a high-resolution detector with a lateral offset has recently been evaluated [3], [4]. Towards becoming a breast cancer screening modality, there are some challenges for CBBCT. Like CBCT systems, CBBCT suffers scatter effects due to the large area of FPDs, resulting in cupping artifacts. Another challenge is the visualization of microcalcifications becomes difficult [5] when the mean glandular dose (MGD) is reduced to 4.15 mGy which corresponds to the MGD from standard 2-view mammography [6]. Thus, reducing dose and scatter radiation is essential to translate CBBCT to be a screening tool.

Several techniques have been developed for scatter reduction or correction in breast CT including tungsten plates [7], [8], narrow beams [9], library-based algorithms [10], forward-projection model [11], and U-Net-based deep-learning (DL) approaches incorporated with MC-generated projections [12]. There were many advanced technologies developed to reduce the radiation dose in CBBCT as well. Advanced image reconstructions such as compressed-sensing (CS) [13]–[15] and DL methods [16], [17] are common approaches. Protocol-based methods such as short-scan, sparse-view [18], and shorter-pulse width [4] and data acquisition-based methods [19] have also been proposed. However, none of them can reduce the dose and scatter simultaneously.

To meet this need, we propose to use a nearly half-sized FPD with a narrow beam to simultaneously reduce the dose and scatter. Half-sized detectors (along fan angle direction) have been demonstrated to significantly reduce the scatter radiation [20]. The radiation dose is reduced by about 50% with a half-sized detector. The challenge is to reconstruct the breast from projections that only have half the information at every angle. Although images can be successfully reconstructed in the fan-beam geometry [21], it is not trivial for the cone-beam geometry. The traditional FDK algorithm with weighting schemes has been developed and applied in our CBBCT system

with an offset truncated detector. The FDK with weighting schemes that leverage the redundancy information from complementary projections to estimate missing information might fail when the number of detector pixels, *m*, is too small. To address this issue, we applied our recently developed self-supervised deep learning (SSDL), attenuation field network (AFN [22]), to overcome this problem. This study aims to investigate the performance and limit of the FDK with weighting schemes and AFN in terms of the number of pixels, *m*, and image quality.

II. MATERIALS AND METHODS

A. Clinical Data and CBBCT System

In this retrospective study, sinograms from 30 women who participated in an institutional review board-approved, HIPPA-compliant clinical study (clinicaltrials.gov: NCT01090687) were used. All projection images were acquired using a pre-FDA approval prototype dedicated CBBCT system (Koning Corp, Norcross, GA, USA). Since the radiation dose and X-ray scatter are the two main reasons for the degraded visualization of microcalcifications in CBBCT, 29 out of 30 sinograms with calcified lesions were chosen in this study.

The CBBCT scanner used an FPD (PaxScan 4030CB, Varex Imaging, Salt Lake City, UT, USA) with dimensions of 40 cm × 30 cm along the fan-angle and cone-angle directions, respectively, and covered the entire field of view (FOV). The 2 × 2 binned mode was used for data acquisition resulting in 1024 × 768 pixels of 0.388 mm pitch along the fan- and cone-angle directions. The x-ray tube (RAD-71SP, Varex Imaging, Salt Lake City, UT, USA) was operated in pulse mode (8 ms), and the x-ray spectrum was 49 kVp (1st half-value layer thickness: 1.4 mm of Al). The source-to-detector distance was 898 mm, and the distance between the source and the axis of rotation (AOR) was 650 mm. There were 300 projections acquired in the full scan trajectory, and the scan time was 10 seconds.

B. Emulation of CBBCT System with an Offset Detector

To emulate the truncated sinograms acquired from an incomplete detector with a narrow beam, we retrospectively truncated each projection along the fan-angle direction into 584, 572, 560, 548, 536, 524, and 512 pixels. The truncated projections were used to investigate the performance limit of the image reconstruction algorithms.

C. Image Reconstruction Algorithms

Two algorithms were investigated. The truncated detector can be reconstructed by FDK with proper weighting schemes [23]–[25] using data redundancy. These schemes leverage the information from complementary projections to estimate the missing information. We have demonstrated this method before [26] and have shown that there is no significant difference in the statistical results among the weighting functions proposed by Cho [23], Schaefer [24], and Wang [25]. Thus, we used FDK with the Wang weighting function, referred to as FDK-W, for the rest of our study. The reconstruction voxel size was 0.2734 mm, and the reconstructed volume had 1024 × 1024 × *Z* voxels (*Z* depends on the length from the chest wall to the nipple) in this study.

Another algorithm used in this study was our recently developed deep-learning (DL) approach, i.e., attenuation field network (AFN) [22]. It is self-supervised and does not require training data other than the data to reconstruct. This algorithm currently has been demonstrated for reconstructing sinograms from an incomplete detector (768 ×768) with a short-scan acquisition via the concept of sinogram inpainting. We modified this technique to the current problem, where the acquisition is a full scan with the dimension of the detector fan-angle direction, *m*, approaching 512 pixels, i.e., a half-sized detector.

A minimum training sample in the self-training procedure of the AFN technique is a ray propagating from the x-ray source $\boldsymbol{s}$ to a detector pixel $\boldsymbol{d}$. Multiple samples are obtained through the ray $\overrightarrow{\boldsymbol{sd}}$, and they can be represented as $\boldsymbol{t}_i = \boldsymbol{s} + \alpha_i(\boldsymbol{d} - \boldsymbol{s})$, $0 < \alpha_i < \alpha_{i+1} < 1, \forall i$. Therefore, the attenuation coefficients of these samples are integrated along the ray path to provide the projection, and can be expressed as:

$$\hat{\boldsymbol{p}}(\boldsymbol{d}) = \sum_i h_\Theta(\boldsymbol{t}_i)|\boldsymbol{t}_{i+1} - \boldsymbol{t}_i|, \tag{1}$$

where $\hat{\boldsymbol{p}}(\boldsymbol{d})$ is the estimated projection and $h_\Theta(\cdot)$ denotes an AFN parameterized by Θ. The error between the estimated projection $\hat{\boldsymbol{p}}(\boldsymbol{d})$ and the real projection $\boldsymbol{p}(\boldsymbol{d})$ serves as the training loss to optimize the network's representation:

$$\underset{\Theta}{\operatorname{argmin}} \sum_{\boldsymbol{d}} \|\hat{\boldsymbol{p}}(\boldsymbol{d}) - \boldsymbol{p}(\boldsymbol{d})\|^2, \boldsymbol{d} \in \mathbb{D}, \tag{2}$$

where $\mathbb{D}$ denotes the set of coordinates of all the acquired projection data. A trained AFN is a 3D representation of the attenuation coefficients of the imaged object. When the same imaging geometry (as that of the acquired data) is emulated, AFN can inpaint missing data. The inpainted projection data are spliced with the acquired projection data to obtain complete projection data, which can be subsequently reconstructed by any desired reconstruction method. In this entire study, the complete hybrid data were reconstructed by FDK.

D. Scatter-to-Primary Ratio (SPR) and MGD Estimate

The reference 3D reconstruction of each breast was segmented [19], [27] into air, skin, adipose, and fibroglandular tissue. Each segmented tissue type was assigned its physical properties including mass, density, and attenuation coefficient followed by Monte Carlo (MC) simulations using the GEANT4-based GATE 8.0 toolkit. 3×10^{10} X-ray photons were simulated with the same spectral conditions used during acquisition, and 30 projections over 360° with 12° angular spacing were generated. The x-ray focal spot was simulated as a 0.3 mm × 0.3 mm square. The x-ray field was collimated so that the beam was limited to detector dimensions of *m* × 768 pixels, where *m* was selected based on the results from the image quality evaluation described below. The one-dimensional horizontal profile of the scatter and primary photon distributions was obtained by averaging over a rectangular region of interest (ROI) [20], and over all 30 projections.

The MGD (in mGy), was determined from the normalized glandular dose coefficient, DgN^{CT} (in mGy/mGy) from MC simulations using the heterogeneous model, followed by multiplying it with the measured air kerma (mGy) at the AOR. The heterogeneous DgN^{CT} is expressed as [28]:

$$DgN_{hete}^{CT} = \frac{E_{g,dep}}{n_g m_g \sum_E \Phi(E)\Theta(E)} \quad (3)$$

where $E_{g,dep}$ is the total energy deposited in all fibroglandular voxels, n_g is the total number of fibroglandular tissue voxels, m_g is the mass of fibroglandular tissue voxel, $\Phi(E)$ is the x-ray fluence as a function of energy, E, and $\Theta(E)$ is the fluence-to-air kerma conversion factor at energy E.

III. RESULTS

A. FDK-W

Shading artifacts can be observed when the dimension along the fan angle is truncated from 1024 to 584 pixels (Fig. 1), and the artifacts become brighter when truncation increases. The center of this artifact is located at the center of the reconstructed FOV, *i.e.*, (512, 512). Thus, this method fails when the truncated dimension is more than 42%. In this study, m=596 was considered as the limit with the FDK-W algorithm.

B. AFN

With AFN, no artifact was observed even when the number of pixels in the fan-angle direction decreased to 536 (Fig. 2). However, artifacts were observed in some of the 30 cases when the number of pixels was less than 520. Conservatively, m=536 was considered as the limit for the AFN algorithm. The calcifications can be visualized clearly in AFN images reconstructed from the 536 x 768 detector (Fig. 3).

C. Quantitative Image Quality Evaluation

Quantitative image quality was measured based on the signal difference-to-noise ratio (SDNR) and full-width at half maximum (FWHM) of microcalcifications in the mediolateral (ML) direction and the superior-inferior (SI) direction. An ROI of 32 × 32 pixels was used. The SDNRs of the reference, FDK-W, and AFN were 6.71 ± 2.53, 4.12 ± 1.51, and 6.62 ± 2.31, respectively. The FWHM (mm) in the ML direction of the reference, FDK-W, and AFN were 1.67 ± 0.57, 1.66 ± 0.50, and 1.64 ± 0.56, respectively. The FWHM (mm) in the SI direction of the reference, FDK-W, and AFN were 1.67 ± 0.75, 1.67 ± 0.77, and 1.63 ± 0.76, respectively.

D. SPR and Radiation Dose

The 1D horizontal SPR profile averaged across all projection views, ROI, and all thirty samples have maximum values of 0.326, 0.258 (21.1% reduction), and 0.217 (33.5% reduction), respectively, for the complete detector, 596-pixel detector, and 536-pixel detector (Fig. 4). The MGD were 8.37 mGy, 5.46, and 5.2 mGy for the complete detector, 596-pixel detector, and 536-pixel detector, respectively, which indicates 34.8% and 37.9% dose reductions, respectively.

IV. DISCUSSION AND CONCLUSIONS

In this study, we investigated FDK-W and AFN for nearly half-sized detectors. With FDK-W, the images were nosier and demonstrated artifacts when $m \leq 596$. Unlike FDK-W, AFN can maintain the SDNR for $m \geq 536$. Both FDK-W and AFN maintain the resolution of the calcifications. The SPR and the MGD were reduced by more than 30% with the use of the 536-detector, and the MGD was lower than the 6 mGy limit currently used for breast cancer screening. We have demonstrated the feasibility of using a nearly half-sized detector along with a novel algorithm, AFN, to maintain image quality and reduce both SPR and radiation dose simultaneously.

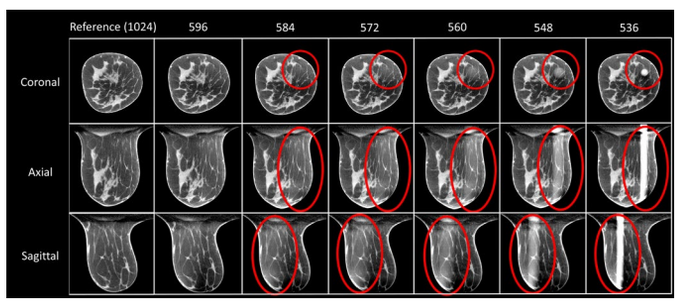


Fig. 1. FDK reference and FDK-W images (central slice; magnified). From top to bottom: coronal, axial, and sagittal views. From left to right are the FDK-W images reconstructed with varying numbers of detector pixels along the fan-angle direction. The reference FDK images use a complete detector (1024 x 768 pixels). All images were displayed as a 5-slice average with a window of [0.2, 0.3] cm^{-1}.

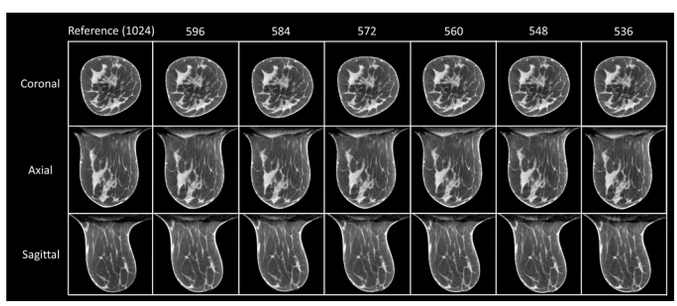


Fig. 2. FDK reference and AFN images (central slice; magnified). From top to bottom: coronal, axial, and sagittal views. From left to right are the AFN images reconstructed with varying numbers of detector pixels along the fan-angle direction. The reference FDK images were obtained with a complete detector (1024 x 768 pixels). All images were displayed as a 5-slice average with a window of [0.2, 0.3] cm^{-1}.

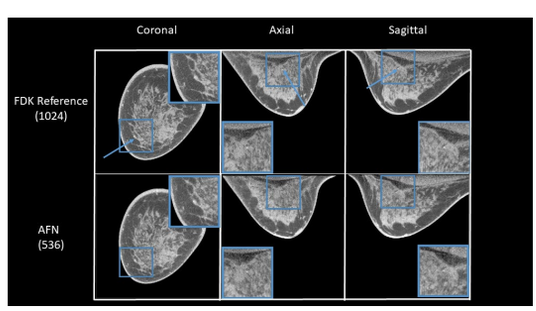


Fig. 3. FDK reference and AFN images (calcification slice; magnified). Blue squares are the calcification ROI. From left to right: coronal, axial, and sagittal views. The reference FDK images were obtained with a complete detector (1024 x 768 pixels). The AFN images were obtained with a truncated detector (536 x 768 pixels). All images were displayed as a 5-slice average with a window of [0.2, 0.3] cm^{-1}.

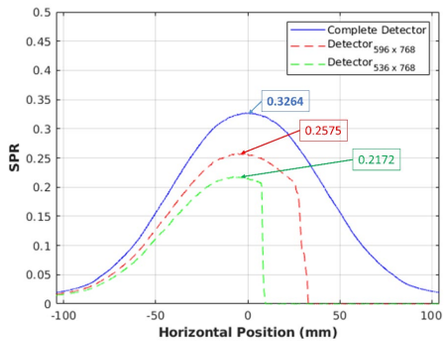


Fig. 4. Horizontal SPR profile. Blue curve: The reference (complete detector with 1024 x 768 pixels). Red curve: detector of 596 x 768 pixels. Green curve: detector of 536 x 768 pixels.

ACKNOWLEDGMENT

This work was supported in part by the National Cancer Institute (NCI) of the National Institute of Health (NIH) grants R01CA199044 and R01CA241709. The contents are solely the responsibility of the authors and do not represent the official views of the NCI or the NIH.

REFERENCES

[1] E. B. Cole, A. S. Campbell, S. Vedantham, E. D. Pisano, and A. Karellas, "Clinical Performance of Dedicated Breast Computed Tomography in Comparison to Diagnostic Digital Mammography (Abstract # SSA01-09)," in *Annual Meeting & Scientific Assembly of Radiological Society of North America, RSNA 2015, Chicago, IL*.

[2] P. S. Sujlana, M. Mahesh, S. Vedantham, S. C. Harvey, L. A. Mullen, and R. W. Woods, "Digital breast tomosynthesis: Image acquisition principles and artifacts," *Clin Imaging*, vol. 55, pp. 188–195, 2019.

[3] R. Trinate, H. W. Tseng, T. Larsen, Z. Fu, and S. Vedantham, "Objective Physics-Based Image Quality Characterization of a Low-Noise, High-Resolution, Dedicated Cone-Beam Breast CT Using an Offset Detector," presented at the American Association of Physicists in Medicine, Annual Meeting of the American Association Physicists in Medicine (AAPM), Houston, Texas, United States, 2023.

[4] H. W. Tseng, Z. Fu, and S. Vedantham, "Radiation dose reduction in cone-beam breast CT using shorter X-ray pulse-width with a self-supervised deep learning algorithm," presented at the American Association of Physicists in Medicine, Annual Meeting of the American Association Physicists in Medicine (AAPM), Houston, Texas, United States, 2023.

[5] K. K. Lindfors, J. M. Boone, T. R. Nelson, K. Yang, A. L. C. Kwan, and D. F. Miller, "Dedicated breast CT: initial clinical experience," *Radiology*, vol. 246, no. 3, pp. 725–733, Mar. 2008.

[6] R. E. Hendrick, E. D. Pisano, A. Averbukh, C. Moran, E. A. Berns, M. J. Yaffe, B. Herman, S. Acharyya, and C. Gatsonis, "Comparison of acquisition parameters and breast dose in digital mammography and screen-film mammography in the American College of Radiology Imaging Network digital mammographic imaging screening trial," *AJR Am J Roentgenol*, vol. 194, no. 2, pp. 362–369, Feb. 2010.

[7] I. Sechopoulos, "X-ray scatter correction method for dedicated breast computed tomography," *Med Phys*, vol. 39, no. 5, pp. 2896–2903, May 2012.

[8] S. Ramamurthy, C. J. D'Orsi, and I. Sechopoulos, "X-ray scatter correction method for dedicated breast computed tomography: improvements and initial patient testing," *Phys Med Biol*, vol. 61, no. 3, pp. 1116–1135, Feb. 2016.

[9] P. Ghazi, G. Fu, T. Ghazi, P. Kazanzides, and I. Iordachita, "Narrow beam breast CT: proof-of-concept," *Medical Physics*, vol. 50, no. 6, pp. 3418–3434, 2023.

[10] L. Shi, S. Vedantham, A. Karellas, and L. Zhu, "Library based x-ray scatter correction for dedicated cone beam breast CT," *Med Phys*, vol. 43, no. 8, p. 4529, Aug. 2016.

[11] L. Shi, S. Vedantham, A. Karellas, and L. Zhu, "X-ray scatter correction for dedicated cone beam breast CT using a forward-projection model," *Med Phys*, vol. 44, no. 6, pp. 2312–2320, Jun. 2017.

[12] J. J. Pautasso, M. Caballo, M. Mikerov, J. M. Boone, K. Michielsen, and I. Sechopoulos, "Deep learning for x-ray scatter correction in dedicated breast CT," *Med Phys*, vol. 50, no. 4, pp. 2022–2036, Apr. 2023.

[13] J. Bian, K. Yang, J. M. Boone, X. Han, E. Y. Sidky, and X. Pan, "Investigation of iterative image reconstruction in low-dose breast CT," *Phys. Med. Biol.*, vol. 59, no. 11, pp. 2659–2685, Jun. 2014.

[14] 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.

[15] 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," *Physica Medica*, vol. 73, pp. 117–124, May 2020.

[16] Z. Fu, H. W. Tseng, S. Vedantham, A. Karellas, and A. Bilgin, "A residual dense network assisted sparse view reconstruction for breast computed tomography," *Scientific Reports*, vol. 10, no. 1, Dec. 2020.

[17] H. Xie, H. Shan, W. Cong, C. Liu, X. Zhang, S. Liu, R. Ning, and G. Wang, "Deep Efficient End-to-end Reconstruction (DEER) Network for Few-view Breast CT Image Reconstruction," *IEEE access : practical innovations, open solutions*, vol. 8, p. 196633, 2020.

[18] 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*, Sep. 2020.

[19] H. W. Tseng, A. Karellas, and S. Vedantham, "Dedicated cone-beam breast CT: Data acquisition strategies based on projection angle-dependent normalized glandular dose coefficients," *Med Phys*, vol. 50, no. 3, pp. 1406–1417, Mar. 2023.

[20] G. Mettivier, P. Russo, N. Lanconelli, and S. L. Meo, "Cone-beam breast computed tomography with a displaced flat panel detector array," *Med Phys*, vol. 39, no. 5, pp. 2805–2819, May 2012.

[21] B. Rutt and A. Fenster, "Split-filter computed tomography: a simple technique for dual energy scanning," *J Comput Assist Tomogr*, vol. 4, no. 4, pp. 501–509, Aug. 1980.

[22] Z. Fu, H. W. Tseng, and S. Vedantham, "An attenuation field network for dedicated cone beam breast CT with short scan and offset detector geometry," *Sci Rep*, vol. 14, no. 1, p. 319, Jan. 2024.

[23] P. S. Cho, A. D. Rudd, and R. H. Johnson, "Cone-beam CT from width-truncated projections," *Comput Med Imaging Graph*, vol. 20, no. 1, pp. 49–57, Feb. 1996.

[24] D. Schaefer, M. Grass, and P. van de Haar, "FBP and BPF reconstruction methods for circular X-ray tomography with off-center detector," *Medical Physics*, vol. 38, no. S1, May 2011.

[25] G. Wang, "X-ray micro-CT with a displaced detector array," *Med Phys*, vol. 29, no. 7, pp. 1634–1636, Jul. 2002.

[26] S. Vedantham, H.-W. Tseng, S. Konate, L. Shi, and A. Karellas, "Dedicated cone-beam breast CT using laterally-shifted detector geometry: Quantitative analysis of feasibility for clinical translation," *J Xray Sci Technol*, vol. 28, no. 3, pp. 405–426, 2020.

[27] S. Vedantham, H. W. Tseng, Z. Fu, and H.-H. S. Chow, "Dedicated Cone-Beam Breast CT: Reproducibility of Volumetric Glandular Fraction with Advanced Image Reconstruction Methods," *Tomography*, vol. 9, no. 6, pp. 2039–2051, Dec. 2023.

[28] H. W. Tseng, A. Karellas, and S. Vedantham, "Radiation dosimetry of a clinical prototype dedicated cone-beam breast CT system with offset detector," *Med Phys*, vol. 48, no. 3, pp. 1079–1088, Mar. 2021.

Preliminary simulation results in model-based iterative DECT with respect to convergence speed, truncation and dose-reduction via double half scan geometry

Richard N.K. Bismark and Georg Rose

***Abstract*—We implemented a model-based iterative dual-energy CT (mbDECT) reconstruction technique proposed by Johnston et al. [4]. It iteratively minimizes a log-likelihood cost function by gradient descent in order to enable a quantitative segmentation of the CT images resulting from two CT scans. The flexibility of the approach is used to showcase two likely use cases for cone-beam flat-panel C-arm CT, namely truncation and dose reduction by applying two 100° half scan protocols. For this we simulated projection data with software phantoms and show that it can estimate material distribution that can distinguish both materials accordingly. For now, the error with respect to the groundtruth is too high to give quantitative results.**

***Index Terms*—Beam hardening, C-arm CT, Cone beam, Material Decomposition, Truncation, Dual-Energy.**

I. Introduction

DUAL-ENERGY CT stands as a widely employed technique in the realm of medical imaging for discerning spatial material distribution through the utilization of two CT scans featuring distinct spectral X-ray properties. Fundamentally, it operates on the premise of employing two different X-ray spectra, leading to disparate attenuation values within the volume space. The crucial divergence in these attenuation values becomes the cornerstone for distinguishing materials based on their photon energy-dependent attenuation.

The primary objective of dual-energy CT (DECT) is the reconstruction of the spatial distribution of materials. This segmentation can be achieved through two principal methodologies: firstly, by manipulating the X-ray projections prior to CT reconstruction, referred to as raw data-based DECT; and secondly, by employing numerical methods on the resultant CT images, known as image-based DECT. Both of these approaches grapple with the challenge of error propagation, wherein the outcome hinges on a direct comparison of either the measurements or intermediate results derived from an image processing pipeline.

The inherent issue of error propagation exacerbates the resolution of challenges such as truncation and limited angular resolution. This complication arises due to conflicts with the assumptions underpinning the standard image processing pipelines employed in CT reconstruction. Enter model-based iterative methods, which offer an alternative paradigm. Unlike their counterparts, these methods amalgamate all measurements into a cohesive pipeline, ensuring that the result aligns consistently with all measured signals simultaneously.

This integration is achieved through a sequential process involving the backprojection of an error signal onto the volume space. This error signal is proportionate to the disparity between the measurements and the estimated measurements, given the current approximation of the material distribution within the volume space. Consequently, the model-based iterative method circumvent the pitfall of error propagation, presenting a more robust and coherent approach to dual-energy CT applications. The trade-off though, is its increased computation requirements.

R. Bismark and G. Rose are with the Institute for Medical Engineering and Research Campus *STIMULATE*, University of Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany, richard.bismark@ovgu.de

II. Methods

The objective of model-based dual-energy CT (mbDECT) is to determine a material distribution of two basis materials, denoted as c_{jk}, which aligns with the measurements Y_i. These measurements are modeled as a function of the volume space $\bar{Y}_i(c_{jk})$. In this context, k serves as the material index, taking values of 1 and 2; i encompasses all measured X-ray intensities, accounting for both scan geometries; and j spans all voxels within the volume space. We have adopted a methodology inspired by Johnston et al.[4]. The iterative update scheme minimizes a log-likelihood cost function K through a gradient descent step in each iteration n:

$$c_{jk}^{n+1} = \left[c_{jk}^{n} - R \frac{\partial}{\partial c_{jk}^{n}} K\left(c_{jk}^{n}\right) \right]_{+} . \tag{1}$$

Here, R is the relaxation parameter, adjustable to influence convergence speed, and the outer brackets impose a non-negativity constraint. The gradient $\frac{\partial}{\partial c_{jk}^{n}}$ applies the backprojection of the error signal. The cost function $K = \left(\bar{Y}_i - Y_i \ln \bar{Y}_i\right)$ is contingent on the distribution of basis material voxels via the forward projection model:

$$\bar{Y}_i = \int X_i\left(E\right) \exp\left\{ - \sum_j a_{ij} \left(\sum_d c_{jd} f_d\left(E\right) \right) \right\} \mathrm{d}E. \tag{2}$$

In this equation, $X_i(E)$ represents a blank scan of X-ray intensity, $f_d(E)$ denotes the energy-dependent X-ray attenuation property of the d-th basis material, and $a_i j$ are the geometry factors. Note that the blank scan is dependent on the index i, thus includes the swap which is characteristic for DECT.

A. Convergence Speed

To increase convergence we changed $R \longrightarrow R^n_{jk}$ to be non-uniform in volume space and to be dependent the material k and on the iteration number n. The idea is inspired by the work of Kim, Elbakri and Fessler [5], [6].

$$R^n_{jk} = \alpha(n) / \left[\sum_i m_{ik}(E_{\text{eff}})^2 a_{ij} \gamma_i Y_i \right]. \tag{3}$$

Here, $m_{ik}(E_{\text{eff}})$ is the energy-dependent mass attenuation coefficient of the k-th basis material at the effective photon energy with respect to the used X-ray spectrum and γ_i is the intersection length of the i-th ray with the reconstruction volume. Additionally we made the relaxation dependent on the iteration n by linearily changing between two given boundaries

$$\alpha(n) = 30 - (3 \cdot n \textbf{mod} 10). \tag{4}$$

This adaption was found to increase the convergence properties of the algorithm dramatically. Note that the denominator of eq. (3) is independent of the iteration and can be pre-computed. Since the algorithm does apply ordered subsets, one iteration means that all views are used once.

B. Truncation

Truncated projection data introduces undersampling in the volume space, leading to a deficiency in information. To address this, we employ a compensatory strategy. We posit that the undersampled outer region of the volume comprises solely the first basis material, such as water. Voxel values in this region are estimated through an iterative model-based CT reconstruction method[2]. The forward projection of this outer region is precomputed, and this information is subsequently incorporated into the forward projection steps during the update process (1) as follows:

$$K\left(c^n_{jk}\right) \longrightarrow K\left(c^n_{jk}; c^1_{j'1}\right) \tag{5}$$

Here, the index $j'1$ denotes the voxel coordinates in the undersampled outer volume parts, consisting solely of the first basis material and remaining iteration-independent. This adaptation is inspired by the approach of Ziegler et al.[3] and was used to great success already in model-based statistical CT reconstruction of a single flat-panel cone-beam data set in head scans[1].

C. Double Half Scan Geometry

Conventional image-based DECT methods typically employ two separate CT scans, yielding two CT images. The comparison of these images facilitates the estimation of the spatial distribution of the two basis materials. Despite both scans employing a full angular sampling of at least 180 degrees around the object, this spatial sampling redundancy is essential. It is necessitated by the fact that image-based DECT methods rely on having two CT images as an intermediate result, which serves as input for the actual DECT image processing pipeline.

The update step (1) is inherently independent of the scanning trajectory. Consequently, if the angular sampling of the volume space proves sufficient, it might be feasible to utilize two adjacent 100-degree half scans. These half scans, when combined, provide the equivalent amount of measurements as a standard circular cone-beam scan.

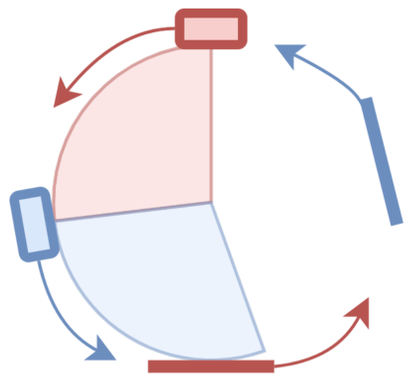


Fig. 1. The double half scan geometry is realized by splitting a 200-degree scan into a high and a low energy interval indicated by the different colors. For the simulations we used source-to-isocenter distances and detector geometries comparable to a flat-panel C-arm system.

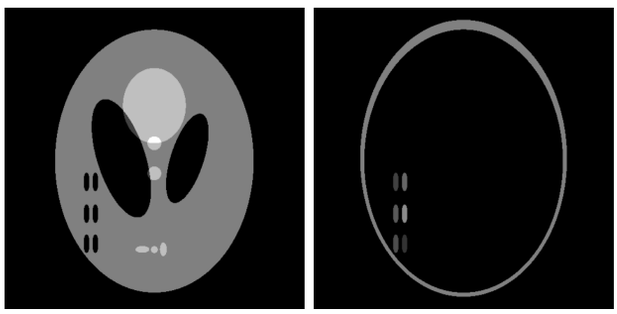
III. Materials

We employed a Shepp-Logan software phantom with a sampling of $512 \times 512 \times 10$ voxels, each measuring $0.5 \times 0.5 \times 2.0\text{mm}^3$. The detector consisted of 616×10 pixels, each measuring $6.16 \times 6.16\text{mm}^2$, and we acquired 500 measurements along a 200-degree circular trajectory. The central slice of the phantom is illustrated in Fig. 6. To assess the convergence properties of the relaxation choice in (3), we simulated two scans with the same geometry using high-energy spectra of 125kV and filtering materials of 0.9mm Al and 0.8mm Sn. The low-energy scan utilized 70kV and 0.9mm Al. We reconstructed the resulting concentrations of the basis materials using a $256 \times 256 \times 10$ grid with a resolution of $1\text{x}1\text{x}2\text{mm}^3$.

We tested the relaxation strategy (3) in three ways. Initially, we used a constant relaxation parameter R for all voxel coordinates, independent of the iteration. Secondly, we set $\alpha = 1$ irrespective of the iteration number but applied the non-uniform denominator. In the third setting, we adjusted α after each iteration that utilized the full angular measurements once. It began as a factor of 30 and reduced itself by 3 after each iteration, returning to 30 after 10 iterations according to Eq. (4).

To explore the effects of truncation, we reduced the number of detector pixels by approximately 30% to 440×10 pixels. For estimating the outer volume material distribution $c^1_{j'1}$, we utilized the high-energy scan and a model-based polychromatic reconstruction technique [6], [2] with an expanded volume space ($512 \times 512 \times 10$ grid and a resolution of $1 \times 1 \times 2\text{mm}^3$). It's important to note that this reconstruction method employs the same forward model (2).

Following reconstruction, we nullify the inner $256 \times 256 \times 10$ voxels and compute forward projections $\sum_j a_{ij} c_{jd}$. This is



(a) (b)

Fig. 2. the software phantom, comprising water and bone concentrations c_1 and c_2, illustrated in (a) and (b) respectively. The window level spans from 90% to 110%, allowing for the visualization of inlets with varying density values, approximately around 10 percent.

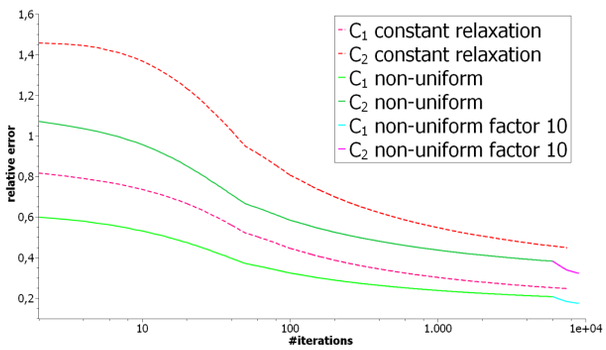


Fig. 3. Relative error as a function of the iteration number n applying the update step eq. (1).

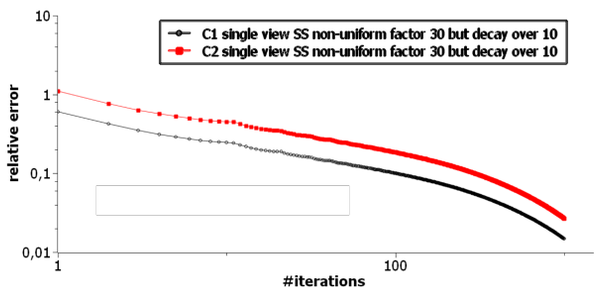


Fig. 4. Relative error as a function of the iteration number n applying the update step that includes relaxation $\alpha(n)$.

feasible because the model allows for swapping the summation order between the material index d and voxel number j, enabling the storage of this information in the projection space. The calculation of the energy integral E is performed later within the update step pipeline.

The simulation of the double half-scan geometry involves neglecting the second half of the low-energy scan and the first half of the high-energy scan. Consequently, the combined measurements can be represented by a single 200-degree circular fan-beam scan that undergoes a change in tube voltage and filtering at the 251-th view, following the settings described above.

IV. Results

A. Convergence Speed

If we initialize the reconstruction volumes with zeros and a constant relaxation parameter R, we observe slow convergence properties. After the first iteration, the update step attributes a significant portion of the observed attenuation in the measurements to the second basis material (cortical bone), owing to its higher effective mass attenuation coefficient. As depicted in Fig. 3, even after 1,000 iterations, the relative error remains above 50%, and it does not dip below this threshold even after 10,000 iterations.

Implementing the non-uniform relaxation of Eq. (3) while keeping α constant leads to a decrease in the relative error but does not address the persistently slow convergence, which still hovers above 40% for the second basis material.

Motivated by the non-uniform relaxation ideas of Kim et al.[5], we opted to amplify the step length by a constant factor of 10 after 6,000 iterations. This resulted in a substantial acceleration and altered the characteristics of the error as a function of the iteration number. The most effective approach to enhance convergence speed and achieve viable reconstruction results, as depicted in Eq. (4), is evident in Fig. 4, where the error approaches single-digit percentages after only a few hundred iterations. Notably, we found it crucial to make α a function of n since failure to do so resulted in strong divergence. It's essential to highlight that we applied only one view per subset, maximizing the number of update steps in one iteration.

B. Truncation

Reconstruction of truncated projection data induces the update step to distribute attenuation to outer regions that minimize the error in projection space. Due to the inherent ambiguity of the forward model in Eq. (2), the gradient descent converges to an estimation that allocates a significant portion of the attenuation to the higher attenuating basis material. Fig. 5 illustrates that nearly all the attenuation is attributed to the second basis material.

Utilizing a polychromatic reconstruction with the high-voltage projection dataset results in the concentration of the first basis material, distributed similarly in the outer region as depicted in Fig. 5(b). The proposed method yields more qualitatively robust results, evident in the small holes observed in the c_1 reconstruction, where corresponding spots in c_2 behave accordingly. Despite these improvements, the relative error for the particular slabs remains around 50%, rendering it insufficient for quantitative analysis even after 2,000 iterations. The values within the VOI of c_1 are similarly erroneous and cannot be reconstructed with an accuracy below a 40% relative error.

C. Double Half Scan Geometry

We initialized the volumes with zeros and executed 10,000 iterations of our proposed relaxed update step. The qualitative results exhibit coherence, but there is a challenge in discerning contrast in c_1 while concurrently overestimating the concentrations of the second basis material c_2.

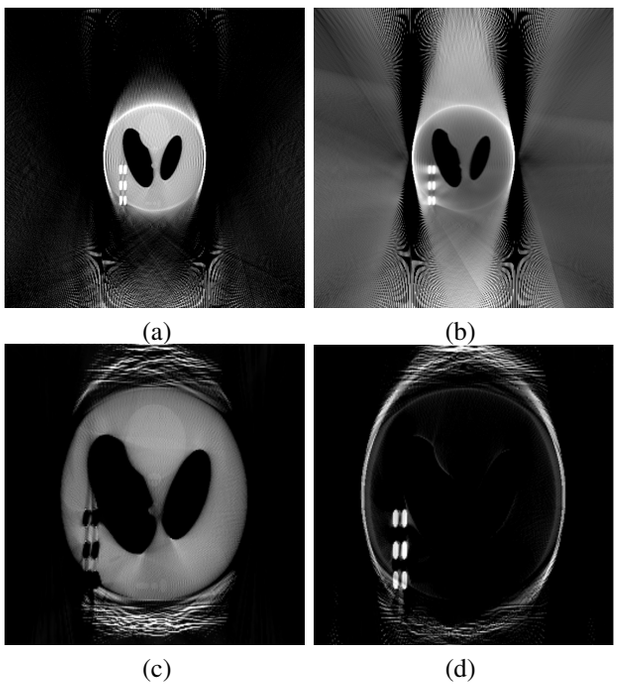


Fig. 5. Reconstructed values of the central slice of a reconstruction volume of 512x512x10 voxels with a resolution of $1 \times 1 \times 2.0\mathrm{mm}^3$ for (a) and (b). (c) and (d) are the results of the fixed outer volume that is estimated by a standard CT reconstruction and a constant offset to the foward projection step involved in the gradient descent update step. Window levels: (a) 0.0 to 1.15, (b) 1.0 to 4.0, (c) 0.0 to 1.1, (d) 0.0 to 1.25.

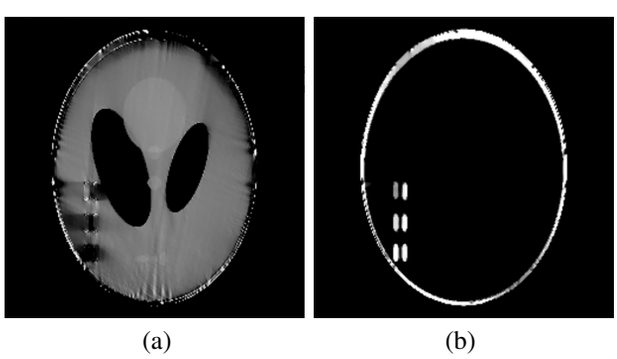


Fig. 6. The reconstructed values of basis material distributions with only half of the projection data show sufficient qualitative consistency. The window level ranges from 50% to 150%. The estimations of the second basis material c_2 overshoots by 25%.

V. Conclusion

We have adapted a model-based iterative dual-energy CT reconstruction technique designed for a flat-panel cone-beam setup. The incorporation of additional non-uniform relaxation, coupled with an interpolating step length between a factor of 30 and 1, has significantly enhanced the convergence of the gradient descent process. In scenarios involving truncation, the algorithm's ability to estimate an outer volume, enabling the correction of attenuation by the lower attenuating basis material in a manner consistent with truncated projections, facilitates the qualitative separation of both materials. This separation would be otherwise unattainable. Although the relative errors remain too high for precise quantitative analysis, the method demonstrates utility for segmentation purposes or can function as prior knowledge for other advanced image-processing pipelines. The influence of the fixed outer volume on the attenuation values of the first basis material requires further investigation, as it stands as a pivotal factor in maintaining accuracy. Specific image processing and fine tuning of the outer volume might help to fix this issue.

Moreover, our findings showcase the method's capability to reconstruct coherent material distributions even when utilizing only the first half of the low-energy scan and the second half of the high-energy scan, effectively halving the X-ray dose. This breakthrough has potential implications for advanced methods that may only require a modest increase in projection angles or are primarily concerned with qualitative estimations, such as material segmentation.

It's worth noting that our approach involves the use of step length amplifiers determined through trial-and-error. Further research in this direction is imperative to gain a deeper understanding of the potential for further accelerating iterative methods, particularly in situations where the cost function may converge toward incorrect minima.

Acknowledgment

The work of this paper is partly funded by the German Ministry of Education and Research (BMBF) within the Forschungscampus *STIMULATE* under grant number '13GW0473A'.

References

[1] R. Bismark, O. Beuing and G. Rose, *Truncation artifacts caused by the patient table in polyenergetic statistical reconstruction on real C-arm CT data*, in The 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine 2019.

[2] R. Bismark, R. Frysch, S. Abdurahman, . Beuing, M. Blessing and G. Rose, *Reduction of beam hardening artifacts on real C-arm CT data using polychromatic statistical image reconstruction*, in Zeitschrift für Medizinische Physik, 0939-3889 , 2019, URL: https://doi.org/10.1016/j.zemedi.2019.10.002

[3] A. Ziegler, T. Nielsen and M. Grass, *Iterative reconstruction of a region of interest for transmission tomography* , in Medical Physics 35 ,1317-1327, 2008.

[4] Samuel M. Johnston, and G. Johnson, G. Allan and Cristian T. Badea, *Temporal and spectral imaging with micro-CT*, in Medical Physics 39, 4943-4958, 2012.

[5] D. Kim, D. Pal, J. -B. Thibault and J. A. Fessler, *Accelerating Ordered Subsets Image Reconstruction for X-ray CT Using Spatially Nonuniform Optimization Transfer*, in IEEE Transactions on Medical Imaging, vol. 32, no. 11, pp. 1965-1978, 2013.

[6] Idris A. Elbakri and J. A. Fessler, *Statistical Image Reconstruction for Polyenergetic X-Ray Computed Tomography*, in IEEE Transactions on Medical Imaging, vol. 21, no. 2, pp. 89-99, 2002.

Performance study of novel silicon photomultiplier devices in scintillator-based photon-counting detector for X-ray CT application

Min Le, Chunguang Zong, Liang Li, *Senior Member, IEEE*

***Abstract*—The dominant architectures of photon-counting detectors for X-ray imaging are CdTe/CdZnTe and Si semiconductor detectors. These detectors have the advantages of small pixel size and high energy resolution. However, their high cost and limitations for high-energy X-ray detection have always been a problem. Photon counting detector based on scintillator detector coupled with silicon photomultiplier (SiPM) is one feasible implementation scheme for CT systems with large detection area and wide energy range. With the development of photoelectric sensor technology, the performance of novel SiPM devices has been improved, making it possible to use them in scintillator-based photon-counting detector for X-ray CT application. We aim to investigate a photon-counting detector based on SiPM-coupled scintillators for wide-energy-band industrial X-ray CT at 100 keV-15MeV. The performance of SiPM can greatly affect the energy response and count rate of the detector, and consequently the energy and spatial resolution of the CT system. This paper measured the performance of new commercial SiPM devices from SensL, HPK, and Broadcom, including dark current, single-photon amplitude, pulse shape, and time resolution, to provide the necessary technical support for subsequent detector design. Future work will couple different scintillator crystals to characterize the energy response and count rate performance of SiPM to determine its influence on the energy resolution and spatial resolution of the overall CT system, and ultimately to optimize the detector design.**

***Index Terms*—X-ray CT, photon-counting detector, scintillator, SiPM.**

I. Introduction

PHOTON-counting CT has developed rapidly in recent years and is widely used in many fields[1]. The photon-counting detector used in photon-counting CT is a novel type of detector, which has a different working mechanism compared with the traditional energy integration detector. It detects individual X-ray photons, analyzes and processes the signal generated by each photon, and obtains its energy and time information[2]. This detection approach has several advantages, including reduced electronic noise, more detailed X-ray spectral information, and the ability to time individual events.

Most of the detectors used in photon-counting CTs that are already in clinical use and in prototype CT systems under investigation are direct-conversion detectors based on semiconductor materials. Semiconductor detectors can generate very fast signals and have high energy resolution and high spatial resolution due to pixel subdivision. These advantages have made semiconductor detectors the mainstream choice for photon-counting detector design, and the most commonly used semiconductor materials at this stage are CdTe, CdZnTe, and Si[3]. However, there is still some resistance to the promotion of semiconductor detectors. On the one hand, the low atomic number and small density of Si-based detectors limit their application in the high-energy band. On the other hand, the high cost of CdTe and CdZnTe makes them too expensive for CT systems with large detection areas[4]. Therefore, another realization of photon-counting detectors has also been proposed: scintillator-based indirect conversion detectors. With the advent of various commercial fast scintillators and the development of SiPM-based photoelectric sensor technology, the scintillator detector coupled with SiPM has become feasible and has been widely used in the field of nuclear medicine (e.g. PET and SPECT). Recent studies suggest that SiPM-based scintillator detectors may be a viable alternative to semiconductor detectors for photon-counting CT[5], [6].

Our goal is to investigate a SiPM-based scintillator photon-counting detector for wide-energy-band industrial X-ray CT ranging from 100 keV to 15MeV. One of the possible applications of this detection system is the inspection of large containers. Because of the large size of the scanned object, the flux of X-rays through it to the detector is very small, and using photon-counting detectors significantly improves the signal-to-noise ratio and avoids signal pile-up caused by excessive photon flux. Furthermore, the energy spectrum information from multi-energy bins can assist in the material recognition. The performance of the SiPM device is a crucial factor that impacts both the energy and spatial resolution of the entire CT system. To guide the next design of detector, this article tested the performance of novel advanced SiPM from SensL, HPK and Broadcom, including dark current, single-photon amplitude and measured the pulse shape and time resolution of 100-photons signal, since the signal of 100 photons is more consistent with the signal generated by actual CT measurement. In future work, we will examine the energy response and count rate performance of SiPM coupled

This work was supported by the grants from Beijing Natural Science Foundation (L222001).

M. Le is with 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.

C. Zong is with Nuctech Co., Ltd, Beijing, China.

L. Li is with 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 (e-mail: lliang@tsinghua.edu.cn).

TABLE I
DETAILED INFORMATION ON THE SIPM DEVICES TESTED IN THE EXPERIMENT

SiPM type	Device size(mm^2)	Cell pitch(μm)	Number of cells	Terminal capacitance(pF)	Breakdown voltage(V)
AFBR-S4N44C013, Broadcom	4×4	30	15060	760	27
S14160-4050HS, HPK	4×4	50	6331	900	38
S14160-6050HS, HPK	6×6	50	14331	2000	38
MICROFC-10035, SensL	1×1	35	504	100	24.5
MICROFJ-30035, SensL	3×3	35	5676	1070	24.5
MICROFJ-40035, SensL	4×4	35	9260	1800	24.5
MICROFJ-60035, SensL	6×6	35	22292	4140	24.5

with different scintillators, evaluate its effect on the energy and spatial resolution of the total CT system, and eventually finalize the design of detector.

II. METHODS

A. Materials

To compare the performance of SiPM with different device types, the tests comprised various advanced commercial SiPMs from SensL, HPK and Broadcom. Table I provides details of the tested devices.

B. Experimental setup

The overview diagram shown in Fig 1 illustrates the experimental setup. A fast laser pulse with a wavelength of 405nm and a pulse width of 40ps (FWHM) was used as the experimental light source. We employed optical attenuator to regulate the intensity of the laser and diffused the laser pulse to uniformly illuminate the SiPM device. The entire setup was placed in a large dark box to prevent interference from natural light. The electrical synchronization signal of the laser was used as the trigger signal of the oscilloscope, and the original current signal of the SiPM was processed by the preamplifier circuit and then input to the oscilloscope. The preamplifier circuit based on THS4303 converts the current signal into voltage signal through 10 ohm load resistor and amplifies it to reduce the measurement error. Both the trigger signal and the amplified SiPM signal were digitized on a LeCroy oscilloscope Waverunner 9404M-MS with 4GHz bandwidth and 20Gs/s sampling rate, and subsequently fed to PC for off-line data analysis.

C. Data acquisition

1) Signal screening: Theoretically, the 100-photons signal amplitude is the single-photon amplitude (SPA) multiplied by 100, which is applied as the criterion for screening the signal. As for SPA measurements, the laser pulse was attenuated to a few photons level and the signal was acquired by oscilloscope at low bandwidth of 20 MHz to minimize the effect of noise. We turned on the persistence display of the oscilloscope and plotted the pulse amplitude spectrum, as shown in Fig 2, to facilitate observation and calculation. From the picture, single-photon to six-photons signal can be clearly distinguished and the difference in amplitude of adjacent signals is the SPA. In the pulse shape and time resolution experiments,

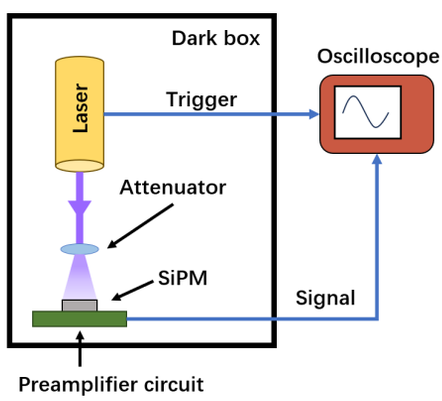


Fig. 1. Schematic diagram of the experimental setup.

the laser intensity was adjusted to approximately 100-photons level and 2000 triggered waveforms were saved for analysis. Considering dark counts, crosstalk, and electronic noise effects in practice, signals within the 97%-103% interval of 100 SPAs were screened out from these 2000 waveforms to obtain the 100-photons pulse shape and time resolution.

2) Pulse shape and Time resolution measurement: By averaging lots of screened waveforms, the 100-photons pulse shape can be easily attained. The 100-photons time resolution is defined as the timing jitter that determines the moment at which 100 visible photons enter the SiPM to generate the signal. Therefore, the measurement requires a trigger signal that is highly synchronized with the optical signal as a time baseline (the test used the electrical synchronization signal of the laser), and leading edge discrimination is used to obtain the arrival time of the 100-photons signal, and then the time distance between the two is counted to calculate the time resolution. To accurately calculate the time resolution,

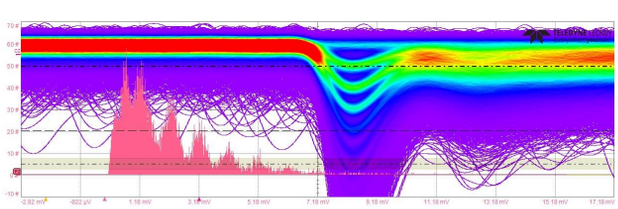


Fig. 2. Oscilloscope snapshots of SPA measurement for HPK S14160-4050HS device. Pulse amplitude spectrum is on the left; Waveforms under persistence display are on the right.

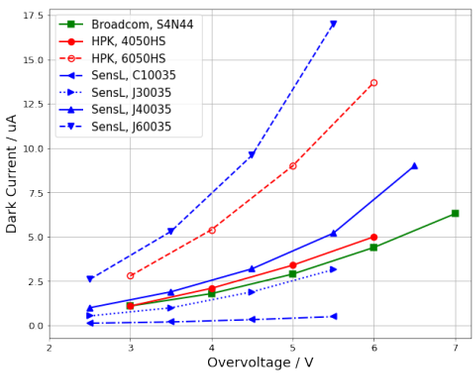


Fig. 3. Dark current versus overvoltage for different SiPMs

the timing jitter of the laser synchronization signal as the time baseline must be removed. We obtained a coincidence time resolution of $\sigma = 15\text{ps}$ by splitting the electrical laser synchronization signal into two channels and inputting to the oscilloscope. Thus, the timing jitter of the time baseline is $\sigma_j = \sigma/\sqrt{2} = 11\text{ps}$, which should be orthogonally subtracted from the 100-photons time resolution.

III. Results

A. Dark current

The dark current can intuitively reflect the noise performance of SiPM, and it is related to the overvoltage (overvoltage = bias voltage − breakdown voltage). In the experiments, we used a Keithley SourceMeter 2400 to provide bias voltage for SiPM and monitor current. The dark current of SiPMs were measured at different overvoltages, and the results are presented in Fig 3.

B. Pulse shape

The energy response and counting rate performance of SiPM are closely dependent on the pulse shape. We tested 100-photons pulse shape for all devices under typical operating bias and draw the original current pulse and also computed their rise time (20%-80%) and fall time (80%-20%). Fig 4a provides an overview of the pulse shape of 100 photons. It is evident that the pulse duration of all devices is less than 250ns, and the pulse height of smaller devices is higher. Fig 4b shows a zoomed view of the rising edge portion, illustrating that the rising edge of the smaller device is sharper. This can be explained by the terminal capacitance of the device. On the one hand, the smaller terminal capacitance increases the SPA ($V = Q/C$), resulting in a higher pulse height. On the other hand, the terminal capacitance can be regarded as an RC low-pass filter[7], and increasing the terminal capacitance will reduce the high-frequency component of the pulse, i.e., the fast component. More precise and detailed results are listed in Table II. For the same device size, the device of HPK has the highest pulse height and the shortest rise time and fall time.

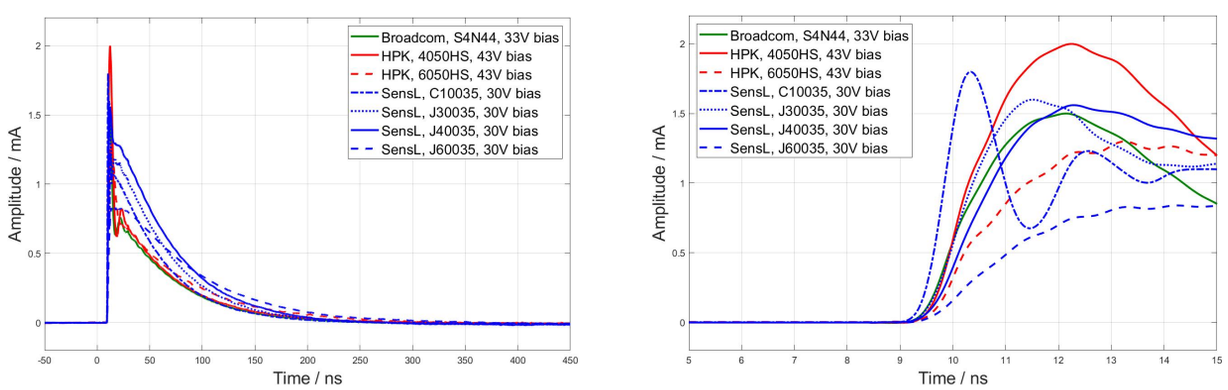
TABLE II
Rise time and fall time of 100-photons pulse shape for all devices

SiPM type and operating bias	Rise time (20%-80%)/ns	Fall time (80%-20%)/ns	SPA /μA
Broadcom, S4N44, 33V bias	1.2	85	15.0
HPK, 4050HS, 43V bias	1.1	80	20.0
HPK, 6050HS, 43V bias	1.6	90	13.0
SensL, C10035, 30V bias	0.5	65	18.0
SensL, J30035, 30V bias	0.9	79	16.0
SensL, J40035, 30V bias	1.2	90	15.6
SensL, J60035, 30V bias	1.9	97	8.4

C. Time resolution

All devices were tested at various discriminator thresholds, and we plotted the 100-Photons Time Resolution vs. Threshold curves to evaluate the time performance of SiPM. The experimental results of SiPMs under typical operating bias are shown in Fig 5. This images illustrate that SiPMs with smaller sizes exhibit better time performance, and that the time resolution first decreases and then increases with increasing threshold. Commonly, the time resolution can be simply expressed as[8]:

$$\sigma_t = \frac{\sigma_a}{\mathrm{d}V/\mathrm{d}t} \tag{1}$$



(a) (b)

Fig. 4. 100-phtons pulse shape (original current) for different SiPMs. (a): Full view of the entire waveform; (b): Zoom in on the rising edge region of the entire waveform.

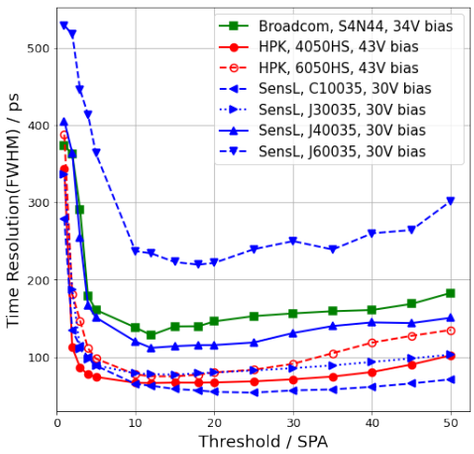


Fig. 5. 100-phtons time resolution versus discriminator threshold for different SiPMs.

Where σ_a denotes the standard deviation of the amplitude fluctuation and $\mathrm{d}V/\mathrm{d}t$ is the slope of the rising edge at the discriminator threshold. Under the same design architecture, increasing device size means larger terminal capacitance and more micro-cells, which leads to slower rising edges and higher dark count noise. The former reduces the slope ($\mathrm{d}V/\mathrm{d}t$) while the latter exacerbates amplitude fluctuations (σ_a), ultimately causing lower time resolution in large-size SiPM devices.

Regarding discriminator threshold, waveform baseline oscillation because of electronic noise and dark counts degrades timing jitter at low thresholds. By increasing the threshold, these disturbances are rapidly reduced, allowing the time resolution to reach an optimal value. As the discriminator threshold continues to increase, the slope of the rising edge will gradually decrease and the time resolution will slowly increase. Test results revealed that an optimal discriminator threshold between 10 and 20 SPAs provides the best 100-photons time resolution of SiPM. Table III summarizes the best 100-photons time resolution and the corresponding discriminator threshold for each device tested.

TABLE III
BEST 100-PHOTONS TIME RESOLUTION AND OPTIMAL DISCRIMINATOR THRESHOLD FOR ALL DEVICES

SiPM type and operating bias	Best 100-photons time resolution (FWHM) / ps	Optimal discriminator threshold / SPA
Broadcom, S4N44, 33V bias	127	12
HPK, 4050HS, 43V bias	65	15
HPK, 6050HS, 43V bias	75	15
SensL, C10035, 30V bias	53	20
SensL, J30035, 30V bias	77	15
SensL, J40035, 30V bias	108	20
SensL, J60035, 30V bias	207	18

IV. CONCLUSION

In this work, we surveyed the time performance of various novel SiPMs from SensL, HPK, and Broadcom with respect to dark current, single-photon amplitude, pulse shape, and time resolution. Based on the experimental results, it can be concluded that the smaller device performs better, exhibiting lower dark current, higher pulse amplitude, faster rising and falling edges, and smaller time resolution. When comparing fixed device sizes across different vendors, the design architecture of HPK device is more excellent and demonstrates superior performance. The current stage of our research is focused on the device. Subsequent work will involve coupling the scintillator crystal on SiPM to form a preliminary detector module, and investigate the energy response and counting rate performance of the detector module. The final goal is to build a complete CT detector system, as seen in Fig 6, realize imaging on this system, and analyze the time resolution and spatial resolution of imaging.

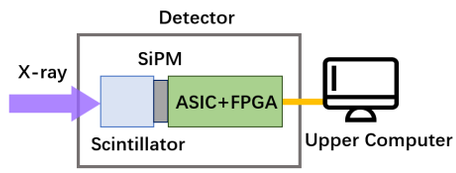


Fig. 6. Schematic of the final CT detector system

REFERENCES

[1] 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.

[2] T. Flohr, M. Petersilka, A. Henning, S. Ulzheimer, J. Ferda, and B. Schmidt, "Photon-counting ct review," *Physica Medica*, vol. 79, pp. 126–136, 2020, 125 Years of X-Rays. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1120179720302738

[3] R. Ballabriga, J. Alozy, F. N. Bandi, M. Campbell, N. Egidos, J. M. Fernandez-Tenllado, E. H. M. Heijne, I. Kremastiotis, X. Llopart, B. J. Madsen, D. Pennicard, V. Sriskaran, and L. Tlustos, "Photon counting detectors for x-ray imaging with emphasis on ct," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 4, pp. 422–440, 2021.

[4] L. Bläckberg, M. Moebius, N. Moghadam, D. Uzun-Ozsahin, E. Mazur, G. El Fakhri, and H. Sabet, "Scintillator-based photon counting detector: Is it feasible?" in *2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD)*, 2016, pp. 1–5.

[5] N. Tuccori and S. J. M. Peeters, "Photon-counting computed tomography and scintillator-based detectors: A simulation analysis with scintillating and reflecting materials currently on the market," *IEEE Transactions on Nuclear Science*, vol. 70, no. 7, pp. 1404–1412, 2023.

[6] S. J. van der Sar, S. E. Brunner, and D. R. Schaart, "Silicon photomultiplier-based scintillation detectors for photon-counting ct: A feasibility study," *Medical physics*, vol. 48, no. 10, pp. 6324–6338, 2021.

[7] F. Acerbi and S. Gundacker, "Understanding and simulating sipms," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 926, pp. 16–35, 2019.

[8] A. Gola, C. Piemonte, and A. Tarolli, "Analog circuit for timing measurements with large area sipms coupled to lyso crystals," *IEEE Transactions on nuclear science*, vol. 60, no. 2, pp. 1296–1302, 2013.

Calibration-based Statistical Framework for Material Decomposition in Photon Counting CT

Okkyun Lee and Joonbeom Kim

***Abstract*—This study aims to develop a calibration-based statistical framework for accurate and precise material decomposition in a photon counting detector (PCD)-based computed tomography (CT). The first step of the framework is to obtain a pseudo-counts generator (PCG). It is a pre-processing step using calibration slab phantoms. It approximates the mean photon counts for a given basis line integrals for each energy bin and PCD pixel. The second step is the maximum likelihood (ML)-based estimation of basis line integrals using the PCG. The estimates are accurate due to the statistical optimum property of the ML, but it suffers from a substantial noise level. The final step improves the precision by applying a likelihood-based bilateral filter (LBF) to the ML-based estimates from the previous step. It calculates statistical weights from the neighborhood estimates and uses them in the weighted average to reduce the noise effectively. We compared the three steps in the proposed framework to the conventional forward modelings, calibration-based material decomposition, and iterative reconstruction method with edge-preserving regularization using a bench-top photon counting cone-beam CT (CBCT) system. The experimental study demonstrates that the proposed method shows the most accurate forward modeling and material decomposition and affects the spatial resolution and tissue textures during the noise reduction less than the conventional methods.**

***Index Terms*—Maximum likelihood, likelihood-based bilateral filter, iterative reconstruction, material decomposition, photon counting CBCT**

I. Introduction

ACCURATE and precise material decomposition is necessary for many photon-counting computed tomography (PCCT) applications, such as coronary artery calcification, beam-hardening artifacts reduction, and K-edge imaging [1–4]. For this purpose, deep learning-based approaches have recently emerged as an active research topic [5–7]. However, the performance depends on the training data, and the general applicability needs to be further studied. This study focuses on the conventional approaches, divided into model-based and calibration-based methods.

The model-based method uses energy-dependent system models such as incident x-ray spectrum and the detector response functions to estimate basis line integrals or images [1, 8]. It searches for a solution that maximizes the likelihood with a given measurement to obtain accurate estimates. It also reduces the difference between neighborhood estimates to improve the precision by applying edge-preserving regularization [2]. However, it is challenging to calculate the accurate energy-dependent system models for each detector pixel and energy bin. Furthermore, the regularization may degrade spatial resolution and background tissue structures.

On the other hand, the calibration-based method collects a set of calibration measurements for the estimation [3, 9]. It is computationally efficient, and the system models are not required. However, how to design the estimator and calibration phantoms is an open question, and the accuracy of the estimates is not guaranteed. Integrating the calibration-based method with a noise reduction process is also challenging.

This study introduces a calibration-based statistical framework that aims for accurate and precise estimation of basis line integrals without system models. It consists of three steps: (1) We make a function of basis line integrals that approximates the mean photon counts. We collect calibration measurements (function values) associated with the thicknesses of the basis materials (function arguments) using slab phantoms. It interpolates or extrapolates the measurements to approximate the mean photon counts for a given basis line integral. We call the function a pseudo-counts generator (PCG) and obtain it for each detector pixel and energy bin. (2) Using the PCG, one can calculate the likelihood of the basis line integrals for a given PCD measurement. Thus, the maximum likelihood (ML)-based estimates can be obtained. The estimates are accurate due to the asymptotic optimal property of the ML, but it suffers from a substantial noise level. (3) We apply the recently developed likelihood-based bilateral filter (LBF) [10] to the ML-based estimates from the second step. The statistical weights in the LBF need to be calculated using neighborhood likelihoods, and the PCG can be used for this purpose.

We demonstrate the accuracy of the PCG compared to other conventional forward models using test slab phantoms. We also demonstrate the accuracy and precision of the second and third steps of the proposed method compared to conventional calibration-based material decompositions and iterative reconstruction method with edge-preserved regularization using calcium hydroxyapatite (CaHA) phantom and chicken leg experiments.

II. Methods

A. Proposed Statistical Framework

1) [1st Step] Pseudo-Counts Generator (PCG): The first step of the proposed framework is to obtain the PCG, the

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00249223).

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).

J. Kim was with the Department of Robotics and Mechatronics Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu, Republic of Korea.

function that approximates the mean photon counts, using calibration slab phantoms. We define the function as $PCG_b(\mathbf{x})$ as follows:

$$\bar{p}_b(\mathbf{x}) = -\ln \frac{\bar{y}_b(\mathbf{x})}{\bar{y}_b(\mathbf{0})} \approx PCG_b(\mathbf{x}), \tag{1}$$

for $b = 1, 2, \cdots, N_b$, where N_b is the number of energy bins. $\bar{y}_b(\mathbf{x})$ is the mean photon counts at the b-th energy bin for a given vector of basis line integrals $\mathbf{x} = [x_1, x_2, \cdots, x_{N_m}]^T$, where N_m is the number of basis materials. Due to the near-linear property of the log-normalized data [9], we approximate the log-normalized data in (1) and calculate the mean photon counts as follows:

$$\hat{\bar{y}}_b(\mathbf{x}) \approx \exp\{-PCG_b(\mathbf{x})\}\bar{y}_b(\mathbf{0}). \tag{2}$$

We intentionally omitted the ray dependency to make the notation simple. We obtained the PCGs for all detector pixels and energy bins. It will be frequently used during material decomposition and noise reduction (second and third steps).

We used slab phantoms of PMMA and aluminum (AL) to obtain a set of calibration measurements $\{\bar{p}_b(\mathbf{x}_c)\}_{c=1}^{N_c}$, where N_c is the number of calibration points. We used PMMA of 0, 20, and 40 mm and AL of 0, 4, 8, 12, and 16 mm; thus, $N_c = 15$. The theoretical expectation of the photon counts was calculated by averaging multiple scans of the slab phantoms. Once $\{\bar{p}_b(\mathbf{x}_c)\}_{c=1}^{N_c}$ is obtained, the PCG interpolates or extrapolates the data to approximate $\bar{p}_b(\mathbf{x})$ for an arbitrary $\mathbf{x}$. We applied spline interpolation for this purpose. Figure 1 shows the calibration measurements for two energy bins (red stars) and examples of $PCG_b(\mathbf{x})$ (black points), interpolated or extrapolated data at various combinations of basis materials.

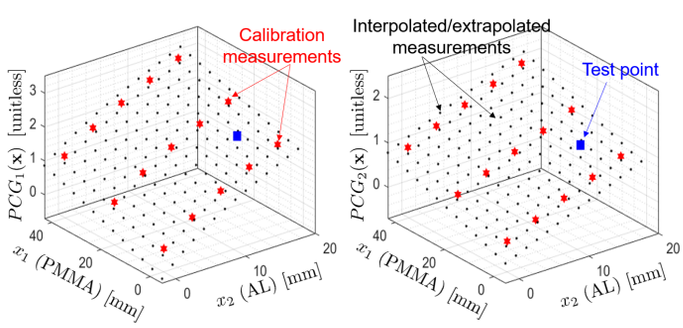


Fig. 1. An example of the PCGs at the center detector pixel. (Left: low energy bin, right: high energy bin; Red star: calibration measurements, black point: interpolated/extrapolated data, blue square: test data used to compare various forward modelings).

2) [2nd Step] Maximum Likelihood (ML)-based Estimator: The log-likelihood for a given PCD measurement is given by,

$$\mathcal{L}(\mathbf{x}|\mathbf{y}) = \sum_{b=1}^{N_b} \{-\bar{y}_b(\mathbf{x}) + y_b \ln(\bar{y}_b(\mathbf{x}))\}, \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^{N_b}$ is the vector of counts, including the quantum noise. The likelihood is the function of basis line integrals. We calculate it using the PCGs by approximating the mean counts, $\bar{y}_b(\mathbf{x})$. One can numerically calculate the derivative of the likelihood or use the derivative-free optimization solver such as Nelder-Mead [11]. However, it takes a substantial amount of time to estimate for all data. We applied the quadratic surrogate (QS)-based approach [12] to accelerate the computation and developed the method using the PCGs. We omit the details for brevity. All the results shown here were from the accelerated version; we also confirmed that the results were almost identical to those using the Nelder-Mead method except for the substantially reduced computation time.

3) [3rd Step] Likelihood-based Bilateral Filtering (LBF): The LBF was previously developed to reduce the noise in the ML-based estimates of basis sinograms using statistical weighting coefficients [10]. However, it was only validated using simulation studies. In this study, we apply it to the experimental data using the PCGs. The noise-reduced basis line integral at the j-th ray is given by,

$$\hat{\mathbf{x}}_j^{LBF} = C(j) \sum_{i \in \mathcal{N}_j} \mathcal{G}(i, j, \sigma_1) \mathcal{S}(i, j, \sigma_2) \hat{\mathbf{x}}_i^{ML}, \tag{4}$$

where $\hat{\mathbf{x}}_i^{ML}$ is the ML-based estimates for the i-th ray from the second step, $\mathcal{N}_j$ is the set of indices for the neighborhood rays (j is the center one), $C(j)$ is a normalization factor to make the sum of the weights to one, and $\mathcal{G}(i, j, \sigma_1)$ is the Gaussian kernel with a variance of σ_1^2. Lastly, $\mathcal{S}(i, j, \sigma_2)$ is the statistical weight given by,

$$\mathcal{S}(i, j, \sigma_2) = \exp\left\{-\frac{\left(\mathcal{L}(\hat{\mathbf{x}}_i^{ML}|\mathbf{y}_j) - \mathcal{L}(\hat{\mathbf{x}}_j^{ML}|\mathbf{y}_j)\right)^2}{\sigma_2^2}\right\}, \tag{5}$$

where σ_2 controls the degree of smoothing of the filter. The statistical analysis of the LBF and its variations can be found in [10]. When $\sigma_2 \to \infty$, it becomes the conventional low-pass filter (LPF). We also compared the LBF to the LPF.

B. Conventional Methods

1) Forward Modelings: We compared the first step of the proposed framework (PCG) to the conventional forward modelings of the polynomial expansion-based one (POLY) [13], semi-empirical model (SEM) [14], and the projected equivalent thickness calibration (PETC) method [3]. We used the same calibration data used in the proposed method for the conventional forward modelings.

2) Calibration-based Material Decomposition: We compared the second step of the proposed method (ML-based estimation) to the conventional polynomial expansion-based material decomposition. We also call it POLY (the input and output of the polynomial expansion-based estimator are reversed from the POLY, the forward modeling). We also implemented the PETC method for the material decomposition. It was less accurate than the proposed method and is not shown here for brevity.

3) Statistical Iterative Reconstruction (SIR): We compared the third step of the proposed method (LBF) to the conventional iterative method using edge-preserved regularization [2]. We call it SIR. We used the pseudo-Huber penalty that depends on the parameter δ, which determines the shape of the penalty function. We also set the regularization parameter β to balance the data fidelity and the regularization terms. In the data fidelity term, we fit the projection of the updated basis

images to the ML-based estimates from the second step of the proposed method. Thus, the accuracy of the SIR follows that of the proposed method. We focus on comparing how effectively it reduces the noise with less affecting the resolution and background tissue.

C. Experimental Settings

We used a micro-focus X-ray source (L9181-02, Hamamatsu) with 130 kV with a collimator and AL prefilter (2 mm), as shown in Fig. 2. We used a PCD array (XC-THOR FX, Direct Conversion) with an active area of $50.6 \times 100.2\,\mathrm{mm}^2$, where each PCD has a pixel size of $100 \times 100\,\mu\mathrm{m}^2$, with two energy thresholds (20 and 50 keV; $N_b = 2$). The distance from the source to the center was 36.9 cm, and the center to the detector was 30.94 cm. The number of views was 900 per 2π, and the total exposure was 15 mAs. After 2×2 detector binning, the rows and channels were 253 and 501, respectively. The volume image was reconstructed with an array of $512 \times 512 \times 375$ voxels with an $80 \times 80 \times 80\,\mu\mathrm{m}^3$ size.

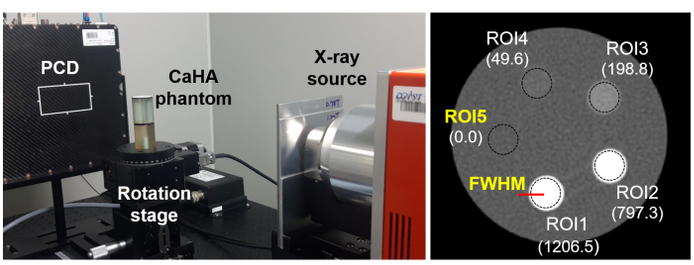


Fig. 2. The bench-top photon counting cone beam CT (CBCT) system and the CaHA phantom used for the validation. Values in the parenthesis indicate the ground truth of the CaHA densities in $\mathrm{mg/cm}^3$. Each ROI has a thickness of 7.2 mm in the vertical direction. The diameter of the phantom is about 32 mm.

We performed two-material decomposition (PMMA and AL) and then transformed the two into epoxy resin and CaHA (by the least squares fitting with attenuation curves of the materials). We focused on the analysis of CaHA for brevity.

III. Results

Figure 3 presents relative error maps (lower energy bin) using various forward modelings. The proposed method (PCG in the first step) shows the smallest error throughout the detector pixels compared to other methods. The higher energy bin showed smaller errors than the lower energy bin for all methods and is not shown here.

Figure 4 illustrates the CaHA images obtained from the POLY and the proposed method (ML; the second step). Both show a substantial noise due to the lack of noise reduction step. The proposed method shows a smaller bias than the POLY for all ROIs. Noise is substantially reduced in the LBF (the third step) and the SIR. The accuracy of the CaHA densities is almost preserved since both methods used the results of the ML (proposed; the second step).

Figure 5 presents the noise-FHWM plots (left) and the 1D NPS (right) for various methods. All the methods show the trade-off between noise and resolution. The LPF and SIR(0.5)

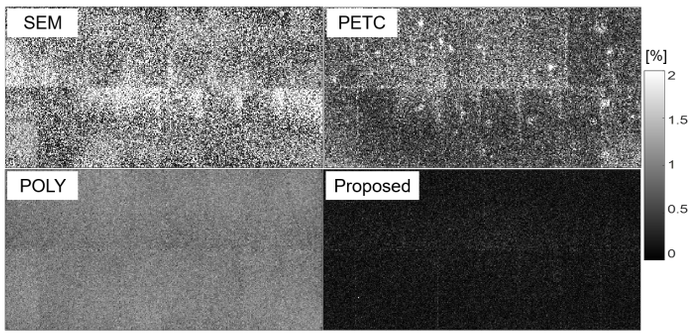


Fig. 3. Relative error maps (# of rows × # of channels) using the various forward modelings for the lower energy bin. Each value in the map is the relative error calculated between the ground truth of $\bar{y}_b(\mathbf{x}_t)$ and its approximation at a specific PCD pixel, where $\mathbf{x}_t$ indicates the test slab phantoms (PMMA 10 mm, AL 14 mm; see the blue square in Fig. 1). It showed a similar tendency for other test phantoms and is not shown here.

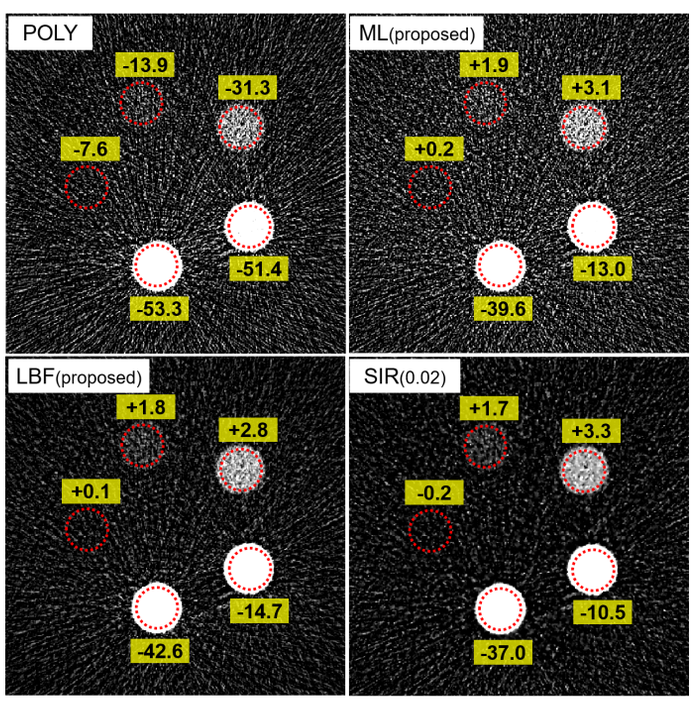


Fig. 4. The CaHA images (horizontal plane; slice thickness of 0.24 mm) reconstructed using various methods. The value in the yellow box indicates the bias (in $[\mathrm{mg/cm}^3]$) at the associated ROI. The value in the parenthesis of the SIR is of the parameter δ. The CaHA images of the LBF(proposed) and the SIR have almost the same noise level at ROI5 (see Fig. 2).

show larger FWHMs than others for the same noise level. The resolution improves as δ decreases in the SIR, and the two ($\delta = 0.02$ and 0.004) present almost the same tendency. The LBF (third step) shows a comparable tendency to the SIR. In the 1D NPS plots, the ML (proposed; the second step) shows the most widespread NPS along the frequencies since it lacks the noise suppression step. The NPS of the LPF was similar to the SIR(0.5), and the values in the low frequencies gradually decrease as δ decreases in the SIR. The proposed method (LBF) shows a more widespread NPS than others.

The CaHA images from the chicken leg experiment are presented in Fig. 6. The SIR and LBF substantially reduced the noise. However, the SIR(0.5) smooths the details in the tissue.

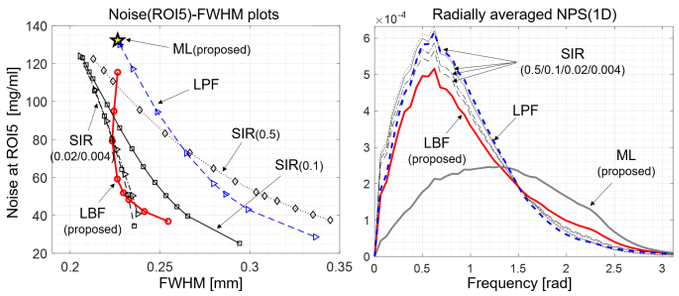


Fig. 5. (Left) Noise (at ROI5)-FWHM (full width at half maximum; see Fig. 2) curves using various methods were obtained by performing a parameter sweep for each method. The value in the parenthesis of the SIR is of the parameter δ. (Right) Radially averaged NPS (noise power spectrum). 2D NPS was calculated at a homogeneous region of the CaHA image, averaged in the z-axis (depth), normalized to have the sum to one, and then radially averaged. The noise level at the homogeneous volume was almost identical for all methods (except the ML).

The SIR(0.004) shows a sharper edge than the SIR(0.5), but the background tissue presents cartoon-like structures (see the zoomed images). On the other hand, the LBF preserved more details in the tissue compared to the SIR.

Table I summarizes the run time for the material decomposition and noise reduction process in the chicken leg experiment. The proposed ML-based estimation with the PCG took about 6 hours (Nelder-Mead algorithm). However, the QS-based approach substantially accelerated the process and took less than 5 minutes. The SIR and LBF are additional processes after the ML. The SIR took about an hour, but the proposed LBF took only about 2 minutes.

TABLE I
SUMMARY OF THE RUN TIME [SEC] FOR THE CHICKEN LEG EXPERIMENT.

Methods	Run time [sec]
ML(proposed)	21,838
ML(proposed; QS-based acceleration)	293
SIR	3,982
LBF(proposed)	125
Data dimension	$253\times501\times900\times2$ (# of rows × channels × views × bins)
Image size	$512\times512\times375$
Hardware specfication	Intel i7-8700K (6 cores), 64GB RAM, GTX 1080 Ti

IV. Conclusion

We introduced a calibration-based statistical framework for the material decomposition in photon counting CT. Energy-dependent system models are not required for this method. The pseudo-counts generator obtained from the calibration process enabled the maximum likelihood-based estimation followed by the likelihood-based bilateral filtering. Experimental studies showed that the PCG was more accurate than conventional forward models, and the proposed ML-based estimation was more accurate than conventional calibration-based material decompositions. The last step of the framework (LBF) was more effective and efficient for reducing noise than the edge-preserved regularization.

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, pp. 100 901(1–19), Oct. 2013.

[2] C. O. Schirra *et al.*, "Statistical reconstruction of material decomposed data in spectral CT," *IEEE Trans. Med. Imag.*, vol. 32, no. 7, pp. 1249–1257, July 2013.

[3] M. A. K. Juntunen *et al.*, "Framework for photon counting quantitative material decomposition," *IEEE Trans. Med. Imag.*, vol. 39, no. 1, pp. 35–47, 2020.

[4] O. Lee *et al.*, "X-ray transmittance modeling-based material decomposition using a photon-counting detector CT system," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 5, no. 4, pp. 508–516, July 2021.

[5] J. F. P. J. Abascal *et al.*, "Material decomposition in spectral CT using deep learning: A Sim2Real transfer approach," *IEEE Access*, vol. 9, pp. 25 632–25 647, 2021.

[6] H. Karlsson *et al.*, "Deep-learning-based denoising for photon-counting CT: Image domain or projection domain?" in *Proc. SPIE, Medical Imaging*, vol. 12031, April 2022, pp. 120 312S:1–18.

[7] R. Nadkarni, A. Allphin, D. P. Clark, and C. T. Badea, "Material decomposition from photon-counting CT using a convolutional neural network and energy-integrating CT training labels," *Phys. Med. Biol.*, vol. 67, pp. 155 003:1–13, 2022.

[8] 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, pp. 4031–4047, 2008.

[9] R. E. Alvarez, "Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis," *Med. Phys.*, vol. 38, no. 5, pp. 2324–2334, May 2011.

[10] O. Lee, "Likelihood-based bilateral filters for pre-estimated basis sinograms using photon-counting CT," *Med. Phys.*, vol. 50, pp. 2733–2758, 2023.

[11] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, "Convergence properties of the Nelder-Mead simplex method in low dimensions," *SIAM J. Optim.*, vol. 9, no. 1, pp. 112–147, 1998.

[12] Y. Long and J. A. Fessler, "Multi-material decomposition using statistical image reconstruction for spectral CT," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1614–1626, 2014.

[13] K. Mechlem *et al.*, "Joint statistical iterative material image reconstruction for spectral computed tomography using a semi-empirical forward model," *IEEE Trans. Med. Imag.*, vol. 37, no. 1, pp. 68–80, Jan. 2018.

[14] ——, "Spectral angiography material decompoisition using an empirical forward model and a dictionary-based regularization," *IEEE Trans. Med. Imag.*, vol. 37, no. 10, pp. 2298–2309, Oct. 2018.

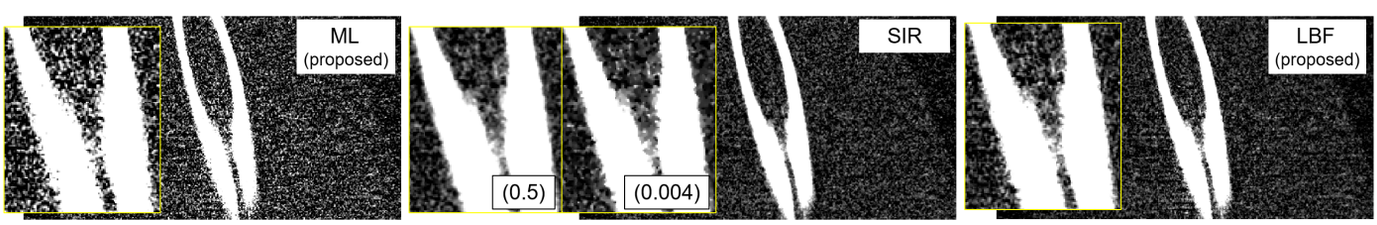


Fig. 6. The CaHA images (vertical plane; slice thickness of 0.08 mm) of the chicken leg experiment using the ML (proposed; without noise suppression), SIR, and LBF (proposed). The zoomed image of the center part is overlaid on the left-hand side (yellow box). For the SIR, two images are shown with different δs (see the value in the parenthesis). The SIR and LBF have almost the same noise level at a near-homogeneous volume.

Partial Angle Motion Compensation for Maxillofacial Cone Beam CT

Abdul Salam Rasmi Asraf Ali, Cristina Sarti, Claudio Landi, and Andrea Fusiello

***Abstract*—Cone Beam Computed Tomography (CBCT) has become essential in dentistry, offering a detailed three-dimensional view of the maxillofacial region. Despite its widespread use, patient motion during CBCT scans poses a significant challenge, leading to motion artifacts in the reconstructions. Existing compensation methods often involve iterative algorithms. In this paper, we build upon our previous work and introduce a novel incremental two-stage partial angle compensation approach to address issues in short-scan reconstructions. Our method begins with an ultra-short-scan reference volume, gradually incorporating additional projections. Regularization techniques enhance the reference reconstruction, mitigating artifacts. The proposed method shows improved image quality in most practical and challenging scenarios. Besides performance, the incremental approach also demonstrates a stable behavior in all tested cases, making it a promising approach for clinical usage. Evaluation metrics, including Structural Similarity Index (SSIM) and Root Mean Square Error (RMSE), highlight substantial enhancements. This approach holds potential for applications involving prolonged and persistent patient movements, providing a valuable contribution to motion compensation in CBCT imaging.**

***Index Terms*—Dental CBCT, motion compensation, partial angle compensation, FDK reconstruction.**

I. Introduction

Cone Beam Computed Tomography (CBCT) is a three-dimensional imaging method that is commonly used in dentistry, providing a way to view the maxillofacial area from any angle, not just the default views [1]. The usage of CBCT has increased significantly in various fields such as orthodontics, endodontics, oral surgery, and restorative dentistry [2]. Unlike clinical helical computed tomography, CBCT systems have acquisition time ranging from 5.4 to 40 seconds [3], which is long enough for significant patient motion to occur. Patients move during the scan for various reasons such as fear of the tube/detector movement [4], people having Parkinson's disease [5], or simply because it is difficult for someone to stay idle for quite a long time, especially children [6]. It is estimated that approximately 21–42% of the in vivo examinations exhibit motion artifacts [6].

Attempts have been undertaken to prevent patients from moving during CBCT acquisition. A common solution is immobilizing the patient using a head strap, chin rest, or a fixed bite block. However, depending only on hardware solutions may not always be successful in avoiding all possible motions [7], since even a small amount of motion displacement of $3mm$ can have a significant influence on the image quality [8]. Several works have been done in the past to compensate for patient motion, e.g., based on a 2D-3D registration process [9]–[11], based on the optimization of data fidelity terms such as local edge entropy [12], [13] or auto-calibration based approaches [14]. However, these solutions often involve complex algorithms that require multiple reconstructions of the complete set of projections, making the process time-consuming [11], [14], [15]. Additionally, these approaches may suffer from inconsistencies resulting from data truncation and scatter. These are common issues encountered in CBCT imaging, which make iterative algorithm convergence more difficult. [16].

In a previous paper [17], we introduced a method for compensating motion in a full 360° scan without needing multiple iterations, effectively addressing the issues mentioned earlier. Expanding on the findings of the previous work, in another paper [18], we proposed a modification to our original approach for handling patient movements in short-scan reconstructions (covering 180° plus the fan angle) with a two-stage approach. Short scans are generally preferred in clinical situations since only the minimum number of projections is collected to achieve good diagnostic quality. This, usually, means a shorter scan time and less dose for the patient. Also, CBCT devices that are designed for short-scan acquisitions have smaller footprints.

Despite these advantages, short-scan scenarios become more complex due to additional artifacts (referred to as short-scan artifacts) arising from the use of fewer projections. Motion can also significantly impact these scenarios due to the reduced number of projections. Although the reduction in projections is beneficial for faster scans and less patient movement, it can lead to increased artifacts in the presence of even minimal motion. In cases where a significant number of projections are impacted by motion in a short-scan reconstruction (as illustrated in Figure 3), the compensation algorithm presented in [18] encounters limitations. To tackle this issue, we introduced the partial angle compensation approach [19], replacing the first stage of motion compensation with an incremental compensation strategy.

II. Methodology

The workflow of the proposed partial angle compensation method is illustrated in Figure 1. A very small subset of nearly motion-free projections is considered to reconstruct an ultra-short-scan reference volume. We have chosen to use a set at the beginning of the scan, for the practical

Abdul Salam Rasmi Asraf Ali is a Ph.D. student at the University of Udine, Italy (e-mail: asrafali.abdulsalamrasmi@spes.uniud.it).

Cristina Sarti is an Imaging Engineer at See Through s.r.l., Brusaporto, Italy (e-mail: cristina.sarti@seethrough.one).

Claudio Landi is the head of the Imaging Team at See Through s.r.l., Brusaporto, Italy (e-mail: claudio.landi@seethrough.one).

Andrea Fusiello is a Professor at the University of Udine, Italy (e-mail: andrea.fusiello@uniud.it).

reason that we observed the motion to be less likely at the very beginning of the scan. Starting with the ultra-short-scan reference volume, we gradually incorporate a small fraction of the remaining projections, reconstructing a new reference volume, and compensating these additional projections as they are included. Each time, the reference volume is regularized using a threshold-based segmentation as explained in Section II-A. After the first stage of compensation using a partial angle scheme, an intermediate reconstruction is performed which is used as the reference volume for the second stage. In the second stage, a one-shot compensation is performed by using the same reference volume for the entire projections. A final high-quality reconstruction is performed using the motion parameters estimated in the second stage.

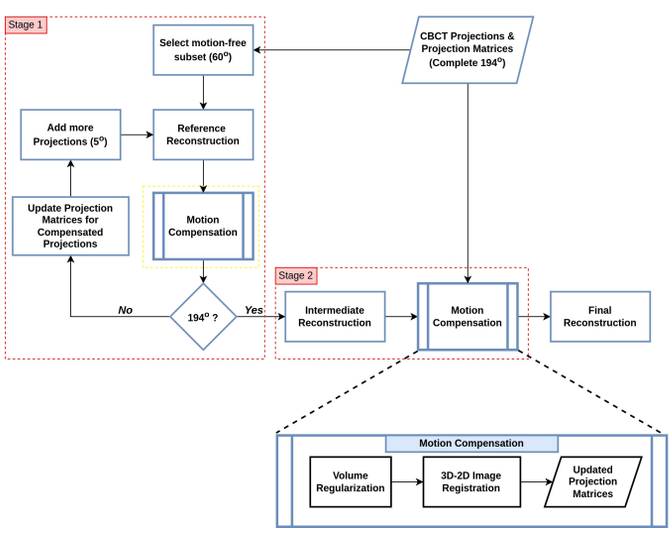


Fig. 1: Pipeline of the proposed method

Our experiments based on the actual device geometry revealed that the minimum number of projections needed for the ultra-short-scan reference volume is 60° of projections, which is less than one-third of the scan. This necessity arises from the absence of strong features in the re-projected ultra-short-scan volume when the angle is below 60°. Furthermore, we observed that adding 5 degrees of projections at a time proved effective for compensation.

A. Regularization

Regularization of the reference reconstruction is necessary for three main reasons. Firstly, as shown in Figure 2a, artifacts arising from the limited angle reconstruction affect the quality of the reference reconstruction leading to ambiguous overlapping structures during reprojection that are absent in real projections. Secondly, small structures in 2D projections are highly sensitive to motion, causing them to appear and disappear with each optimization step of the motion parameter modification. Thirdly, Poisson noise in the original projections further weakens low-contrast structures.

Consequently, reprojecting the reference reconstruction and registering them directly with original projections becomes challenging. To address this issue, we focus on identifying the most prominent structures, such as jawbones and teeth, by employing implicit regularization with a threshold-based segmentation of the reference reconstruction as shown in Figure 2b. During this particular step, our focus is on segmenting solely the hard tissues within the reference reconstruction. To accomplish this, we substitute any attenuation values exceeding a specific threshold with a constant value that closely approximates the average attenuation of the hard tissues. Any values below the threshold are discarded. This ensures only rigid structures are remaining in the reprojection (Figure 2c).

B. Similarity Cost Function

After the regularization step, the extracted strong edges are reprojected to generate the synthetic projections. We can therefore use a gradient strategy for image registration. We define our cost function based on Gradient Information, $G(r, a)$, which is the sum of the minima of gradient magnitudes computed for corresponding pixels in the synthetic image (r) and the actual image (a). This function is multiplied by a weighting function w and can be expressed as follows:

$$G(r,a) = \sum_{(\mathbf{x},\mathbf{x}')\in(r\ \cap\ a)} w_{\mathbf{x},\mathbf{x}'}\ \min\left(|\nabla\mathbf{x}|, |\nabla\mathbf{x}'|\right) \quad (1)$$

Where,

- $\mathbf{x}$ and $\mathbf{x}'$ refer to the corresponding sample points (pixels) in r and a, respectively.
- $|\nabla\mathbf{x}|$ and $|\nabla\mathbf{x}'|$ are the gradient magnitudes of $\mathbf{x}$ and $\mathbf{x}'$, respectively.

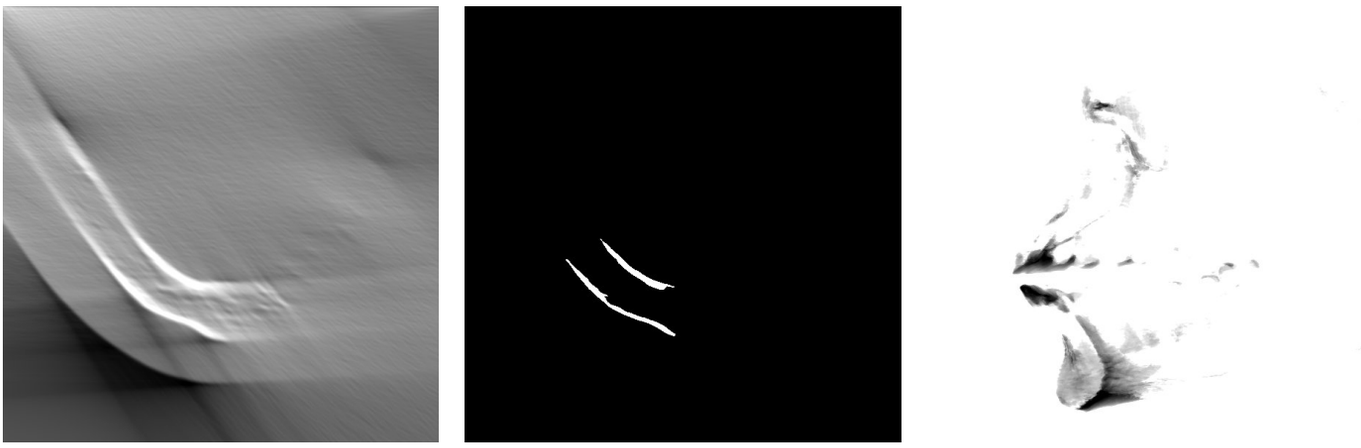
The minima operator is applied to exclude strong extraneous gradients [20]. The weighting function w in Equation 1 is the gradient orientation which is computed as the cosine of the angle between each corresponding gradient of r and a expressed as follows:

$$w_{\mathbf{x},\mathbf{x}'} = \cos(\theta) = \frac{\nabla\mathbf{x}\cdot\nabla\mathbf{x}'}{|\nabla\mathbf{x}|\,|\nabla\mathbf{x}'|}. \quad (2)$$

III. Experimental Results

A. Data

In this experiment, we used phantoms containing the skull and teeth of deceased patients, cast in a uniform plastic resin that approximates the X-ray attenuation of human soft tissue. Realistic motions were simulated using the strategy described in our previous paper [21]. In [18], we explained two difficult scenarios where the previous algorithm failed. We assessed the current algorithm's performance in these difficult scenarios. Two non-returning motions were considered as illustrated in Figure 3. These scenarios are challenging because non-returning motions create double contours in the reconstruction (Figure 4), affecting the majority of projections and complicating diagnosis. To evaluate the quality of the final reconstruction, we performed a motion-free short-scan reconstruction of 194° to serve as a ground truth for comparison.



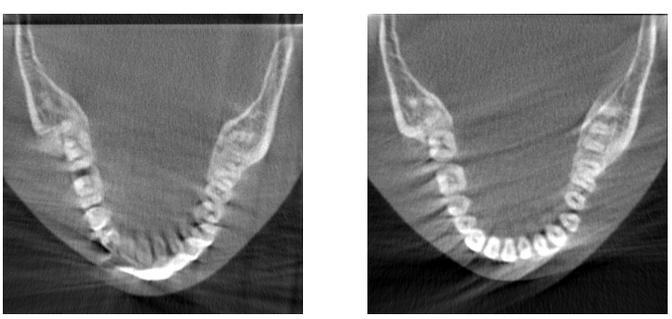
(a) Reconstruction (60°) (b) After Regularization (60°) (c) Forward Projection (60°)

Fig. 2: Regularization (The threshold is set to 1400 HU)



(a) View - 1 (b) View - 2

Fig. 3: Illustration showing the motion-affected and motion-free subsets for nodding motion (non-returning). 'red' indicates patient movement, while 'gray' and 'yellow' signify instances where the patient remains stationary but in distinct poses.



(a) View - 1 (b) View - 2

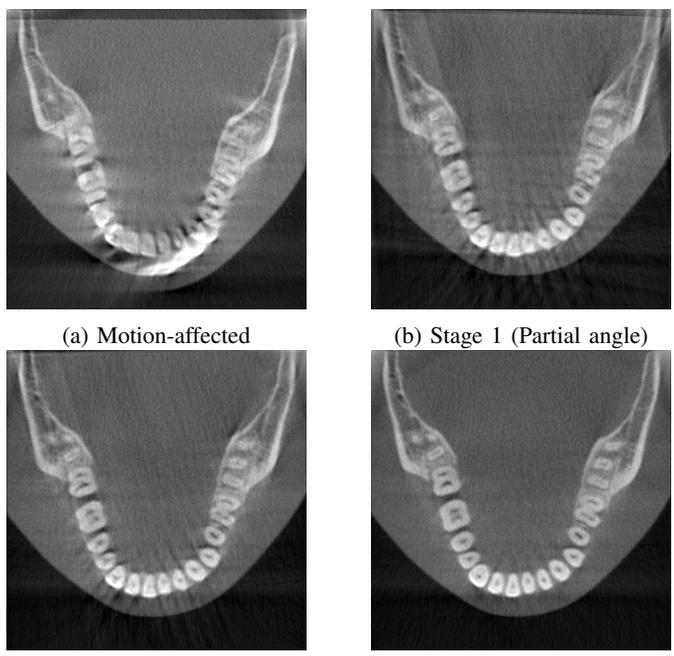
Fig. 4: Results of the difficult cases using our previous approach explained in [18]

B. Results

The performance of the partial angle method is detailed in Table I, where its effectiveness is assessed using SSIM and RMSE metrics. Notably, there is a substantial enhancement observed in both evaluation metrics. Furthermore, the motion-compensated reconstructions showcased in Figures 5 and 6 exhibit a marked improvement, successfully mitigating nearly all artifacts. This suggests that the proposed approach holds promise for applications dealing with significant patient movements, showcasing its potential to be a valuable solution in such scenarios.

TABLE I: SSIM and RMSE values for nodding motions (non-returning) using Partial Angle Method [MA = motion-affected, MC = motion-compensated]

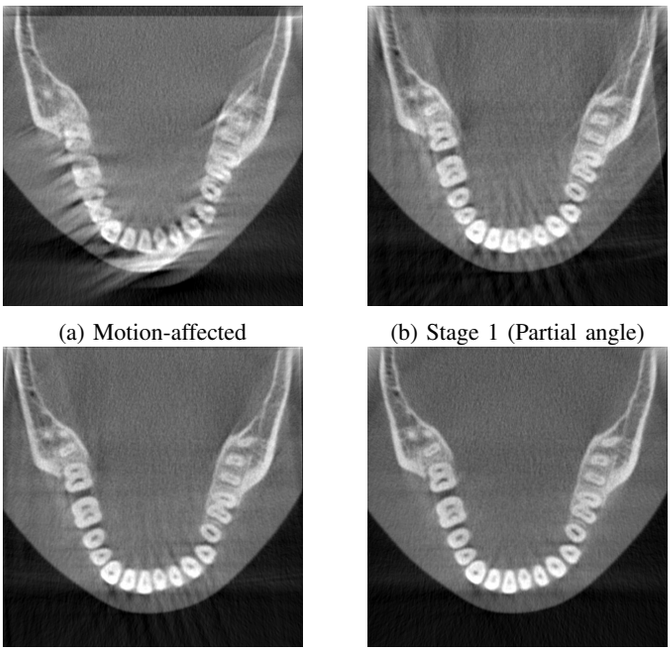
	SSIM (↑)			RMSE ($\times 10^{-2}$)(↓)		
View	MA	MC	Diff %	MA	MC	Diff %
View - 1	0.628	0.774	+14.52	7.53	3.22	-4.31
View - 2	0.726	0.814	+8.78	5.95	2.51	-3.44



(a) Motion-affected (b) Stage 1 (Partial angle)

(c) Stage 2 (d) Ground truth

Fig. 5: Results for non-returning nodding motion (View-1)



(a) Motion-affected (b) Stage 1 (Partial angle)

(c) Stage 2 (d) Ground truth

Fig. 6: Results for non-returning nodding motion (View-2)

IV. Conclusion

The proposed partial angle motion compensation method emerges as a promising solution to address challenges in maxillofacial cone-beam computed tomography (CBCT) imaging, particularly in prolonged patient movements. The method extends the applicability of our previous approach to the most challenging scenarios. Utilizing an incremental compensation strategy that starts with an ultra-short-scan reference volume, the proposed method provides a robust reference for the second stage of the algorithm. After the second stage, the quality of the final motion-compensated reconstruction is effectively improved. This suggests the potential applicability of the proposed approach in clinical scenarios.

Currently, a limitation is the need for a small subset of nearly motion-free projections. Future research is necessary to fully understand the limits of the algorithm if the constraint of the nearly motion-free subset is relaxed. Since the partial angle motion compensation only applies to the first stage of the algorithm, our focus is to generate a precompensated reconstruction which is a reliable reference for the second stage.

Acknowledgment

The authors thank Michele Antonelli, Ivan Tomba, Andrea Delmiglio, Lorenzo Arici, Luca Fracassetti from See Through s.r.l., and Prof. Giovanni Di Domenico from the University of Ferrara, Italy for providing the projection data and the reconstruction libraries.

References

[1] S. Patel, C. Durack, F. Abella, H. Shemesh, M. Roig, and K. Lemberg, "Cone beam computed tomography in e ndodontics–a review," *International endodontic journal*, vol. 48, no. 1, pp. 3–15, 2015.

[2] B. W. Small, "Cone beam computed tomography." *General dentistry*, vol. 55, no. 3, pp. 179–181, 2007.

[3] A. Nemtoi, C. Czink, D. Haba, and A. Gahleitner, "Cone beam ct: a current overview of devices," *Dentomaxillofacial Radiology*, vol. 42, no. 8, p. 20120443, 2013.

[4] E. Yıldızer Keriş, "Effect of patient anxiety on image motion artefacts in cbct," *BMC oral health*, vol. 17, no. 1, pp. 1–9, 2017.

[5] K. Donaldson, S. O'Connor, and N. Heath, "Dental cone beam ct image quality possibly reduced by patient movement," *Dentomaxillofacial Radiology*, vol. 42, no. 2, p. 91866873, 2013.

[6] R. Spin-Neto, L. H. Matzen, L. Schropp, E. Gotfredsen, and A. Wenzel, "Factors affecting patient movement and re-exposure in cone beam computed tomography examination," *Oral surgery, oral medicine, oral pathology and oral radiology*, vol. 119, no. 5, pp. 572–578, 2015.

[7] T. Hanzelka, J. Dusek, F. Ocasek, J. Kucera, J. Sedy, J. Benes, G. Pavlikova, and R. Foltan, "Movement of the patient and the cone beam computed tomography scanner: objectives and possible solutions," *Oral surgery, oral medicine, oral pathology and oral radiology*, vol. 116, no. 6, pp. 769–773, 2013.

[8] R. Spin-Neto, L. H. Matzen, L. W. Schropp, T. S. Sørensen, and A. Wenzel, "An ex vivo study of automated motion artefact correction and the impact on cone beam ct image quality and interpretability," *Dentomaxillofacial Radiology*, vol. 47, no. 5, p. 20180013, 2018.

[9] W. Wein, A. Ladikos, and A. Baumgartner, "Self-calibration of geometric and radiometric parameters for cone-beam computed tomography," in *11th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 2, 2011, pp. 1–4.

[10] S. Ouadah, M. Jacobson, J. W. Stayman, T. Ehtiati, C. Weiss, and J. H. Siewerdsen, "Correction of patient motion in cone-beam ct using 3d–2d registration," *Physics in Medicine & Biology*, vol. 62, no. 23, p. 8813, 2017.

[11] T. Sun, R. Jacobs, R. Pauwels, E. Tijskens, R. Fulton, and J. Nuyts, "A motion correction approach for oral and maxillofacial cone-beam ct imaging," *Physics in Medicine & Biology*, vol. 66, no. 12, p. 125008, 2021.

[12] J. Wicklein, Y. Kyriakou, W. A. Kalender, and H. Kunze, "An online motion-and misalignment-correction method for medical flat-detector ct," in *Medical Imaging 2013: Physics of Medical Imaging*, vol. 8668. SPIE, 2013, pp. 466–472.

[13] A. Sisniega, J. W. Stayman, J. Yorkston, J. Siewerdsen, and W. Zbijewski, "Motion compensation in extremity cone-beam ct using a penalized image sharpness criterion," *Physics in Medicine & Biology*, vol. 62, no. 9, p. 3712, 2017.

[14] S. Maur, D. Stsepankou, and J. Hesser, "Cbct auto-calibration by contour registration," in *Medical Imaging 2019: Physics of Medical Imaging*, vol. 10948. SPIE, 2019, pp. 413–421.

[15] L. Birklein, S. Niebler, E. Schömer, R. Brylka, U. Schwanecke, and R. Schulze, "Motion correction for separate mandibular and cranial movements in cone beam ct reconstructions," *Medical physics*, vol. 50, no. 6, pp. 3511–3525, 2023.

[16] T. Würfl, M. Hoffmann, A. Aichert, A. K. Maier, N. Maaß, and F. Dennerlein, "Calibration-free beam hardening reduction in x-ray cbct using the epipolar consistency condition and physical constraints," *Medical physics*, vol. 46, no. 12, pp. e810–e822, 2019.

[17] A. S. R. Asraf Ali, C. Sarti, and C. Landi, "Compensation for patient movements in cbct imaging for dental applications," in *Image Analysis and Processing – ICIAP 2023*, G. L. Foresti, A. Fusiello, and E. Hancock, Eds. Cham: Springer Nature Switzerland, 2023, pp. 540–551.

[18] A. S. R. Asraf Ali, C. Sarti, C. Landi, and A. Fusiello, "Motion compensation in short-scan cbct reconstructions for dental applications," to appear at SPIE 2024, 18-22 February 2024, San Diego (USA).

[19] 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.

[20] T. De Silva, A. Uneri, M. Ketcha, S. Reaungamornrat, G. Kleinszig, S. Vogt, N. Aygun, S. Lo, J. Wolinsky, and J. Siewerdsen, "3d–2d image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch," *Physics in Medicine & Biology*, vol. 61, no. 8, p. 3009, 2016.

[21] A. S. R. Asraf Ali, A. Fusiello, C. Landi, C. Sarti, and A. A. P. Siswadi, "Motion artifacts detection in short-scan dental cbct reconstructions," *arXiv preprint arXiv:2304.10154*, 2023.

On the Influence of Smoothness Constraints in Computed Tomography Motion Compensation

Mareike Thies, Fabian Wagner, Noah Maul, Siyuan Mei, Mingxuan Gu, Laura Pfaff, Nastassia Vysotskaya, Haijun Yu, and Andreas Maier

***Abstract*—Computed tomography (CT) relies on precise patient immobilization during image acquisition. Nevertheless, motion artifacts in the reconstructed images can persist. Motion compensation methods aim to correct such artifacts post-acquisition, often incorporating temporal smoothness constraints on the estimated motion patterns. This study analyzes the influence of a spline-based motion model within an existing rigid motion compensation algorithm for cone-beam CT on the recoverable motion frequencies. Results demonstrate that the choice of motion model crucially influences recoverable frequencies. The optimization-based motion compensation algorithm is able to accurately fit the spline nodes for frequencies almost up to the node-dependent theoretical limit according to the Nyquist-Shannon theorem. Notably, a higher node count does not compromise reconstruction performance for slow motion patterns, but can extend the range of recoverable high frequencies for the investigated algorithm. Eventually, the optimal motion model is dependent on the imaged anatomy, clinical use case, and scanning protocol and should be tailored carefully to the expected motion frequency spectrum to ensure accurate motion compensation.**

***Index Terms*—cone-beam computed tomography, patient motion, motion frequency, splines.**

I. Introduction

FOR computed tomography (CT), it is crucial that the patient does not move throughout the acquisition of the scan. Patient motion leads to a mismatch between the measured sinogram and the geometry settings assumed during reconstruction which results in artifacts in the reconstructed image. Motion compensation approaches consequently aim to eliminate these artifacts in a post-processing step.

Whereas some methods act only on the reconstructed, motion-affected image [1], many works explicitly estimate the underlying motion curves and perform a motion-compensated reconstruction given an updated geometry [2], [3], [4], [5]. Explicit modelling of the motion patterns ensures that the motion-compensated reconstruction is consistent with the measured sinogram and yields interpretable solutions where estimated motion curves can be inspected for plausibility.

Most approaches, which estimate motion curves explicitly, enforce a type of smoothness constraint on the motion patterns over time [3], [4], [5]. Usually, this is achieved by either adding a regularization term to the objective function which implicitly punishes non-smooth motion patterns, or explicitly approximating the motion patterns with a smooth function. Using such smoothness constraints makes sense intuitively, since realistic motion patterns are temporally correlated especially considering the high frame rate at which projection images are acquired in modern scanners. Moreover, it can simplify the optimization problem of motion estimation by smoothing the loss landscape or by reducing the number of free parameters which need to be estimated. However, the influence of these modelling decisions on the type of motion patterns which can be recovered by the particular method is often not investigated.

All authors are with the Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, Correspondence e-mail: mareike.thies@fau.de.

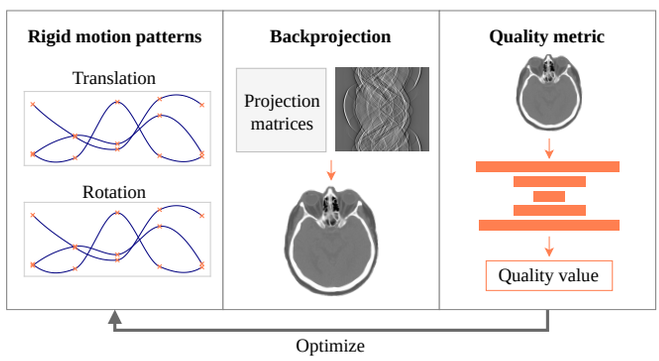


Fig. 1. High level overview of the motion compensation algorithm studied in this paper.

In this work, we perform a frequency analysis for our previously published motion compensation algorithm [6] which features an explicit spline-based representation of the motion patterns. Specifically, we formulate the following research questions:

- Which frequencies in the patient's motion patterns can be reliably recovered by the algorithm?
- How do these frequencies relate to the choice of motion model parameterizing the recovering motion patterns?
- Can we loosen the smoothness constraint to increase the range of recovering motion patterns that can be fitted?

II. Methods

A. Motion Compensation Algorithm

This work builds up on our previously developed rigid motion compensation algorithm for cone-beam CT systems [6]. It estimates motion patterns for the six rigid motion parameters (t_x, t_y, t_z, r_x, r_y, r_z) over the scan range. Here, $t_\square$ represents translation along one axis and $r_\square$ represents rotation around one axis (Euler angles). The estimation of the motion patterns is performed by formulating an optimization problem. The target function is inspired by the autofocus principle where intermediate reconstructions are computed given the

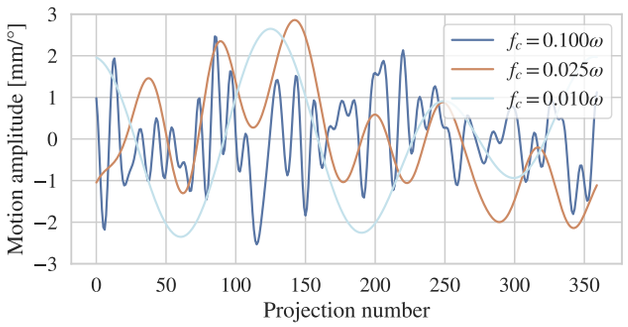


Fig. 2. Exemplary motion patterns obtained after limiting a random signal to the cutoff frequency f_c. Motion patterns with a low cutoff frequency oscillate slower than those with high cutoff frequencies.

current estimate of the motion parameters. A quality metric quantifies the amount of left-over artifacts in the intermediate reconstruction and guides the optimization toward motion patterns yielding reconstructions with maximal quality. For the intermediate reconstruction, we use a backprojector for which an analytic Jacobian is computed. The Jacobian connects the space of intermediate reconstructions and the space of cone-beam geometries (parameterized by projection matrices) in a differentiable manner [7]. Based on this, gradients of the quality metric to the free parameters of the recovering motion patterns are obtained, and gradient-based optimization is applied. The quality metric is a reference-free mapping which grades the quality of the intermediate reconstruction given only the motion-affected volume. It is parameterized by a neural network trained to regress the visual information fidelity (VIF) [8] as a spatially resolved volumetric map from the input reconstruction followed by a simple average resulting in a scalar quality value. A high-level graphical summary of the most important steps of the motion compensation algorithm is presented in Fig. 1 and we refer the interested reader to our paper [6] for a more in-depth description.

B. Spline-Based Motion Model

The motion compensation algorithm in [6] enforces smoothness of the separate rigid motion parameters. This is achieved by approximating the parameter's values over time with smooth Akima splines [9]. Each spline is parameterized by N_n nodes which are distributed over the full scan range with equidistant spacing and with one node aligning with the first and last projection of the scan, respectively. Given that a scan has N_p projections, the spline-based formulation limits the number of free parameters which need to be fitted during motion estimation from $6N_p$ for the unconstrained setting to $6N_n$ for the setting with smoothness constraint ($N_n < N_p$).

C. Frequency Analysis

To analyze the frequency behavior, data is perturbed with random motion signals of different band-limitation in frequency domain. For each of the six rigid parameters, an unconstrained[1] random motion pattern x is sampled. Subsequently, a Fourier transformation $\mathcal{F}$ is applied to the randomly sampled pattern and all frequencies higher than a certain cutoff frequency f_c are set to zero. After an inverse Fourier transform $\mathcal{F}^{-1}$, the signal is rescaled to a predefined maximal amplitude. The resulting motion pattern x_c is

$$x_c = r(\mathcal{F}^{-1}(b(\mathcal{F}(x), f_c))) \ , \tag{1}$$

where $b(\cdot,\cdot)$ performs the band-limitation of the signal in Fourier domain with cutoff frequency f_c and r rescales the amplitude of the signal. Three examples of band-limited motion patterns along with their corresponding cutoff frequency are shown in Fig. 2. All investigations in this paper are performed independently of the scanner-specific sampling rate ω, measured in $1/\mathrm{s}$, at which projections are acquired. For an entirely random motion pattern without smoothness constraint, the maximal frequency that can be measured given a certain scanner-dependent sampling rate ω is $f_{max} = 0.5\omega$, which follows from the fundamental Nyquist-Shannon sampling theorem. Any cutoff frequency $f_c < f_{max}$ smaller than this maximal frequency reduces the highest frequency present in the motion signal. Cutoff frequencies are sampled such that they are equidistant in logarithmic space. $f_c = 0.005\omega$ is the smallest investigated frequency and $f_c = f_{max}$ is the highest investigated frequency which corresponds to unconstrained motion.

III. Data and Experiments

The experiments are based on the publicly available CQ500 data set [10] which contains reconstructed head CT volumes acquired on different helical multi-detector CT scanners. For the investigations in this paper, we use a subset of ten scans which are reconstructed with a small slice thickness of $0.625\,\mathrm{mm}$. These scans have not been incorporated in the training or development of the quality metric regression network which is part of the motion compensation pipeline. Cone-beam sinograms are obtained from the reconstructed volumes via forward projection using the software *PyroNN* [11]. We simulate a (potentially motion-perturbed) circular full scan with 360 projections, a source-to-isocenter distance of $785\,\mathrm{mm}$, a source-to-detector distance of $1200\,\mathrm{mm}$, and a detector of shape 500×700 pixels of isotropic pixel spacing ($0.64\,\mathrm{mm}$). The motion compensation algorithm is run on reconstructed volumes of 128^3 voxels with an isotropic spacing of $2\,\mathrm{mm}$. For evaluation, we reconstruct the data on grids with 256^3 voxels and a spacing of $1\,\mathrm{mm}$ in all dimensions.

Motion compensation is performed by executing the following steps: First, a random, band-limited motion signal is sampled as described in section II-C using a maximal amplitude of $5\,\mathrm{mm}$ for translation and $5°$ for rotation. This signal is used to rigidly perturb the ideal circular trajectory and a volume from the data set is forward projected onto this motion-affected trajectory. The motion compensation algorithm is initialized with the ideal circular trajectory. It estimates the perturbing motion pattern as precisely as possible based on the spline-based representation explained in section II-B.

[1]In this study, we always refer to unconstrained *observable* patterns, i.e., sampled with the scanner-specific sampling rate.

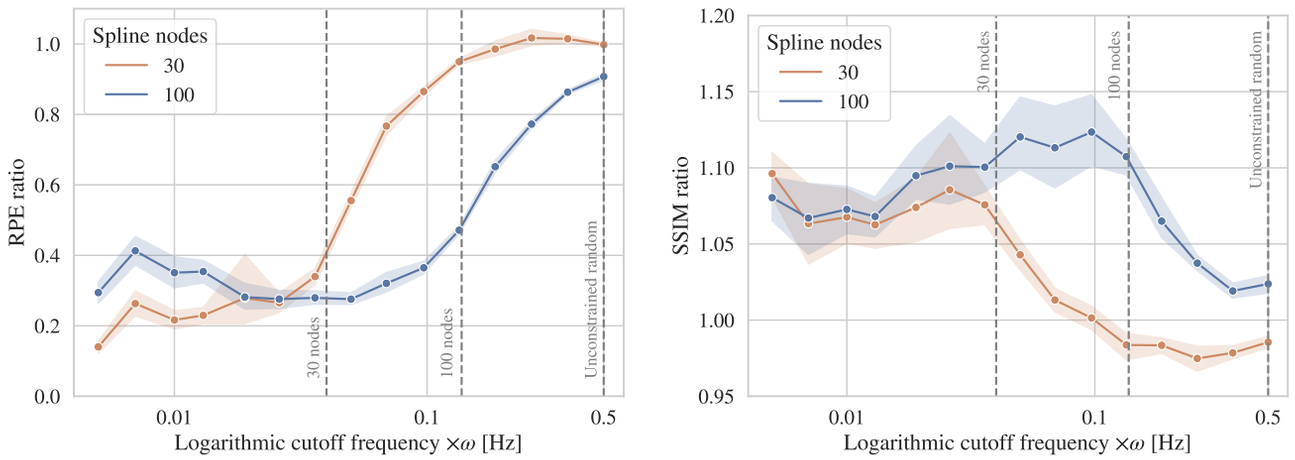


Fig. 3. Plot of the change in reprojection error (RPE) (left) and structural similarity index measure (SSIM) (right). Both are displayed as their ratio of after/before motion compensation. While smaller values for RPE and higher values for SSIM are better, respectively, a ratio close to 1 reflects no improvement after motion compensation for either metric. The ratio is drawn as a function of the cutoff frequency f_c in the perturbing motion pattern. Note that the x-axis is scaled logarithmically from small to high frequencies. The maximum measurable frequencies according to the Nyquist-Shannon theorem are indicated by gray dashed lines for both spline types and the unconstrained setting. The two curves show the results obtained with splines with 30 nodes and 100 nodes for the estimated motion signal. 95 % confidence intervals are computed over ten different scans.

IV. Results

The motion compensation algorithm is applied to ten scans. 15 different perturbing motion patterns with varying cutoff frequencies f_c are applied to each sample. Each combination is recovered using 30 and 100 nodes for the spline modelling the recovering signal leading to $10 \cdot 15 \cdot 2 = 300$ runs in total. We compute the reprojection error (RPE) which measures deviations between projected points on the detector planes and the structural similarity index measure (SSIM) on the reconstructed volumes both before and after motion compensation. The ratio of RPE and SSIM after and before compensation depending on the cutoff frequency is plotted in Fig. 3. The RPE ratio (left) is clearly below 1.0 for lower cutoff frequencies indicating a strong improvement of the RPE over the initial state. For the lowest investigated frequencies, the spline with 30 nodes yields slightly better results than the spline with 100 nodes. Both spline types perform on par for intermediate low frequencies around $f_c = 0.02\omega$. The results obtained with 30 nodes become worse (increase in RPE ratio) when approaching the maximum node-dependent measurable frequency. The spline with 100 nodes shows the same behavior, but the increase in RPE happens at higher cutoff frequencies compared to the spline with fewer nodes. For completely unconstrained motion patterns, both splines lead to no, or only very minor improvement in RPE.

The SSIM ratio plot (right) confirms these findings. In this case, higher values indicate an improved image quality after motion compensation. Again, we observe similar performance of both spline types for low frequencies up to the maximal frequency that can be recovered by 30 nodes. The SSIM ratio drops for the spline with 30 nodes, but stays high for the spline with 100 nodes until the respective critical frequency for 100 nodes is reached. Both splines do not yield a substantial improvement for unconstrained motion patterns.

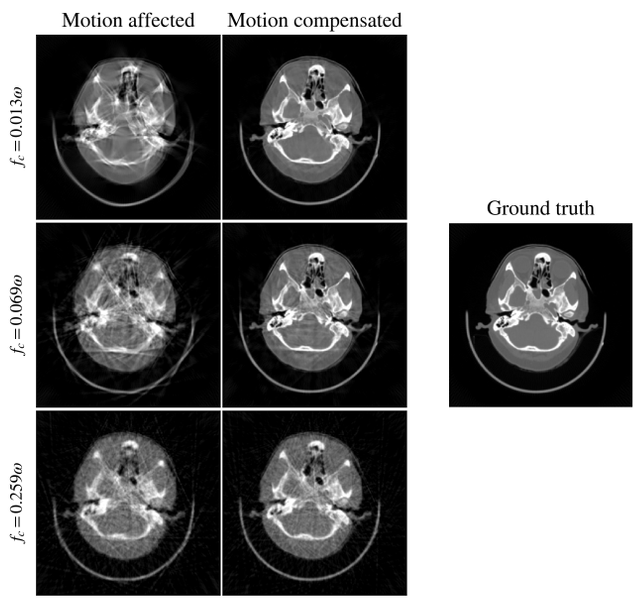


Fig. 4. Reconstruction results for an axial slice of one example scan from three different cutoff frequencies for the perturbing motion pattern and a spline with 100 nodes for the estimated signal. The ground truth is identical for all three cases. Motion-compensated results appear best for the lowest frequency and worst for the highest frequency.

Fig. 4 shows reconstructed results obtained from the same scan but with three different cutoff frequencies for the perturbing motion pattern. Looking at the motion-affected slices, we see that the frequency of the motion pattern has a strong influence on the appearance of the motion artifacts. Low frequencies lead to distortions of the shapes and misaligned edges, but some areas are still mostly homogeneous. High

frequencies, in contrast, result in overall blurring of the reconstructed image and pronounced streak artifacts which affect all regions of the image. The motion-compensated results obtained with 100 nodes recover these artifacts well for the two lower frequencies and yield reconstructed images which resemble the ground truth closely. However, for $f_c = 0.069\omega$ some inhomogeneities remain especially for soft tissue areas. The highest frequency in the figure $f_c = 0.259\omega$ is above the critical frequency for 100 nodes according to the Nyquist-Shannon theorem. Therefore, the motion-compensated result sharpens the bone structures to some extent, but can not recover the blur and streaking artifacts caused by the high-frequency motion components.

V. Discussion

The presented results confirm that the choice of motion model has a profound influence on the types of motion patterns which can be recovered by a certain motion compensation algorithm. For the investigated method, the frequencies that can robustly be recovered are limited by the sampling rate according to Nyquist-Shannon given the number of nodes of the respective spline. In the plot showing the RPE ratio, it can be observed that the performance of the motion compensation begins to decrease for frequencies slightly lower than the critical recoverable frequency. This confirms that the optimization in the investigated motion compensation algorithm is able to fit the spline-based motion model almost to its theoretical limit in terms of recoverable frequencies. For low-frequency motion patterns, the spline with fewer nodes leads to larger improvements in RPE, but these do not seem to translate into the reconstructed image as the SSIM ratio is similar for 30 and 100 nodes even for the lowest frequencies. On the other hand, a higher number of nodes increases the frequency range that can be recovered. As a result, for the investigated method, it seems reasonable to prefer a larger number of nodes if the maximal expected frequency in the motion signal can not reliably be estimated.

We did not test the method with splines of more than 100 nodes. In the setting of this study with 360 projections in total, such a spline has a node for every $360/99 = 3.64$ acquired projection images. Considering existing research about the characteristics of head motion during a scanning procedure [12], we believe that this setting is sufficient to cover even the high-frequency motion patterns that could occur in practice.

The network regressing the quality metric from an intermediate, motion-affected reconstruction in the investigated motion compensation algorithm was trained on slow motion patterns parameterized by splines with ten nodes. Interestingly, this does not seem to have a detrimental influence on the target function for higher-frequency motion patterns. We conclude that the quality regression network generalizes well to other frequencies than those present in the training data.

This study is limited to rigid head motion and assumes that the motion occurs inter-frame, i.e., in between the projections and not during the acquisition of a single projection. Furthermore, we only test an explicit, spline-based motion model. A comparison to an implicit smoothness constraint that regularizes the target function is a valuable future extension of this work.

VI. Conclusion

In this study, we assess the influence of the smooth motion parameterization on the motion-compensated reconstruction depending on the frequency spectrum of the motion signal. Our findings indicate that, for the investigated motion compensation algorithm, an increased spline node count accommodates higher motion frequencies while maintaining robustness for low frequencies. This observation validates the algorithm's efficacy under a relaxed smoothness constraint. Ultimately, the choice of motion model has a substantial impact on the recoverable motion patterns and should be chosen with care in the context of the clinical scan protocol and application.

Acknowledgment

The research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC Grant No. 810316). The authors gratefully acknowledge the scientific support and HPC resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). The hardware is funded by the German Research Foundation (DFG).

References

[1] Y. Ko, S. Moon, J. Baek, and H. Shim, "Rigid and non-rigid motion artifact reduction in X-ray CT using attention module," *Med. Image Anal.*, vol. 67, p. 101883, 2021.

[2] A. Aichert, M. Berger, J. Wang, N. Maass, A. Doerfler, J. Hornegger, and A. Maier, "Epipolar Consistency in Transmission Imaging," *IEEE Trans. Med. Imaging*, vol. 34, no. 11, pp. 2205–2219, 2015.

[3] M. Berger, K. Müller, A. Aichert, M. Unberath, J. Thies, J.-H. Choi, R. Fahrig, and A. Maier, "Marker-free motion correction in weight-bearing cone-beam CT of the knee joint," *Med. Phys.*, vol. 43, no. 3, pp. 1235–1248, 2016.

[4] A. Sisniega, J. W. Stayman, J. Yorkston, J. Siewerdsen, and W. Zbijewski, "Motion compensation in extremity cone-beam CT using a penalized image sharpness criterion," *Phys. Med. Biol.*, vol. 62, no. 9, p. 3712, 2017.

[5] A. Preuhs, M. Manhart, P. Roser, E. Hoppe, Y. Huang, M. Psychogios, M. Kowarschik, and A. Maier, "Appearance Learning for Image-Based Motion Estimation in Tomography," *IEEE Trans. Med. Imaging*, vol. 39, no. 11, pp. 3667–3678, 2020.

[6] M. Thies, F. Wagner, N. Maul, H. Yu, M. Meier, L.-S. Schneider, M. Gu, S. Mei, L. Folle, and A. Maier, "A gradient-based approach to fast and accurate head motion compensation in cone-beam CT," *arXiv*, 2024.

[7] M. Thies, F. Wagner, N. Maul, L. Folle, M. Meier, M. Rohleder, L.-S. Schneider, L. Pfaff, M. Gu, J. Utz *et al.*, "Gradient-based geometry learning for fan-beam CT reconstruction," *Phys. Med. Biol.*, vol. 68, no. 20, p. 205004, 2023.

[8] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, pp. 430–444, 2 2006.

[9] H. Akima, "A new method of interpolation and smooth curve fitting based on local procedures," *Journal of the ACM*, vol. 17, no. 4, pp. 589–602, 1970.

[10] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N. G. Campeau, V. K. Venugopal, V. Mahajan, P. Rao, and P. Warier, "Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study," *The Lancet*, vol. 392, no. 10162, pp. 2388–2396, 2018.

[11] C. Syben, M. Michen, B. Stimpel, S. Seitz, S. Ploner, and A. Maier, "PYRO-NN: Python reconstruction operators in neural networks," *Med. Phys.*, vol. 46, no. 11, pp. 5110–5115, 2019.

[12] A. Wagner, K. Schicho, F. Kainberger, W. Birkfellner, S. Grampp, and R. Ewers, "Quantification and Clinical Relevance of Head Motion during Computed Tomography," *Invest. Radiol.*, vol. 38, pp. 733–741, 2003.

Motion-compensated time-enhancement curve estimation for intra-operative brain perfusion assessment using a standard x-ray system with no gantry rotation

Katsuyuki Taguchi,[1] Shalini Subramanian,[1] and W. Paul Segars[2]

I. Introduction

Many interventional procedures aim at changing soft tissue perfusion or blood flow. For example, endovascular mechanical thrombectomy (EVT) for acute ischemic stroke is to retrieve an ischemic-causing thrombus to recover blood flow to the downstream brain tissues. One problem at present is that soft tissue perfusion and its changes cannot be assessed volumetrically in an interventional suite. An x-ray angiogram (two-dimensional projections) provides a good temporal resolution, but it lacks volumetric information, and therefore, it does not work when lesions are overlapped or if sub-sections of a lesion have different time-enhancement curves (TECs). Cone-beam computed tomography (CT) takes 4–10 s per volume scan and is too slow for perfusion assessment. Interventional radiologists cannot assess if the procedure they just performed on a patient was successful and they are ready to close it, or if it was not successful and an additional procedure is necessary. These are unmet clinical needs we plan to address ultimately.

We have developed an innovative method called IPEN for **I**ntra-operative 4-dimensional soft tissue **PE**rfusion using a standard x-ray system with **N**o gantry rotation [1]. Although IPEN works without moving the gantry and uses standard x-ray angiography images, IPEN is built on the same principle as x-ray CT, making accurate soft tissue perfusion imaging possible. The key is to divide the object into multiple regions-of-interest (ROIs) and estimate the mean value of each ROI. When the object consists of N ROIs, N x-ray beams are sufficient to estimate N values (one per ROI) even if the ROIs are overlapped (Fig. 1). Thus, estimating ~100 ROI values from one x-ray projection that has 1,000×1,000 pixels is an extremely over-determined, well-posed problem. The IPEN is built on a solid tomography theory (presented in the next section), which enables a new imaging paradigm of imaging by dividing the object into N ROIs with known shapes and locations.

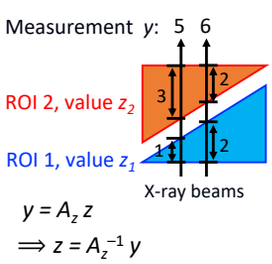


Fig. 1. IPEN principle: N x-ray beams are sufficient to estimate N ROI values. A_z is a projection matrix

We have developed the IPEN for brain perfusion and a computer simulation study showed that the approach worked well and was quantitatively accurate. It will be presented at this conference by Shalini Subramanian [2].

One of the remaining issues is the effect of patient motion. When patient motion is present, both x-ray CT and x-ray digital subtraction angiography (DSA) images motion artifacts (Fig. 2), which could degrade the accuracy of perfusion assessment. A 2-D angiography analysis method [3] that assesses time-pixel value curves obtained by DSA images for each pixel would also fail because brain tissues where x-ray beams pass through are inconsistent over time.

Head motion artifacts occur relatively frequently even if the patient head is strapped to the couch to restrict the head motion. Respiratory and cardiac motion artifacts are also present for in abdominal and cardiac angiography images as well.

The purpose of this study is to propose a novel method to tackle patient motion within IPEN's framework. It is desirable to further develop the IPEN algorithm such that it can handle patient motion. We derived the algorithm and performed computer simulations to show that the core idea of the algorithm works.

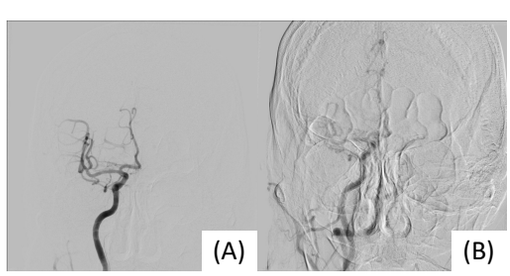


Fig. 2. Digital subtraction angiography (DSA) images without (A) and with (B) patient motion

II. Algorithms and Evaluation

We describe the algorithms and outline a computer simulation study in this section. In this study, we assumed that the object orientation relative to the x-ray angiography system was estimated for each x-ray projection using a 2D–3D registration algorithm, similar to Step 2 of the 4-step algorithm for brain

1. The Johns Hopkins University. 2. Duke University

perfusion [2] (Fig. 3). We also assumed that un-subtracted native angiography images were available.

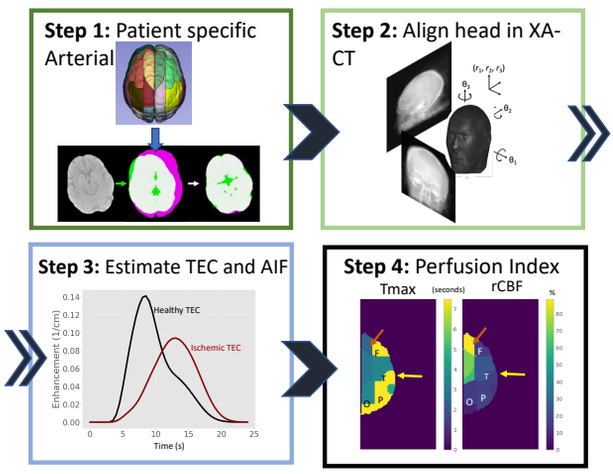


Fig. 3. IPEN workflow. Motion-compensated time-enhancement curve (MC–TEC) estimation from native x-ray angiography (XA) images in Step 3 is the topic of this work.

A. *Original IPEN algorithm*

Below we outline the original IPEN algorithm that does *not* take into account patient motion. We use bold letters to denote vectors and matrices.

Suppose that an object (e.g., a head) consists of small voxels. Let $\boldsymbol{x}_t$, $0 \le t \le t_{end}$, be a vector of voxel values (i.e., linear attenuation coefficients) at time t, and $\underline{\boldsymbol{x}}_{0\to t}$ be a vector for the enhancement, i.e., changes of voxel values from time 0 to t. Let also $\boldsymbol{z}_t$ and $\underline{\boldsymbol{z}}_{0\to t}$ be vectors for the ROI values and the enhancement values, respectively. We assume that the enhancement is homogeneous for voxels inside the same ROI. Then $\boldsymbol{x}_t$ can be expressed as a summation of pre-contrast-enhanced value and the enhancement value as

$$\boldsymbol{x}_t = \boldsymbol{x}_0 + \underline{\boldsymbol{x}}_{0\to t} = \boldsymbol{x}_0 + \boldsymbol{A}_{z2x}\underline{\boldsymbol{z}}_{0\to t}\,, \quad (1)$$

where a projection matrix $\boldsymbol{A}_{X2Y}$ maps a vector $\boldsymbol{X}$ to a vector $\boldsymbol{Y}$. Cone-beam projection of the object at time t, $\boldsymbol{y}_t$, can be expressed by

$$\boldsymbol{y}_t = \boldsymbol{A}_{x2y}\boldsymbol{x}_t = \boldsymbol{A}_{x2y}\boldsymbol{x}_0 + \boldsymbol{A}_{x2y}\boldsymbol{A}_{z2x}\underline{\boldsymbol{z}}_{0\to t} = \boldsymbol{y}_0 + \boldsymbol{A}_{z2y}\underline{\boldsymbol{z}}_{0\to t}\,, \quad (2)$$

where $\boldsymbol{y}_0$ is native angiography data at t=0 and Eq. (1) was used during the derivation of Eq. (2). Using measured angiography data $\tilde{\boldsymbol{y}}_{\boldsymbol{0}}$ and $\tilde{\boldsymbol{y}}_{\boldsymbol{t}}$, the ROI enhancement vector $\underline{\boldsymbol{z}}_{0\to t}$ can be estimated via minimization:

$$\hat{\underline{\boldsymbol{z}}}_{0\to t} = argmin_{\underline{\boldsymbol{z}}_{0\to t}} \left\| \boldsymbol{A}_{z2y}\underline{\boldsymbol{z}}_{0\to t} + \tilde{\boldsymbol{y}}_0 - \tilde{\boldsymbol{y}}_t \right\|_W^2 . \quad (3)$$

The unenhanced voxel values $\boldsymbol{x}_0$ can be heterogenous as in many clinical cases. As the measured unenhanced projection $\tilde{\boldsymbol{y}}_{\boldsymbol{0}}$ fully contained $\boldsymbol{x}_0$, and therefore, the effect of heterogenous $\boldsymbol{x}_0$ was eliminated from the estimation problem. Heterogenous ROI enhancement makes the theory not exact; however, the previous study showed that the IPEN could work well even when the homogeneity assumption was violated [1].

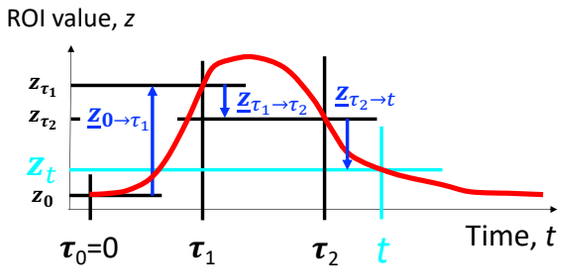


Fig. 4. The proposed MC–IPEN principle

B. *Proposed MC–IPEN algorithm*

Suppose that the head motion occurred multiple times within a short time ε right before $t = \tau_1, \tau_2, \tau_3, \ldots$, but the head stayed still otherwise (Fig. 4). We first divide the time into multiple segments between motion occurrences. Let T_i, $i = 0,1,2 \ldots$, denote the stationary time segment $[\tau_i, \tau_{i+1} - \varepsilon]$ and U_i, $i = 1,2 \ldots$, denote the motion segment $[\tau_i - \varepsilon, \tau_i]$. We then express the object at time $t \in T_i$ as

$$\begin{aligned}\boldsymbol{x}_{t\in T_i} = \boldsymbol{x}_0 + \underline{\boldsymbol{x}}_{0\to\tau_1-\varepsilon} + \underline{\boldsymbol{x}}_{\tau_1-\varepsilon\to\tau_1} + \cdots \\ + \underline{\boldsymbol{x}}_{\tau_{i-1}\to\tau_i-\varepsilon} + \underline{\boldsymbol{x}}_{\tau_i-\varepsilon\to\tau_i} + \underline{\boldsymbol{x}}_{\tau_i\to t}\end{aligned}\,. \quad (4)$$

Adding up the enhancement during motion-free segments and motion segments separately, we get

$$\boldsymbol{x}_{t\in T_i} = \boldsymbol{x}_0 + \sum_{j=1}^{i}\left(\underline{\boldsymbol{x}}_{\tau_{j-1}\to\tau_j-\varepsilon} + \underline{\boldsymbol{x}}_{\tau_j-\varepsilon\to\tau_j}\right) + \underline{\boldsymbol{x}}_{\tau_i\to t}. \quad (5)$$

The mapping $\boldsymbol{A}_{z2x}$ is an operation within the object; thus, it is independent of head motion and consistent over time. From Eq. (1) and Eq. (5), we get

$$\underline{\boldsymbol{z}}_{0\to t\in T_i} = \sum_{j=1}^{i}\left(\underline{\boldsymbol{z}}_{\tau_{j-1}\to\tau_j-\varepsilon} + \underline{\boldsymbol{z}}_{\tau_j-\varepsilon\to\tau_j}\right) + \underline{\boldsymbol{z}}_{\tau_i\to t}. \quad (6)$$

In the following, we aim at estimating $\underline{\boldsymbol{z}}_{0\to t\in T_i}$.

We generalize Eq. (1) and the object during a stationary segment can be expressed using the voxel values at the beginning of the stationary segment, $\boldsymbol{x}_{\tau_i}$, and the changes as

$$\boldsymbol{x}_{t\in T_i} = \boldsymbol{x}_{\tau_i} + \underline{\boldsymbol{x}}_{\tau_i\to t} = \boldsymbol{x}_{\tau_i} + \boldsymbol{A}_{z2x}\underline{\boldsymbol{z}}_{\tau_i\to t}. \quad (7)$$

In the following, we use mapping operations specific to each stationary time segment and use subscript i to clarify which stationary time segment T_i they belong to: $\boldsymbol{A}_{x2y,T_i}$ for voxel-to-projection mapping and $\boldsymbol{A}_{z2y,T_i}$ for ROI-to-projection mapping. They are a function of time segment because the head motion changes the orientation of the head relative to the x-ray angiography system; but they are consistent within the stationary time segment.

The corresponding angiography data can be expressed by

$$\begin{aligned}\boldsymbol{y}_{t\in T_i} = \boldsymbol{A}_{x2y,T_i}\boldsymbol{x}_t = \boldsymbol{A}_{x2y,T_i}\left(\boldsymbol{x}_{\tau_i} + \boldsymbol{A}_{z2x}\underline{\boldsymbol{z}}_{\tau_i\to t}\right) \\ = \boldsymbol{A}_{x2y,T_i}\boldsymbol{x}_{\tau_i} + \boldsymbol{A}_{x2y,T_i}\boldsymbol{A}_{z2x}\underline{\boldsymbol{z}}_{\tau_i\to t} = \boldsymbol{y}_{\tau_i} + \boldsymbol{A}_{z2y,T_i}\underline{\boldsymbol{z}}_{\tau_i\to t}\,.\end{aligned} \quad (8)$$

The ROI enhancement vector from t=τ_i, $\underline{\boldsymbol{z}}_{\tau_i\to t\in T_i}$ can then be estimated by

$$\hat{\underline{\boldsymbol{z}}}_{\tau_i\to t\in T_i} = argmin_{\underline{\boldsymbol{z}}_{\tau_i\to t}} \left\| \boldsymbol{A}_{z2y,T_i}\underline{\boldsymbol{z}}_{\tau_i\to t} + \tilde{\boldsymbol{y}}_{\tau_i} - \tilde{\boldsymbol{y}}_t \right\|_W^2. \quad (9)$$

Estimating $\underline{\boldsymbol{z}}_{\tau_{j-1}\to\tau_j-\varepsilon}$ for $j = 1, \ldots, i$ by substituting $t = \tau_i - \varepsilon$ into Eq. (9), we get the first and the last term of Eq. (6).

We estimate the enhancement during the short motion segment, $\underline{\boldsymbol{z}}_{\tau_i-\varepsilon\to t\in U_i}$, by interpolating the time-enhancement curve during the proceeding and succeeding segments, T_{i-1} and T_i, respectively:

$$\underline{\hat{\boldsymbol{z}}}_{\tau_i-\varepsilon\to t\in U_i} = interp\left(\underline{\hat{\boldsymbol{z}}}_{t\in T_{i-1}}, \underline{\hat{\boldsymbol{z}}}_{t\in T_i}, t\right). \quad (10)$$

Using Eqs. (6), (9)–(10), and substituting $t = \tau_j - \varepsilon$ into Eq. (9) and $t = \tau_{i+1}$ into Eq. (10), we get

$$\underline{\hat{\boldsymbol{z}}}_{0\to t\in T_i} = \sum_{j=1}^{i}\left(\underline{\hat{\boldsymbol{z}}}_{\tau_{j-1}\to\tau_j-\varepsilon} + \underline{\hat{\boldsymbol{z}}}_{\tau_j-\varepsilon\to\tau_j}\right) + \underline{\hat{\boldsymbol{z}}}_{\tau_i\to t}. \quad (11)$$

For minimization, Eq. (9), one could add a temporal roughness penalty and use a non-negativity constraint, if necessary. The minimum was found using a separable paraboloidal surrogate (SPS) optimizer [4, 5]. The SPS approximates the Poisson likelihood by a quadratic function with an appropriate statistical weighting that is easier to minimize, and the SPS further uses surrogate functions to separate all of the unknowns for ROIs such that they can be updated simultaneously.

C. Computer simulation settings

We used the 4-dimensional brain perfusion extended NURBS-based cardiac-torso (XCAT) phantom [6] and simulated eight acute ischemic stroke cases with thrombus occlusion at one of the four major arteries with an intra-arterial contrast injection over 4–6 s. A single-plane angiography system with a flat-panel detector of 1024×1024 pixels/frame was used to acquire images for 20–25 s at 5 frames/s. The x-rays were monoenergetic at 70 keV at the mean number of initial photons of 10^5/pixel/frame, from which Poisson noise was added.

The head motion occurred twice, once for 1.5–2.1 s after the start of contrast injection and the other for 4.5–5.1: thus, $U_1 = (1.5, 2.1)$ and $U_2 = (4.5, 5.1)$). They are during the wash-in and wash-out phases of the time-enhancement curves, respectively. See Fig. 3. The head motion was nodding of ±10˚ and lateral motion of ±10 mm for the former motion; azimuthal head rotation of ±10˚, lateral swing of ±10˚ and vertical motion of ±10 mm for the latter.

We performed three algorithms: (1, 2) the proposed MC–IPEN and the original IPEN to estimate time-enhancement curves, and (3) angiography pixel value analysis to obtain time-pixel value curves. We then calculated two pseudo perfusion indices—the maximum slope of the rising part of the curve and full-width-at-half-maximum of the curve (FWHM)—which were associated to the true perfusion indices, cerebral blood flow (CBF) and mean transmit time (MTT), respectively.

The accuracy of the estimated time-enhancement curves and pseudo perfusion indices was assessed by root-mean-square deviation (RMSD) from the true values.

D. Computer simulation results

Unfortunately, we have not completed the computer simulation study yet. We would need a few more weeks. We plan to continue to work on it this spring and will present these results at the conference.

III. Discussion

Typical perfusion assessment is a 4-D (or 3-D plus time) image reconstruction problem. The voxel locations are stationary but the voxel values change over time due to wash-in and wash-out of the contrast agent. Four-D image reconstruction methods have been investigated [7, 8], which attempt to reconstruct a 4-D (or 3-D plus time) object. Reconstructing a 4-D object from a series of gantry rotating angiography images, however, is an ill-posed problem.

The original IPEN algorithm has formulated the 4-D reconstruction problem into a rigorous inverse problem framework such that exact tomography is possible when a few assumptions are satisfied [1]: The enhanced part of the object consists of multiple ROIs; voxels within each ROI has homogeneous enhancement; and the shapes and locations of ROIs are known. A study showed that even when these assumptions are violated, the accuracy of estimated TECs and perfusion indices is within an acceptable range (e.g., ±10% error).

Perfusion assessment with patient motion is a 5-D (or 3-D plus two time-axes) image reconstruction problem. The temporal speed for perfusion and patient motion is very different from each other—several seconds for perfusion and a fraction of second for motion—making it even harder to address the issue. To our knowledge, an exact 5-D image reconstruction from a series of 2-D projections has not been successful yet.

In this study, the IPEN framework allowed us to extend it and address the patient motion with the brain perfusion. The novel formulation is theoretically sound, and the proposed MC–IPEN is built based on the solid tomography principle. No additional assumptions were necessary. Although the simulation study is ongoing, we expect the results of MC–IPEN with head motion to be comparable to the original IPEN with no motion [2]. We will complete the simulation study and present the results at the conference.

We plan to start the assessment of IPEN algorithms using clinical data soon. Once fully developed and validated, we believe that the original IPEN when motion is absent and the MC–IPEN with motion present will provide quantitative brain perfusion assessment from standard x-ray angiography images.

IV. Conclusions

We proposed the MC–IPEN algorithm for brain perfusion assessment with patient motion. Ongoing computer simulation is expected to validate the algorithm soon.

V. Acknowledgments

This research was funded by R01 NS126256 and R21 EB029049.

References

[1] K. Taguchi *et al.*, "Three-dimensional regions-of-interest–based intra-operative four-dimensional soft tissue perfusion imaging using a standard x-ray system with no gantry rotation: A simulation study for a proof of concept," *Medical Physics,* vol. 47, no. 12, pp. 6087-6102, 2020, doi: https://doi.org/10.1002/mp.14514.

[2] S. Subramanian *et al.*, "CT brain tissue perfusion assessment using an x-ray angiography images," presented at the The 8th International Meeting on Image Formation in X-Ray Computed Tomography, Bamberg, Germany, August, 2024. [Online]. Available: https://ct-meeting.org/scientific-program/.

[3] T. Murray, T. Rodt, and M. J. Lee, "Two-Dimensional Perfusion Angiography of the Foot:Technical Considerations and Initial Analysis," *Journal of Endovascular Therapy,* vol. 23, no. 1, pp. 58-64, 2016, doi: 10.1177/1526602815621289.

[4] H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Physics in Medicine and Biology,* vol. 44, no. 11, pp. 2835-2851, 1999. [Online]. Available: http://stacks.iop.org/0031-9155/44/i=11/a=311.

[5] D. Hao *et al.*, "Multi-resolution statistical image reconstruction for mitigation of truncation effects: application to cone-beam CT of the head," *Physics in Medicine & Biology,* vol. 62, no. 2, pp. 539–559, 2017. [Online]. Available: http://stacks.iop.org/0031-9155/62/i=2/a=539.

[6] S. E. Divel, S. Christensen, W. P. Segars, M. G. Lansberg, and N. J. Pelc, "A dynamic simulation framework for CT perfusion in stroke assessment built from first principles," *Medical Physics,* vol. 48, no. (published online, April 20, 2021, doi.org/10.1002/mp.14887), pp. 3500-3510, 2021, doi: https://doi.org/10.1002/mp.14887.

[7] M. T. Manhart *et al.*, "Dynamic iterative reconstruction for interventional 4-D C-arm CT perfusion imaging," (in eng), *IEEE Trans Med Imaging,* vol. 32, no. 7, pp. 1336-48, Jul 2013, doi: 10.1109/tmi.2013.2257178.

[8] A. Fieselmann *et al.*, "Interventional 4-D C-arm CT perfusion imaging using interleaved scanning and partial reconstruction interpolation," (in eng), *IEEE Trans Med Imaging,* vol. 31, no. 4, pp. 892-906, Apr 2012, doi: 10.1109/tmi.2011.2181531.

Motion Compensation for 4D Digital Subtraction Angiography via Deep Autofocus and Implicit Neural Motion Models

H. Huang, A. Lu, F. Gonzalez, T. Ehtiati, J. H. Siewerdsen, A. Sisniega

Abstract— **4D-DSA offers time-resolved 3D information of contrast concentration in vascular structures and has seen increased use in the neuro-interventional suite. Most current approaches to 4D-DSA are implemented using interventional cone-beam CT systems and involve the acquisition of a non-contrast-enhanced reference mask and a contrast-enhanced volume during contrast administration. The final 4D-DSA dataset is obtained under assumptions of perfect stationarity of the patient during both CBCT scans. However, cone-beam CT often shows moderately long acquisition time, making it susceptible to motion artifacts that could significantly degrade the accuracy of 4D-DSA. Recent developments in motion compensation with image-based autofocus, integrating learning-based autofocus metrics and implicit neural representation of the motion trajectory, has demonstrated reliable and robust performance in solving complex motion in CBCT images. In this work, we propose a novel framework that incorporates motion compensation into the 4D-DSA pipeline. 4D-DSA is obtained via a multi-stage approach involving deep autofocus motion estimation acting on the non-contrast-enhanced mask, simultaneous motion estimation and registration of the contrast-enhanced volume, and estimation of the time-dependent volumetric contrast distribution.**

The proposed methodology was evaluated on simulation studies of anatomically realistic brain/head phantoms and contrast distribution, featuring rigid motion with amplitude ranging from 4 mm to 12 mm. Motion compensation showed consistent improvement of image quality, yielding average SSIM of 0.939 compared to 0.877 without motion compensation, and average error in the estimated motion trajectory of 0.822 mm. The final 4D-DSA volumes showed 10.15x increase in DICE score between the estimated vascular distribution and the ground truth, and 98% reduction in the error of estimated time-dependent contrast distribution. The results open a way for increasing usability and accuracy of 4D-DSA in presence of patient motion, with potential reduction of repeated scans and improvement in the quantification of blood flow in vasculature.

Index Terms—**Motion Compensation, 4D-DSA, Digital Subtraction Angiography, Radiology, Cone-Beam CT**

I. Introduction

Traditional digital subtraction angiography (DSA) is obtained via subtraction of an x-ray planar image, denoted mask, and a series of contrast-enhanced x-ray fluoroscopy views. The two-dimensional angiography views obtained in this manner have high temporal resolution, but present obvious limitations associated with anatomical overlay inherent to projection views. This anatomical overlap can be mitigated via 3D-DSA approaches [1], in which a 3D distribution of contrast-enhanced vascularity is obtained via an analogous approach in which the mask and contrast-enhanced images are volumetric reconstructions, commonly obtained with interventional cone-beam CT (CBCT) systems. However, due to its lack of temporal information, 3D-DSA faces limitations in certain applications where the characterization of the temporal distribution of contrast is necessary. Direct quantification of the temporal variation of the contrast distribution can be achieved via 4D-DSA [2], in which a time-resolved 3D-DSA sequence is reconstructed directly from data acquired via analogous procedures to 3D-DSA.

Common to all DSA techniques is the assumption of perfect alignment between the mask and the contrast-enhanced distributions, with misalignments resulting in significant drops in image quality. Displacements between the mask and contrast-enhanced distributions can be mitigated via registration approaches. However, 3D- and 4D-DSA face further challenges associated with distortion of anatomical structures and artifacts caused by intra-scan motion, prevalent in CBCT due to its moderate acquisition time [3].

Recent work has demonstrated the potential of deep autofocus for estimation and compensation of motion in interventional CBCT, with application to rigid motion in head/brain imaging [3] and deformable motion in abdominal CBCT [4]. Deep autofocus methods are purely image-based approaches that estimate the underlying motion trajectory of anatomical structures during the scan via optimization of a learned autofocus metric acting on reconstructed image for candidate motion trajectories. Compared to conventional autofocus, deep autofocus metrics are trained to provide quantification of motion severity and anatomical plausibility of the image content via observation of large collections of motion-corrupted datasets [3,4]. A second key component of deep autofocus is the parametric model used to describe the motion trajectory which are often parametrized by relatively simple dimension-reduction models, such as b-splines and thin plate splines. However, the complexity of motion in clinical often leads to large number parameters to optimize leading to suboptimal convergence. Recent work has proposed leveraging the universal approximator property of deep neural networks to represent complex continuous functions, such as those describing the spatial-temporal characteristics of motion, via

H. Huang, A.Lu, and A. Sisniega are with the Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD. F. Gonzalez is with the Department of Neurosurgery, Johns Hopkins Hospital, Baltimore, MD. T. Ehtiati is with Siemens Medical Solutions USA Inc. Advanced Therapies, Malvern, PA. J.H. Siewerdsen is with the Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX.

Corresponding Author: A. Sisniega (email: asisnie1@jhu.edu)

implicit neural representations (INRs) [5]. INRs have found application in modeling complex spatial-temporal deformation fields for image registration tasks [6,7], demonstrating its ability to encode complex motion patterns.

This work presents a comprehensive framework to 4D-DSA for neuro-interventional applications in presence of patient motion. The proposed approach leverages deep autofocus motion compensation and INR models of motion trajectories to assists the estimation of 4D-DSA distributions from motion-corrupted CBCT datasets.

II. Material And Methods

A. Framework for Motion Compensated 4D-DSA

The proposed framework for motion-compensated 4D-DSA is illustrated in Fig. 1. The underlying 4D-DSA pipeline follows a fundamental principle similar to [8] but includes assumptions of intra- and inter-scan motion that results in distortion and artifacts in the mask and contrast-enhanced volumes, and in differences in pose between the two volumes. The proposed approach consists of the following stages: i) estimation of the intra-scan motion trajectory of the mask scan, $\boldsymbol{\phi}_{NC}(t)$; ii) joint estimation of the intra- and inter-scan motion for the contrast-enhanced volume, denoted $\boldsymbol{\phi}_{CE}(t)$, followed by subtraction of the aligned mask to define a volumetric vascular tree; and, iii) estimation of temporal variation of contrast concentration inside the defined vascular region.

The two motion trajectories are represented as time-dependent rigid transformations with 6 degrees-of-freedom (DoFs), and they are modelled using a continuous and fully differentiable INR. The INR defines a continuous function $\boldsymbol{\phi}(t)$: $\mathbb{R}^1 \mapsto \mathbb{R}^6$ to map any arbitrary point in time domain to a 6 DoF transformation. We parameterize $\boldsymbol{\phi}$ using the SIREN architecture, consisting of a multilayer perceptron (MLP) with 3 hidden layers of 20 neurons and sine activations.

Motion estimation in the mask volume is achieved via maximization of a previously developed, anatomy-aware deep autofocus metric (DL-VIF) for quantification of motion in head/brain CBCT [3]. DL-VIF is based on a convolutional neural network acting on motion-corrupted CBCT and trained to provide reference-free estimations of visual information fidelity (VIF) with a virtual motion-free counterpart.

Following previous work for compensation of rigid and deformable motion estimation in CBCT, DL-VIF was included into a fully differentiable deep autofocus framework [4]. To allow for efficient end-to-end gradient propagation to the motion trajectory, candidate reconstructions $u_{NC}(\boldsymbol{\phi}_{NC}(t))$ were obtained as a sum of N_{PAR} partial angle reconstructions (PARs), each incorporating the backprojection of N_θ/N_{PAR} consecutive CBCT projections from the motion-corrupted stack of projections in the mask scan (p_{NC}), with N_θ denoting the total number of projections in p_{NC}. The complete deep autofocus cost function included a combination of DL-VIF and a regularization term minimizing the entropy of the image gradient [9]. The final motion compensated mask reconstruction is denoted u^*_{NC}.

The second stage performs a joint deep autofocus optimization yielding a motion trajectory $\boldsymbol{\phi}_{CE}(t)$ that compensates for intra-scan motion in the contrast-enhanced volume (u_{CE}) and for pose differences between u_{CE} and u^*_{NC}. The optimization is achieved in an analogous fashion as described above, but with the motion-corrupted contrast-enhanced CBCT projections (p_{CE}) as input and deep autofocus cost function consisting of a registration loss term based on the mean squared error between $u_{CE}(\boldsymbol{\phi}_{CE}(t))$ and u^*_{NC}. The final motion compensated contrast-enhanced volume, u^*_{CE}, thus includes compensation of intra-scan motion and pose variations with the mask volume.

The motion compensated volumes are then used as the basis for: i) generation of a 3D-DSA vascular mask via subtraction of u^*_{NC} from u^*_{CE}, followed by binarization; and, ii) generation of a set of 2D-DSA projections obtained as the subtraction between p_{CE} and a registered set of mask projection views, p'_{NC}, obtained via forward projection of u^*_{NC}, incorporating the trajectory defined by $\boldsymbol{\phi}_{CE}(t)$.

The final 4D-DSA is subsequently generated via backprojection of the sequence of 2D-DSA views onto the volumetric vascular mask.

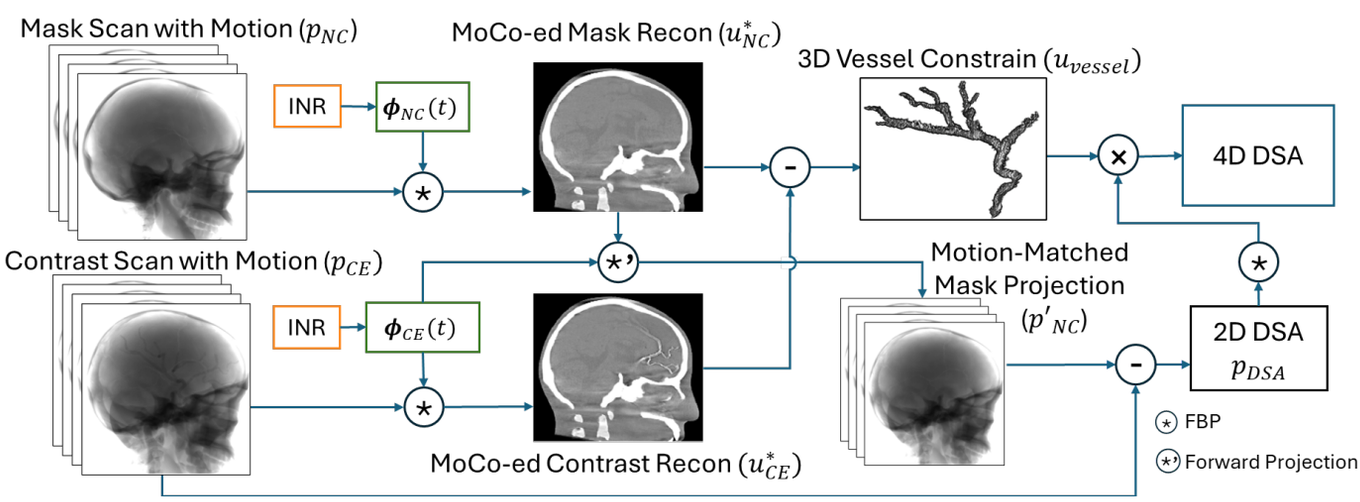


Figure 1. Deep autofocus framework for Motion Compensated 4D-DSA. The proposed framework features a combination of deep autofocus with learned anatomy-aware metrics of motion severity acting on the motion-corrupted mask volume, a joint motion compensation and registration stage between the contrast-enhanced volume and the mask, and consistency-based 4D-DSA estimation.

B. Assessment in Simulation Studies

The proposed approach was evaluated on simulated CBCT datasets based on a digital phantom derived from the CT scan of an anatomically realistic head phantom (Angiographic CT Head Phantom ACS, Kyoto Kagaku, Japan). The phantom contained contrast-enhanced vascularity, surrogate for left-anterior and left-middle cerebral arteries, and for the internal carotid artery.

For simulation of the time- (projection-) dependent contrast flow, the contrast-enhanced vascular tree was segmented, and its centerline was delineated using standard methodology, as implemented in the Vascular Modeling Toolkit [10,11]. The root of the vascular tree was selected as the entry point for the contrast distribution. The contrast injection was simulated via variation of the CT number of each voxel in the vascular tree, according to the following expression:

$$I(x,y,z,t) = \begin{cases} 1200\ HU, & \text{if } d(x,y,z) \le \dfrac{t}{T} \\ 0\ HU, & \text{if } d(x,y,z) > \dfrac{t}{T} \end{cases} \qquad (1)$$

where $I(x,y,z,t)$ is the intensity of a voxel of the vascular tree at location (x,y,z) and at time t, and $d(x,y,z)$ is the distance of (x,y,z) to the entry point, measured along the vascular tree. Finally, T is the total scan time.

Static and motion-corrupted mask and dynamic contrast-enhanced CBCT were simulated via forward projection of the contrast-free and the -dynamic digital phantoms using a geometrical configuration pertinent to robotic interventional C-arms, such as the Siemens Artis Zeego, with a source-to-axis distance of 785 mm, source-to-detector distance of 1200 mm and a flat-panel detector with 864x660 pixels, and 0.616 mm isotropic pixel size. A total of 304 projections with 260 degree-span were simulated for the total duration of the scan T = 12 s. Motion corrupted scans were generated using a projection-dependent rigid transformation with 6 DoFs. Each DoF followed a sinusoidal trajectory covering approximately 1 period/scan. Motion-corrupted contrast-free mask and dynamic contrast-enhanced scans were simulated for random motion directions and phase, with amplitude ranging from 4 mm to 12 mm in 2 mm steps. To establish a performance benchmark for comparison, mask and contrast scans without any motion were also generated to be used directly in standard 4D-DSA pipeline.

The performance of the proposed approach was evaluated via metrics of fidelity of the estimated motion trajectory (mean squared error with the ground truth trajectory), metrics of image quality (structural similarity, SSIM, against the motion-free reference), and accuracy of the estimated 4D-DSA. The accuracy of 4D-DSA was evaluated as the spatial alignment between the estimated 3D-DSA mask and the ground truth vessel mask, quantified via DICE scores, and as the temporal alignment between the 4D-DSA and the time-evolving ground-truth contrast distribution, quantified via the mean square error at 15 evenly spaced time points during the entire scan.

III. Results

A. Motion Compensation and Registration

Fig. 2(A) shows SSIM values before and after motion compensation for both the mask and contrast-enhanced volumes, as function of motion amplitude. Without motion compensation, SSIM decreases monotonically with increasing motion amplitude, indicating increasing severity of motion-induced distortion and artifacts. Motion compensation achieved consistent improvement in SSIM values for the mask volume, yielding an average SSIM = 0.946 ± 0.007 for the complete range of motion amplitudes. The joint motion compensation and registration applied to the contrast-enhanced volumes resulted in similarity on par with the static reference, yielding average SSIM = 0.933 ± 0.011, across motion amplitudes.

Further quantification on the performance of motion estimation is illustrated in Fig. 2(B), that shows the mean absolute error between the estimated and ground-truth motion trajectories. Both the mask and contrast-enhanced volumes showed errors ~ 1 mm across the complete range of motion trajectories with MAE = 0.747 ± 0.068 mm for mask scans and MAE = 0.846 ± 0.183 for contrast-enhanced volumes. It is worth noting the degradation in motion estimation with increasing amplitude, yielding a rate of increase in MAE of 0.575x and 0.783x for the mask and contrast-enhanced volumes, respectively. This rate of increase is obviously low comparing to the 3-fold rate of increase in motion amplitude.

Fig. 3 shows example image results from a case with 8 mm motion amplitude. Prior to motion compensation, both the mask (Fig. 3(A)) and contrast-enhanced volumes (Fig 3(C)) show streak artifacts, particularly conspicuous at the skull base, and

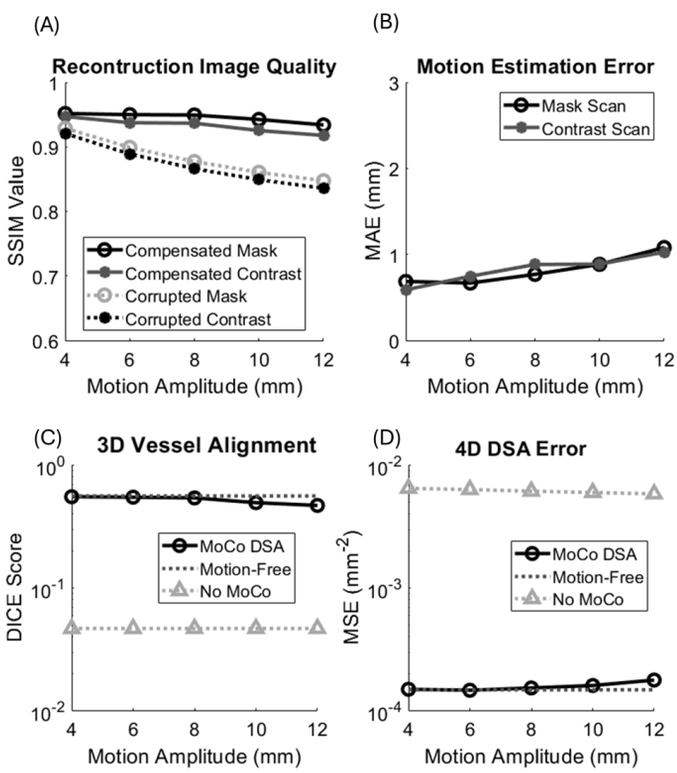


Figure 2. Quantitative analysis of motion estimation fidelity and DSA fidelity. (A) SSIM values between motion compensated volumes and the static reference. (B) Error between the estimated and ground-truth motion trajectories. Quality of 4D-DSA distribution assessed by (C) spatial alignment of the 3D vascular distribution, quantified as the DICE score compared to the ground truth vessel tree, and (D) 4D estimation error referenced to the ground-truth simulated contrast flow.

the blurring and distortion of the contrast-enhanced vasculature. Motion compensation, Fig. 3(B)-3(D) resulted in noticeable reduction of motion artifacts.

B. 4D-DSA Image Quality

The spatial alignment between the estimated and ground truth vascular distributions is illustrated in Fig. 2(C). The distribution obtained in motion-free simulations (dashed black line) offers an upper boundary, i.e. benchmark, to the performance in presence of motion. Gray triangular markers indicate the performance in absence of motion compensation. The 4D-DSA framework with motion compensation achieved a mean DICE score of 0.522 ($\sigma = 0.037$), close to the benchmark value of 0.561. This performance improves by one order of magnitude that achieved without motion compensation, which resulted in average DICE = 0.047.

A similar trend was observed for fidelity of the estimated 4D contrast distribution, shown in Fig. 2D. Motion compensated 4D-DSA achieved near benchmark results with mean of absolute error of 1.577×10^{-4} ($\sigma = 0.124\times10^{-4}$) mm^{-2}, compared to 1.482×10^{-4} mm^{-2} obtained using motion-free projections, and to 6.156×10^{-3} ($\sigma = 2.555\times10^{-4}$) mm^{-2} without compensation which is more than one order of magnitude larger.

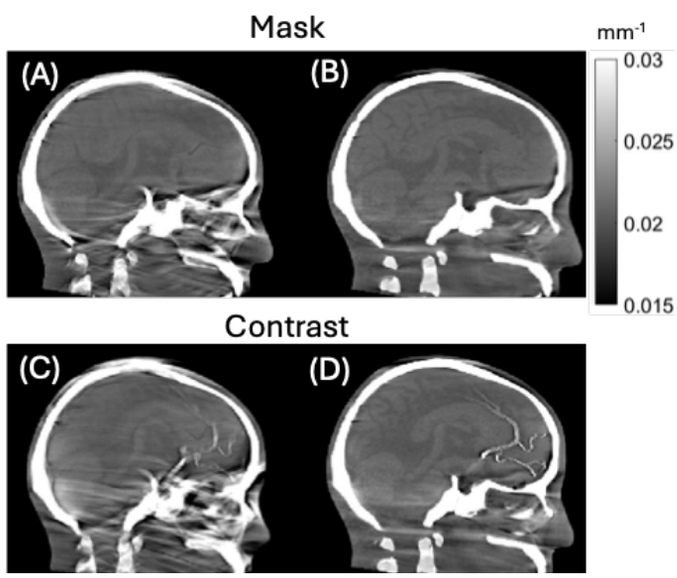


Figure 3. Image results in an example case with 8-mm motion amplitude. (A)&(C) Reconstruction without motion compensation for the non-contrast mask and for the contrast-enhanced volume, respectively. (B)&(D) Corresponding images after autofocus motion compensation and joint autofocus and registration.

Example results of a 4D-DSA sequence with 8-mm motion are shown in Fig. 4. Each sequence consists of 6 frames extracted evenly between 3s and 9s during the scan. As shown in Fig. 4(C), without motion compensation, the standard 4D-DSA produced unusable results. The proposed approach (Fig. 4(D)) resulted in a better depiction of contrast enhanced vascularity and a better agreement with the upper-boundary for performance obtained in motion-free scenarios, Fig. 4(B) and with the ground truth distribution in Fig. 4(A).

IV. Discussion And Conclusion

This work presented a novel comprehensive approach for estimation of head/brain 4D-DSA in presence of patient motion. The proposed method leverages recent work on deep autofocus motion compensation based on anatomy aware learned autofocus metrics and fully differentiable implicit neural representations of the motion trajectory.

The method was validated in simulation studies covering a wide range of motion amplitudes yielding consistent increase in image quality, sub-voxel error in estimated motion trajectories, and ~10-fold improvement in 4D-DSA accuracy, compared to conventional approaches not integrating motion compensation mechanisms.

This novel framework demonstrated a feasible solution to 4D-DSA in presence of patient motion, common in cone-beam CT. The results open a way for increasing usability and accuracy of 4D-DSA, with potential reduction of repeated scans and improvement in imaging reliability.

References

[1] R. Fahrig, et al, AJNR, (18) 1507-14 (1997)
[2] B. Davis, et al, AJNR, (34) 1914-21 (2013)
[3] H. Huang, et al, Phys. Med. Biol (67) 125020 (2022)
[4] H. Huang, et al, Proc. SPIE 1246314, (2023)
[5] V. Sitzmann, et al, arXiv:2006.09661 (2020)
[6] J. Wolterink, et al, PMLR, (172) 1349-1359 (2022)
[7] A. Lu, Proc. SPIE Medical Imaging (2024, accepted)
[8] K.L. Ruedinger, et al, AJNR, (42) 214-20 (2021)
[9] S. Ouadah, Phys. Med. Biol, (66) 055010 (2021)
[10] L. Antiga, et al, T Med Imaging, (23) 704-13 (2004)
[11] M. Piccinelli, et al, T Med Imaging, (28) 1141-55 (2009)
[12] S. Ouadah, et al, Phys. Med. Biol (61) 2613 (2016)

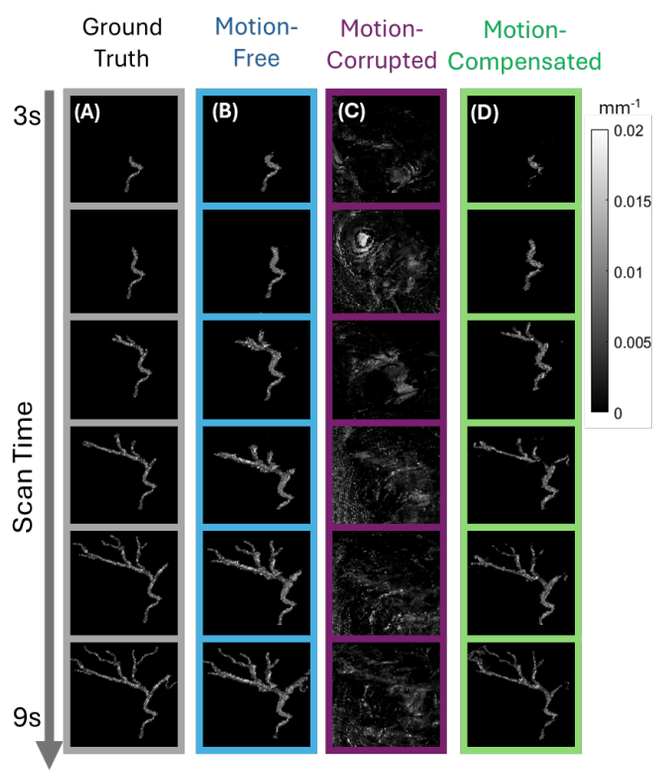


Figure 4. (A) Ground truth contrast flow used during simulation. (B) Benchmark 4D-DSA results using data without motion corruption (C) Motion compensated 4D-DSA results in presence

Dynamic CBCT Reconstruction using Prior Model-Free Spatiotemporal Implicit Neural Representation (PMF-STINR)

Hua-Chieh Shao, Tielige Mengke, Tinsu Pan, and You Zhang

***Abstract*— Dynamic CBCT is desired in clinics to capture high-spatial-resolution, time-varying dynamic images, which can guide motion monitoring, patient setup, and adaptive planning of radiotherapy. However, dynamic CBCT reconstruction is a highly ill-posed spatiotemporal inverse problem, as each CBCT volume in the dynamic sequence is only captured by one or a few X-ray projections, due to combined slow imaging speed and fast anatomical motion (e.g., breathing). We developed a machine learning-based technique (PMF-STINR) that employs a joint image reconstruction and registration approach to address the under-sampling challenge, enabling dynamic CBCT reconstruction from conventional 3D CBCT scans. Specifically, PMF-STINR uses spatial implicit neural representation (INR) to reconstruct a reference CBCT volume; and it applies temporal INR to represent the intra-scan dynamic motion with respect to the reference CBCT to yield dynamic CBCTs. To represent time-varying deformable motion, PMF-STINR couples the temporal INR with a B-spline-based, data-driven motion model. The spatial INR, the temporal INR, and the B-spline model are all learned on the fly during reconstruction, without using any patient-specific prior knowledge, surrogate signal, or motion sorting/binning. PMF-STINR was evaluated via phantom studies and on a multi-institutional dataset of real patient cone-beam projections featuring various imaging protocols. The results showed that the 'one-shot' learning-based PMF-STINR makes an accurate and robust reconstruction technique of dynamic CBCTs. It can capture regular/irregular motion with high temporal (~0.1s) resolution and sub-millimeter accuracy.**

***Index Terms*—Motion modeling, dynamic CBCT, deformable registration, image reconstruction, implicit neural representation**

I. Introduction

Cone-beam computed tomography (CBCT) is widely used in clinical practices. For CBCT applications in radiotherapy, the imaging speed is often limited by the gantry rotation speed which is generally restricted to ~6°/second (s) for patient safety. Due to the long acquisition time, patient anatomic motion, mostly respiration (3-5 s per breathing cycle), results in artifacts and blurriness in the reconstructed CBCTs. To address the artifacts and resolve the underlying motion, four-dimensional (4D)-CBCT was developed as the current clinical state-of-the-art. 4D-CBCT sorts the projections into a pre-defined set of motion bins and stacks semi-static CBCTs reconstructed from each bin to represent an averaged motion pattern. The motion sorting assumes the underlying anatomical motion is periodic and regular, which is in general false [1]. Correspondingly, 4D-CBCT cannot capture irregular motion which may significantly impact patient setup and dose delivery accuracy [2]. Moreover, 4D motion sorting is usually based on surrogates (e.g., surface markers), and can be inaccurate due to limited surrogate-anatomy motion correlation.

A fundamental solution to the limitations of 4D-CBCT is to reconstruct a *dynamic* sequence of CBCTs (one CBCT for one projection), which eliminates the uncertainties from motion sorting to capture both regular and irregular motion. Dynamic CBCTs can reveal anatomy variations with the utmost spatial and temporal resolutions. However, dynamic CBCT is not yet clinically available, mostly due to the lack of a robust technique to reconstruct CBCTs from singular X-ray projections. The information captured by a single 2D projection is extremely limited for CBCT reconstruction, as conventional reconstruction methods require hundreds of projections. Although there are numerous studies on 4D-CBCT imaging and reconstruction (e.g., [3]), the corresponding studies on dynamic CBCT reconstruction are very scarce [4-6]. The few available methods are either limited to preliminary studies based on simplified non-cone-beam geometries; or relying on specifically crafted, prior anatomical/motion models that are susceptible to differing CBCT acquisition conditions and patient anatomy/motion variations. Most of the studies were evaluated on simulated data, which did not fully demonstrate their clinical applicability and translation potential.

Our proposed study marks a general, prior model-free dynamic CBCT reconstruction approach to address the above-mentioned challenges. Specifically, our contributions include:

- We developed a 'one-shot' learning-based dynamic CBCT reconstruction technique that can reconstruct hundreds of motion-resolved CBCT volumes from a conventional 3D CBCT scan, without using any prior modeling, surrogate signal, or motion binning/sorting. The technique offers a clear advantage over the current 4D-CBCT approach and can reveal both regular and irregular volumetric motion to better inform treatment strategies and decisions.
- The proposed dynamic CBCT reconstruction technique was comprehensively evaluated using a digital phantom simulation study, a physical phantom measurement study, and a patient study using multi-institutional clinical cone-beam projections featuring combined anatomical, motion, and imaging variations.

II. Materials and Methods

A. The PMF-STINR method

PMF-STINR adopts a joint reconstruction and deformable registration approach for dynamic CBCT reconstruction. The approach de-couples the spatiotemporal inverse problem into reconstructing a reference CBCT $\boldsymbol{I}_{ref}(\boldsymbol{x})$, and solving intra-scan motion with respect to $\boldsymbol{I}_{ref}(\boldsymbol{x})$, which can be described as a time sequence of deformation vector field $\boldsymbol{d}(\boldsymbol{x}, t)$. Here, $\boldsymbol{x}$ and t denote the voxel coordinate of the dynamic CBCT and the frame index of the dynamic sequence, respectively. The dynamic CBCT sequence $\boldsymbol{I}(\boldsymbol{x}, t)$ can be obtained by warping

The study was supported by the National Institutes of Health (R01 CA240808, R01 CA258987, R01 CA280135), and Varian Medical Systems.

H.-C. Shao, T. Mengke, and Y. Zhang are with the Department of Radiation Oncology, UT Southwestern Medical Center, Dallas, TX 75390, USA. T. Pan is with the Department of Imaging Physics, UT MD Anderson Cancer Center, Houston, TX, 77030, USA.

$\boldsymbol{I}_{ref}(\boldsymbol{x})$ with $\boldsymbol{d}(\boldsymbol{x},t)$ as

$$\boldsymbol{I}(\boldsymbol{x},t) = \boldsymbol{I}_{ref}(\boldsymbol{x} + \boldsymbol{d}(\boldsymbol{x},t)) . \quad (1)$$

To reduce the dimensionality of the solution space, PMF-STINR leveraged the inherent redundancy of anatomical motion, taking a low-rank representation for $\boldsymbol{d}(\boldsymbol{x},t)$:

$$\boldsymbol{d}(\boldsymbol{x},t) = \sum_{i=1}^{3} \boldsymbol{w}_i(t) \times \boldsymbol{e}_i(\boldsymbol{x}) , \quad (2)$$

where $\boldsymbol{e}_i(\boldsymbol{x})$ and $\boldsymbol{w}_i(t)$ respectively represent the spatial components and temporal coefficients of $\boldsymbol{d}(\boldsymbol{x},t)$. The spatial components $\boldsymbol{e}_i(\boldsymbol{x})$ act as a basis set spanning a solution subspace for $\boldsymbol{d}(\boldsymbol{x},t)$; therefore they are referred as *motion basis components*. Without utilizing any prior information, PMF-STINR directly learned $\{\boldsymbol{e}_i(\boldsymbol{x})\}_{i=1}^{3}$ from the acquired projection data. A B-spline-based parametrization of $\boldsymbol{e}_i(\boldsymbol{x})$ was used to reduce the dimensionality. The above strategies of dimension reduction render the dynamic CBCT reconstruction a significantly more tractable problem.

1) Overview of the PMF-STINR workflow

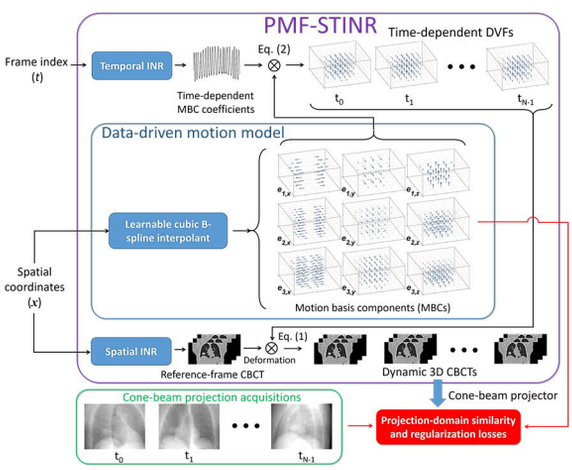


Fig. 1. Workflow of PMF-STINR. PMF-STINR solves a sequence of dynamic CBCTs, by using a spatial INR to reconstruct a reference CBCT and a temporal INR as well as a data-driven B-spline-based motion model to represent the intra-scan motion. DVFs: deformation vector fields.

Figure 1 illustrates the PMF-STINR workflow, which comprises a spatial implicit neural representation (INR), a temporal INR, and a learnable cubic B-spline interpolant for data-driven motion modeling. The spatial INR reconstructs a reference CBCT $\boldsymbol{I}_{ref}(\boldsymbol{x})$ of the dynamic sequence, and the temporal INR, along with the data-driven motion model, solves the time-dependent motion field $\boldsymbol{d}(\boldsymbol{x},t)$. The motion model learns the data-driven motion components $\{\boldsymbol{e}_i(\boldsymbol{x})\}_{i=1}^{3}$ from the cone-beam projections, while the temporal INR captures the corresponding time-varying coefficients $\{\boldsymbol{w}_i(t)\}_{i=1}^{3}$. PMF-STINR jointly trains the spatial and temporal INRs and the motion model, reconstructing $\boldsymbol{I}_{ref}(\boldsymbol{x})$ and solving the intra-scan motion simultaneously. The network training is driven by projection-domain data fidelity loss and spatiotemporal regularization losses. The sections below elaborate on each PMF-STINR component and the training strategy.

2) Spatiotemporal implicit neural representation

Figure 2 presents details on the spatial INR (Fig. 2a) and temporal INR (Fig. 2b). The spatial INR takes a voxel coordinate $\boldsymbol{x}$ of the reference CBCT as input, and maps $\boldsymbol{x}$ to the attenuation coefficient $\boldsymbol{I}_{ref}(\boldsymbol{x})$. The temporal INR takes a frame index t as input, and the outputs are the time-dependent coefficients $\{\boldsymbol{w}_i(t)\}_{i=1}^{3}$. The spatial INR comprises a learnable multiresolution hash encoding [7] and a multilayer perceptron (MLP) acting as a non-parametric universal function approximator. The temporal INR comprises a hash encoding and nine MLPs, each of which corresponds to a Cartesian component (x, y, z) of $\{\boldsymbol{e}_i(\boldsymbol{x})\}_{i=1}^{3}$. As MLP alone is ineffective in learning high-frequency details, hash encoding helps to facilitate fine detail learning [7]. The hash encoding maps a queried input ($\boldsymbol{x}$ or t) to a high-dimensional feature vector, via a multilevel encoding scheme and learning-based hash tables [7]. With hash encoding, the MLPs of the INRs can be made compact and thus memory efficient.

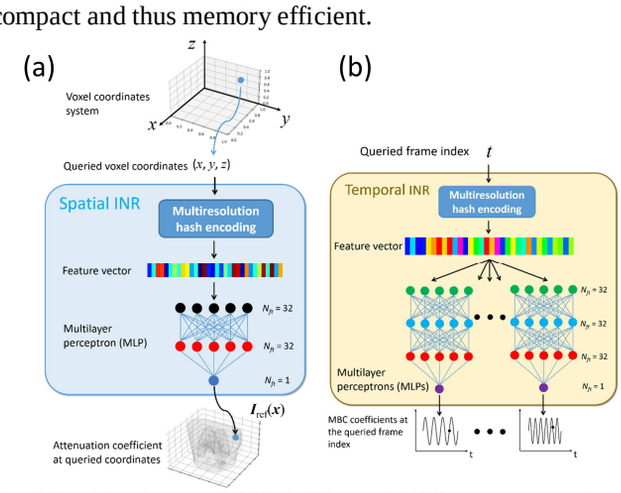


Fig. 2. Spatial and temporal INRs. (a) The spatial INR reconstructs a reference CBCT $\boldsymbol{I}_{ref}(\boldsymbol{x})$ by learning a continuous mapping from the voxel coordinate system to the linear attenuation coefficient, utilizing multiresolution hash encoding and a multilayer perceptron (MLP). (b) The temporal INR maps a frame index t to the temporal coefficient $\boldsymbol{w}_i(t)$ of $\{\boldsymbol{e}_i(\boldsymbol{x})\}_{i=1}^{3}$, using multiresolution hash encoding and nine MLPs. Each of the nine MLPs corresponds to a Cartesian component of $\{\boldsymbol{e}_i(\boldsymbol{x})\}_{i=1}^{3}$.

3) Data-driven motion modeling

Coupled with the temporal INR, the data-driven motion model of PMF-STINR learns $\{\boldsymbol{e}_i(\boldsymbol{x})\}_{i=1}^{3}$ directly from the cone-beam projections. PMF-STINR approximates $\{\boldsymbol{e}_i(\boldsymbol{x})\}_{i=1}^{3}$ using cubic B-spline interpolants with learnable control points, ensuring a continuous and smooth representation. To account for the complexity of anatomical motion, a hierarchical multiresolution strategy is employed, assigning $\boldsymbol{e}_i(\boldsymbol{x})$ (i = 1-3) to different spatial resolutions and motion scales (Fig. 1). Each Cartesian component of $\boldsymbol{e}_i(\boldsymbol{x})$ is represented by a B-spline interpolant of a different grid resolution. The learned $\{\boldsymbol{e}_i(\boldsymbol{x})\}_{i=1}^{3}$ of all resolution levels are weighted and summed by the solved $\boldsymbol{w}_i(t)$ (Eq. 2) to capture both global and local motion.

4) The progressive training strategy and loss functions

PMF-STINR uses a training strategy that progressively increases the learning complexity through five stages with two spatial resolutions to avoid local optima. Starting from a low-resolution scale, Stages I to III reconstruct a down-sampled reference CBCT and the corresponding motion model. The learned motion model and the reference CBCT are further fine-tuned at the targeted high-resolution scale in Stages IV and V. The high-resolution scale doubles the resolution from the lower-scale reconstruction. Specifically, Stage I initializes the low-resolution spatial INR by using all projections for

reconstruction, without considering the temporal motion. In Stage II, the temporal INR and the data-driven B-spline motion model are further initialized, with respect to the reference CBCT (spatial INR) obtained from Stage I. The spatial INR is frozen at Stage II to prevent the interplay between the spatial and temporal INRs. Stage III performs joint training, unfreezing all components in PMF-STINR for simultaneous image reconstruction and registration. In Stage IV, the temporal INR and motion model are frozen to optimize the spatial INR at a targeted, higher-resolution scale. Finally, at Stage V, the spatial and temporal INRs and the motion model are jointly trained to further fine-tune the reconstruction at the high-resolution scale.

B. Evaluation datasets and schemes

PMF-STINR was evaluated using three datasets: the extended cardiac torso (XCAT) digital phantom; the dynamic thorax CIRS 008A physical phantom; and a multi-institutional dataset of real patient scans.

1) XCAT simulation study

We simulated an XCAT phantom covering the thoracic-abdominal anatomy. A spherical tumor 30-mm in diameter was inserted into the lower right lung lobe for motion tracking and assessment. Six free-breathing scenarios (X1-X6) featuring various motion patterns were simulated to assess the accuracy of PMF-STINR to capture different motion. The scan took 60 s, covering a 360° scan angle (6°/s gantry speed) and using a frame rate of 11 fps.

The reconstructed dynamic CBCTs (with an isotropic 2-mm resolution) were quantitatively evaluated using the structural similarity index (SSIM) and the image relative error (RE). The accuracy of solved motion was evaluated by contour-based metrics: tumor center-of-mass error (COME) and Dice similarity coefficient (DSC).

2) CIRS measurement study

A dynamic thorax physical phantom (CIRS) was employed in a clinical measurement study. A spherical tumor with electron density similar to that of the phantom body was placed in the left lung. The tumor motion is driven by an actuator and can be customized. Six motion trajectories (C1-C6) were used in the CIRS phantom study. The cone-beam projections of the dynamic phantom were acquired on a Varian VitalBeam system (Varian Medical Systems, Inc.) in the half-fan mode. Each scan took about 1 min for a 360° scan angle, acquiring 894-896 projections (15 fps).

The CBCTs were reconstructed with an isotropic 2-mm resolution. The solved tumor motion was compared against the programmed trajectories. Tumor COMEs in three Cartesian directions were evaluated. The Pearson correlation coefficient between the solved and the 'ground-truth' tumor motion of the superior-inferior (SI) direction was computed.

3) Patient study

PMF-STINR was further evaluated on a multi-institutional patient dataset (Table I). A total of 12 cone-beam projection sets from eight patients were curated from three sources. The CBCTs were reconstructed with an isotropic 2-mm resolution.

TABLE I. Summary of imaging parameters of the patient study.

Patient ID[a]	Scan mode	Projection size[b]	Pixel size (mm^2)	kVp/mA/mS	SAD[c] (mm)/ SDD (mm)
P1	Full fan	512×384×1983	0.776×0.776	120/80/25	1000/1500
P2	Full fan	512×384×2729	0.776×0.776	120/80/25	1000/1500
P3	Full fan	512×384×1653	0.776×0.776	120/80/25	1000/1500
P4	Full fan	512×512×1015	0.8×0.8	125/20/20	1000/1536
P4-S	Full fan	512×512×340	0.8×0.8	125/20/20	1000/1536
P5	Full fan	512×512×1005	0.8×0.8	125/20/20	1000/1536
P5-S	Full fan	512×512×340	0.8×0.8	125/20/20	1000/1536
P6	Half fan	1024×768×895	0.388×0.388	125/15/20	1000/1500
P7	Half fan	1006×750×2416	0.388×0.388	120/20/20	1000/1500
P7-S	Half fan	1006×750×679	0.388×0.388	120/20/20	1000/1500
P8	Half fan	1006×750×2918	0.388×0.388	120/20/20	1000/1500
P8-S	Half fan	1006×750×677	0.388×0.388	120/20/20	1000/1500

[a]Patient ID with a suffix 'S' indicates a sparsely-sampled set extracted from the fully-sampled one [3].

[b]width (in pixel number) × height (in pixel number) × N_p (number of projections).

[c]SAD: source-to-axis distance. SDD: source-to-detector distance.

Since the patient study had no 'ground-truth' 3D motion for evaluation, the reconstructed reference CBCTs were visually inspected and the accuracy of intra-scan motion was evaluated in re-projected 2D planes. For each reconstructed dynamic CBCT, we re-projected it into a 2D digitally reconstructed radiograph (DRR) to match the corresponding cone-beam projection's imaging geometry. Motion tracked on the cone-beam projections and the DRRs was then compared by two methods: **(1)** *Structure-based motion evaluation.* The tracked structures include the diaphragms, lung nodules, and other lung features, applying the Amsterdam Shroud (AS) technique [8]. **(2)** *Feature point-based motion evaluation.* The second method tracked the motion of feature points that were automatically and independently extracted from both cone-beam projections and DRRs [9]. The feature point motion difference along the SI direction (localization error, LE [9]) between the cone-beam projections and DRRs was evaluated.

III. RESULTS

Due to limited space, we presented quantitative evaluation results of XCAT and CIRS studies to demonstrate the accuracy of PMF-STINR. For the patient study, both qualitative and quantitative results were presented. It took ~3 hours to reconstruct a sequence of 2000-frame CBCTs.

A. The XCAT and CIRS study results

Figure 3 compares the tracked lung tumor motion (dashed lines, by PMF-STINR) and the 'ground-truth' references (solid lines) of selected XCAT motion scenarios (X1-X6), and Table II summarizes the image quality and tumor localization accuracy metrics of the reconstructed dynamic CBCTs. PMF-STINR captured the irregular motion with a sub-voxel accuracy, and accurately reconstructed the dynamic CBCTs.

Table III presents the CIRS tumor center-of-mass localization errors in the left-right (LR), anterior-posterior (AP), and SI directions and the Pearson correlation coefficients (CC) between the solved and the 'ground-truth' SI trajectories.

A tracking error within or around 1 mm was achieved for all three Cartesian directions.

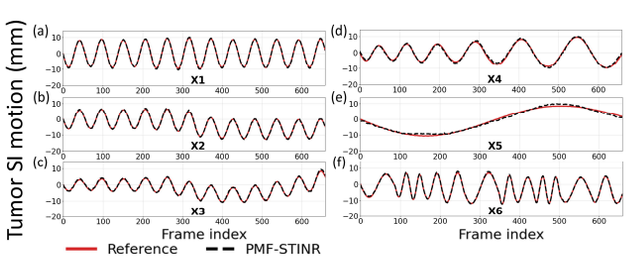


Fig. 3. Comparison of tumor superior-inferior (SI) trajectories of the XCAT study (X1-X6 scenarios), between PMF-STINR and the 'ground-truth' references.

TABLE II. Image quality and lung tumor tracking accuracy metrics (Mean± S.D.) of the reconstructed dynamic CBCTs of the XCAT study.

Motion scenario	Image quality		Tumor motion tracking	
	RE	SSIM	COME (mm)	DSC
X1	0.166±0.014	0.993±0.002	0.8±0.5	0.941±0.017
X2	0.169±0.011	0.993±0.001	0.8±0.4	0.932±0.012
X3	0.163±0.006	0.994±0.001	0.9±0.5	0.931±0.017
X4	0.198±0.014	0.989±0.002	1.4±0.8	0.921±0.029
X5	0.189±0.019	0.990±0.003	0.8±0.3	0.938±0.010
X6	0.174±0.007	0.992±0.001	0.7±0.3	0.943±0.011

TABLE III. Lung tumor tracking accuracy of the CIRS study.

Motion scenario	Pearson CC (SI trajectory)	Tumor localization error		
		LR (mm)	AP (mm)	SI (mm)
C1	0.997	0.4±0.4	1.1±0.7	1.2±0.7
C2	0.999	0.4±0.2	1.0±0.7	1.0±0.3
C3	0.999	0.5±0.6	1.0±0.6	1.0±0.4
C4	0.999	0.3±0.3	1.1±0.9	1.0±0.2
C5	0.959	0.4±0.4	0.2±0.2	1.1±1.1
C6	0.979	0.2±0.3	1.1±0.4	0.9±0.6

B. The patient study results

Figure 4 presents the reference CBCTs reconstructed for the patient study. The reference CBCTs of the fully- and sparsely-sampled acquisitions (P4, P5, P7, and P8) are comparable in image quality, showing that PMF-STINR allows dynamic CBCT reconstruction from sparsely-sampled 3D CBCT scans (~340 total projections).

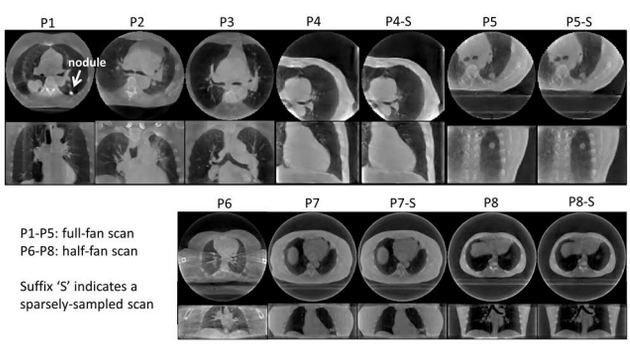


Fig 4. Reference CBCTs recon-ed by PMF-STINR for the patient study.

Figure 5 compares the tracked and the reference SI trajectories of various lung anatomies (e.g. lung nodule, diaphragm) by the structure-based motion evaluation (AS image-based tracking). Table IV shows the accuracy of motion solved by PMF-STINR, using both the AS image-based and the feature point-based tracking methods. The localization errors were calculated in projected 2D planes, thus including the magnification factor (~1.5) due to the imaging trajectory. On average a sub-millimeter accuracy was achieved, after accounting for the factor. Figure 6 plots the dynamic CBCTs reconstructed by PMF-STINR for the P1 patient.

TABLE IV. Motion tracking accuracy of the patient study.

Patient ID	Pearson CC (SI trajectory)	Localization error (mm)	
		AS image-based tracking	Feature point tracking
P1	0.962	1.2±1.3	0.8±1.0
P2	0.985	1.2±1.3	0.7±1.4
P3	0.945	1.0±1.3	1.4±2.1
P4	0.979	1.0±0.9	1.7±1.6
P4-S	0.962	1.3±1.2	1.5±1.5
P5	0.981	2.2±2.1	1.4±1.9
P5-S	0.984	1.9±1.7	1.2±1.6
P6	0.989	1.2±1.0	1.0±0.7
P7	0.994	1.3±1.3	2.0±1.8
P7-S	0.992	1.5±1.4	1.8±1.7
P8	0.994	1.2±1.0	1.5±1.3
P8-S	0.992	1.4±1.5	1.5±1.2

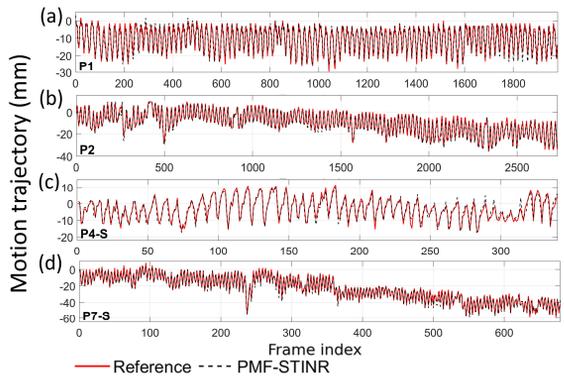


Fig. 5. Comparison between tracked and reference SI trajectories of four cases using the AS image-based tracking method, between the PMF-STINR curves (extracted from the DRRs from reconstructed dynamic CBCTs) and the reference curves (extracted from the cone-beam projections).

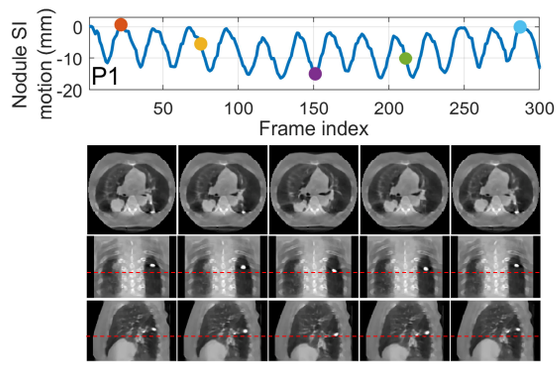


Fig. 6. Dynamic CBCT recon-ed by PMF-STINR for the P1 patient. First row shows the motion curves along the SI direction, with the dots indicating the motion states selected for plotting. In the following rows, PMF-STINR CBCTs of the selected motion states were shown.

IV. Conclusion

A prior model-free, spatiotemporal INR method was developed to reconstruct dynamic CBCTs from conventional 3D-CBCTs in a 'one-shot' fashion, which can capture various motion irregularities to inform better motion management.

References

[1] K. Yasue *et al.*, "Quantitative analysis of the intra-beam respiratory motion with baseline drift for respiratory-gating lung stereotactic body radiation therapy," *J Radiat Res*, vol. 63, no. 1, pp. 137-147, Jan 20, 2022.

[2] N. Clements *et al.*, "The effect of irregular breathing patterns on internal target volumes in four - dimensional CT and cone - beam CT images in the context of stereotactic lung radiotherapy," *Medical physics*, vol. 40, no. 2, pp. 021904, 2013.

[3] C. C. Shieh *et al.*, "SPARE: Sparse-view reconstruction challenge for 4D cone-beam CT from a 1-min scan," *Med Phys*, vol. 46, no. 9, pp. 3799-3811, Sep, 2019.

[4] H. Gao *et al.*, "Principal component reconstruction (PCR) for cine CBCT with motion learning from 2D fluoroscopy," *Medical physics*, vol. 45, no. 1, pp. 167-177, 2018.

[5] C. Jailin *et al.*, "Projection-based dynamic tomography," *Physics in Medicine & Biology*, vol. 66, no. 21, pp. 215018, 2021.

[6] Y. Zhang *et al.*, "Dynamic cone-beam CT reconstruction using spatial and temporal implicit neural representation learning (STINR)," *Phys Med Biol*, vol. 68, no. 4, Feb 6, 2023.

[7] T. Muller *et al.*, "Instant Neural Graphics Primitives with a Multiresolution Hash Encoding," *Acm T Graphic*, vol. 41, no. 4, Jul, 2022.

[8] L. Zijp *et al.*, "Extraction of the respiratory signal from sequential thorax cone-beam X-ray images," *ICCR*, pp. 507-509, 2004.

[9] R. Wei *et al.*, "Real-time tumor localization with single x-ray projection at arbitrary gantry angles using a convolutional neural network (CNN)," *Physics in Medicine and Biology*, vol. 65, no. 6, pp. 065012, Mar, 2020.

Detector Deadtime Estimation in Photon Counting CT: A Novel Image-Domain Approach

Linying Zhan, Guang-Hong Chen, Ke Li

I. Introduction

THE deadtime of a photon counting detector (PCD) refers to the period during which the detector is unable to register additional photon events after detecting one. In the realm of medical PCD-CT imaging, understanding PCD deadtime is crucial for several reasons. First, it plays a critical role in correcting count losses induced by pulse pileup. This correction ensures that the pixel values in reconstructed PCD-CT images remain consistent across varying X-ray flux levels. Second, deadtime serves as a metric for assessing the performance of the PCD, enabling researchers and engineers to evaluate the efficiency of the detector under diverse conditions and implement necessary improvements. For instance, minimizing deadtime becomes a key objective in choices related to sensor material selection and the design of the application-specific integrated circuit for a PCD. Third, a thorough understanding of deadtime is vital for optimizing data acquisition protocols, including the selection of tube current (mA) and detector operation mode. This knowledge is instrumental in refining parameters to enhance the overall efficiency of clinical PCD-CT imaging processes.

In determining the deadtime, denoted as τ, for a particular PCD at a specific energy threshold, the classical approach involves measuring PCD counts at various photon flux or mA levels. This process relies on establishing a parametric model that relates the mean output counts ($\overline{N}_{\text{out}}$) to the flux. The value of τ is then derived through curve fitting or solving an objective function based on this model. For nonparalyzable PCDs, as utilized in existing Siemens whole-body PCD-CTs [1], the parametric model is expressed as follows:

$$\overline{N}_{\text{out}} \approx \frac{nt}{1+n\tau}, \tag{1}$$

where n is the expected number of input photon events per unit time per detector pixel, and t is the measurement time per PCD frame. Based on Eq. (1), τ can be estimated via:

$$\hat{\tau} = \arg\min_{\tau} \left\| n\tau + 1 - nt\overline{N}_{\text{out}}^{-1} \right\|_2^2. \tag{2}$$

A similar objective function is available for paralyzable PCDs. The traditional approach requires access to PCD counts, namely N_{out}. **While this was feasible for users of standalone PCD systems, the situation is different for end users of clinical PCD-CT systems. In current PCD-CT systems used in routine clinical settings, detector counts are typically encrypted and accessible only to scanner developers and a limited number of close collaborators. Consequently, the majority of PCD-CT users remain unaware of the deadtime of their scanner, a critical parameter for evaluating scanner performance.** The purpose of this study was to develop a new method that relies solely on reconstructed PCD-CT images (specifically, their noise properties) to estimate τ, thereby eliminating the need for raw counts access.

L. Zhan, G.-H. Chen, K. Li are with the Department of Medical Physics and the Department of Radiology, University of Wisconsin-Madison, USA.

II. Theoretical Methods

To develop the desired image-domain deadtime estimation methods for PCD-CT systems, it is imperative to understand how the deadtime impacts experimentally measurable quantities in the reconstructed images. For this purpose, this section first presents how the noise variance of the log-processed sinogram data depends on both the tube current (mA) and deadtime. This is discussed in the first subsection. The second and third subsections elaborate on how the noise variance of the reconstructed CT image is influenced by mA and deadtime without or with pileup corrections. Finally, we present a method to experimentally estimate a critical parameter from noise power spectrum (NPS) measurements. This enables deadtime estimation without the need for raw counts access.

A. Analyzing the Impact of Deadtime and mA on the Noise Variance in Sinogram Projection Data

Carloni et al. [2] and Muller [3] previously demonstrated that, for a nonparalyzable PCD, the variance of PCD output, N_{out}, is related to the deadtime τ by

$$\sigma^2_{N_{\text{out}}} \approx \frac{nt}{(1+n\tau)^3}. \tag{3}$$

PCD counts N_{out} can be used to generate sinogram projection data as $P = \ln \frac{\overline{N}^{\text{air}}_{\text{out}}}{N_{\text{out}}}$, where $\overline{N}^{\text{air}}_{\text{out}}$ represents the expected number of PCD counts in an empty scan (air scan). Utilizing the Taylor series expansion of $\ln N_{\text{out}}$, the variance of P can be approximated by

$$\sigma^2_P \approx \frac{1}{nt} \cdot \frac{1}{1+n\tau}. \tag{4}$$

For a given PCD-CT scanner, image object, gantry rotation speed, tube potential (kV), and beam filtration, the true photon event rate (n) is linearly proportional to the tube current (mA):

$$n = \beta(\text{mA}), \tag{5}$$

where β is a factor dependent on the image object, scanner, kV, and beam filter. Consequently, the variance of the sinogram in Eq. (4) can be expressed as

$$\sigma^2_P \approx \frac{1}{\beta(\text{mA})t} \cdot \frac{1}{1+\beta(\text{mA})\tau}. \tag{6}$$

Hence, the variance of PCD-CT sinogram data combines the classical sinogram noise model, $[\beta(\mathrm{mA})t]^{-1}$, with a deadtime-dependent factor, $[1+\beta(\mathrm{mA})\tau]^{-1}$. As $\tau \to 0$, Equation (6) simplifies to the classical sinogram noise model, i.e, $\sigma_P^2 \approx \frac{1}{\beta(\mathrm{mA})t}$.

B. Analyzing the Impact of Deadtime and mA on the Variance of PCD-CT Images: Without Pileup Correction

If a linear or quasi-linear algorithm is used to reconstruct a PCD-CT image, $I(x,y)$, from projections acquired at M view angles, the image can be written as a weighted summation of the projection data, namely

$$I(x,y) = \sum_{m=1}^{M}\sum_{s=1}^{S} \lambda_{ms}(x,y)P(m,s) = \sum_{k=1}^{MS} \lambda_k(x,y)P_k, \quad (7)$$

where m denotes the view index, s denotes the detector element index, S denotes the total number of detector elements in a given detector row, λ_{ms} is a system geometry- and filter-dependent weighting factor. For a given point (x,y) in the image, a cascaded systems analysis reveals that the image variance, σ_I^2, is related to τ as follows:

$$\sigma_I^2 \approx \sum_{k=1}^{MS} \frac{\lambda_k^2}{n_k t}\cdot\frac{1}{1+n_k\tau} = \frac{1}{\mathrm{mA}}\sum_{k=1}^{MS}\frac{\lambda_k^2}{\beta_k t}\cdot\frac{1}{1+\beta_k \mathrm{mA}\tau}, \quad (8)$$

where the summation over k includes all rays contributing to the reconstruction of image pixels. When $\tau \to 0$, Eq. (8) reduces to the classical CT noise-mA scaling law of $\sigma_I^2 \propto \frac{1}{\mathrm{mA}}$.

When a uniform cylindrical phantom is used jointly with a bowtie filter (always available in clinical scanners), n_k can be approximated as a constant n for all k, leading to the approximation of Eq. (8):

$$\sigma_I^2 \approx \frac{\alpha}{nt}\cdot\frac{1}{1+n\tau} = \frac{\alpha}{\beta\mathrm{mA}t}\cdot\frac{1}{1+\beta\mathrm{mA}\tau}, \quad (9)$$

where $\alpha = \sum_k \lambda_k^2$. Compared with the classical CT noise model of $\sigma_I^2 \approx \frac{\alpha}{\beta\mathrm{mA}t}$, Eq. (9) introduces a novel multiplicative correction factor of $\frac{1}{1+\beta\mathrm{mA}\tau}$ due to the finite deadtime τ of the PCD.

C. Analyzing the Impact of Deadtime and mA on the Variance of PCD-CT Images: With Pileup Correction

For a nonparalyzable PCD, according to Eq. (1), the true input event number, nt, is related to the expected value of N_{out} by

$$nt = \frac{\overline{N_{\text{out}}}}{1-\frac{\tau}{t}\overline{N_{\text{out}}}}. \quad (10)$$

Clinically, with only a single measurement of N_{out}, a statistical estimator of nt can be constructed as

$$N_c(N_{\text{out}}) = \frac{N_{\text{out}}}{1-\frac{\tau}{t}N_{\text{out}}}. \quad (11)$$

N_c represents the counts post-pileup correction. Using the first-order Taylor expansion of $N_c(N_{\text{out}})$ at $\overline{N_{\text{out}}}$, the variance of N_c can be approximated as

$$\sigma_{N_c}^2 \approx nt(1+n\tau) = \beta\mathrm{mA}t[1+\beta\mathrm{mA}\tau]. \quad (12)$$

Furthermore, the variance of the post-pileup correction sinogram data, P_c, is given by

$$\sigma_{P_c}^2 = \sigma_{\ln N_c}^2 = \frac{1}{nt}+\frac{\tau}{t} = \frac{1}{\beta(\mathrm{mA})t}+\frac{\tau}{t}. \quad (13)$$

Analysis of the noise propagation from the sinogram to the reconstructed image leads to the following model for the variance of the corrected PCD-CT image, I_c:

$$\sigma_{I_c}^2 \approx \sum_k \frac{\lambda_k^2}{n_k t}+\frac{\tau}{t}\sum_k \lambda_k^2. \quad (14)$$

When a bowtie filter and a uniform cylindrical phantom are used, $n_k \approx n$ for all k, leading to the following simplification of Eq. (14):

$$\sigma_{I_c}^2 \approx \frac{\alpha}{nt}+\frac{\alpha\tau}{t} = \frac{\alpha}{\beta\mathrm{mA}t}+\frac{\alpha\tau}{t}, \quad (15)$$

which introduces a novel additive offset of $\frac{\alpha\tau}{t}$ that deviates from the classical CT noise model.

D. Estimation of the β Parameter to Enable the Proposed Image-Domain Detector Deadtime Measurement Method

As indicated in Eq. (9) for the case without pileup corrections and in Eq. (15) for the case with pileup corrections, **as discussed in the previous two subsections, estimating the proportionality parameter β between the photon flux n and the tube current (mA), as shown in Eq. (5), is crucial. This estimation allows for the use of Eq. (9) and Eq. (15) as theoretical foundations to estimate detector deadtime from image noise variance.** It is important to note that the parameter α is a global constant that can be scaled out in the measured noise variance level.

In this section, we introduce a novel method to estimate the parameter β. This method involves measuring the NPS. By extracting this parameter β, we can estimate the detector deadtime from the reconstructed images using Eqs. (9) and (15).

1) PCD-CT Systems Without Pileup Correction: Using Eq. (5), we can eliminate the parameter β in Eq. (9) to obtain the following equation:

$$(nt\sigma_I^2)^{-1} \approx \frac{\tau}{\alpha}n+\frac{1}{\alpha}. \quad (16)$$

The parameter t in the above equation can be obtained by dividing the gantry rotation speed with M (the number of projection frames per rotation). For a given tube potential (kV), mA, gantry rotation speed, bowtie filter, and image object, the value of n can be estimated based on the quantitative relationship between nt and the CT noise power spectrum (NPS). As previously shown by Hanson [4], at low flux levels where the PCD outputs follow Poisson statistics,

$$\mathrm{NPS}(k) = \frac{\pi\Delta\rho}{Mnt}|k||G(k)G_{\text{det}}G_{\text{int}}|^2, \quad (17)$$

where $|k| = \sqrt{k_x^2+k_y^2}$ is the radial frequency, $\Delta\rho$ denotes the PCD pixel size along the axial direction, $G(k)$ is the window or apodizing function used in FBP, G_{det} presents the normalized noise correlation function of the detector, and

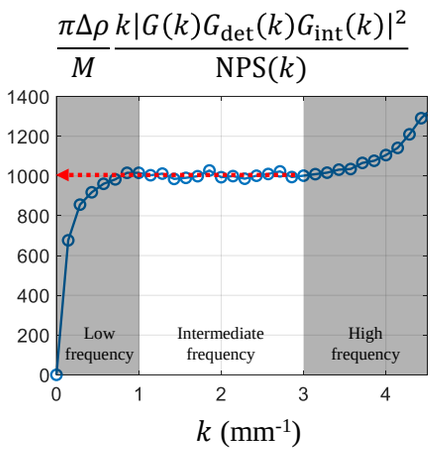


Fig. 1: Plot of the R.H.S of Eq. (18). At intermediate frequencies, the mean magnitude of $\frac{\pi\Delta\rho}{M}\frac{k|G(k)G_{\text{det}}G_{\text{int}}|^2}{\text{NPS}(k)}$ is 1,003, which is consistent with the true value of n (1,000).

G_{int} is the interpolation function used during backprojection. Note that Eq. (17) ignores the effect of discrete sampling and the associated noise aliasing, therefore it is valid only at intermediate frequencies away from 0 or the Nyquist.

Based on Eq. (17), nt can be calculated from the measured NPS as follows:

$$nt = \frac{\pi\Delta\rho}{M}\frac{|k||G(k)G_{\text{det}}G_{\text{int}}|^2}{\text{NPS}(k)}. \tag{18}$$

In practice, $|G(k)G_{\text{det}}G_{\text{int}}|^2$ can be estimated based on the knowledge that $\frac{k|G(k)G_{\text{det}}G_{\text{int}}|^2}{\text{NPS}(k)}$ should be k-independent at intermediate frequencies. An example is shown in Fig. 1.

Once nt is estimated at a low mA level with negligible pileups, its values at other mA levels can be calculated based on the mA ratio.

With n and t in hand, τ can be estimated based on Eq. (16) as follows:

1) Acquire PCD-CT images at different mA levels;
2) Measure PCD-CT variance σ_I^2 at each mA level from the acquired images;
3) Generate a scatter plot of $(nt\sigma_I^2)^{-1}$ vs. n;
4) Perform a linear regression between variables $(nt\sigma_I^2)^{-1}$ and n. According to Eq. (16), the slope of the linear fit gives $\frac{\tau}{\alpha}$, while the y-intercept of the linear fit gives $\frac{1}{\alpha}$;
5) Divide the slope by the y-intercept to obtain an estimation of τ.

2) PCD-CT Systems With Pileup Correction: Similar to the case without pileup corrections, after eliminating the β parameter in Eq. (15), one obtains the following equation:

$$\sigma_{I_c}^2 t \approx \alpha(n)^{-1} + \alpha\tau. \tag{19}$$

Therefore, to estimate τ, one can experimentally measure PCD-CT variance $\sigma_{I_c}^2$ at different n levels. One can then generate a scatter plot of $\sigma_{I_c}^2 t$ vs. n^{-1} before performing a linear regression. Based on Eq. (19), the slope of the linear fit gives α, while the y-intercept of the linear fit gives $\alpha\cdot\tau$. Therefore, by dividing the y-intercept by the slope, one obtains an estimation of τ from PCD-CT images with the pileup correction method shown in Eq. (11).

III. Validation Methods

A. Simulation Methods

Prior to our experimental study, numerical simulations with known ground truth of input count rates and PCD deadtime were performed to validate the noise models and the proposed deadtime estimation method.

In the simulation, the first step was to simulate the time intervals, Θ, between photon events, for a given expected photon event rate n. The probability density function (PDF) of Θ is $f_\Theta(\theta) = ne^{-n\theta}$. The corresponding cumulative distribution function (CDF) is $F_\Theta(\theta) = 1-e^{-n\theta}$. Based on this CDF, a random generator of Θ values was written in Matlab.

The next step was to simulate the output counts of a nonparalyzable PCD with a deadtime of τ. According to the definition of nonparalyzable detectors, a photon can be registered only if the previously-registered photon arrived at least τ earlier in time. Based on this requirement, a for-loop was implemented to sequentially go through the N incident photon events. To determine whether a given photon event can trigger an output count, its stochastic time interval relative to the preceding registered photon event was compared with τ. If the time interval exceeds τ, the digital counter of the simulated nonparalyzable PCD incremented by one count; otherwise, the counter remained in an idle state.

To generate data for testing the proposed variance models, the deadtime and data collection time were fixed at $\tau = 20$ ns and $t = 0.1$ ms, respectively, while the incident photon count rate, n, varied from 5×10^6 cps/pixel to 1×10^8 cps/pixel, where the percent count loss reached 67%. For each incident count rate, the simulation of PCD outputs was performed $1,000$ times. The resulting output counts underwent log transformation, or pileup correction and log transformation. The FBP algorithm with an apodized ramp kernel was used to reconstruct PCD-CT images from the sinogram. Variances in the projection data and PCD-CT images were measured before the proposed deadtime estimation method was applied.

B. Experimental Methods

Experimental data acquired from a clinical PCD-CT system (NAEOTOM Alpha, Siemens Healthineer) were used to validate the proposed noise models and deadtime estimation method. The data were measured at fourteen different mA levels, ranging from 40 mA to 1000 mA, with a fixed tube potential (120 kV) and rotation time ($t = 0.25$ s). The low-energy threshold of the PCD is fixed at 20 keV. The imaging object is an ACR Accreditation phantom. At each mA level, PCD-CT image noise was measured in four square regions-of-interest in the uniform section (Module 3) of the phantom.

IV. Results

The discrete data points in Figure 2 represent the variances of the simulated PCD-CT images. The continuous curves in Figs. 2 are based on the proposed theoretical variance models and align well with the discrete data points at all simulated flux levels. Using variances of the simulated PCD-CT images and the proposed method, the deadtime was estimated to be 19.3 ns (without pileup correction) or 18.9 ns (with pileup

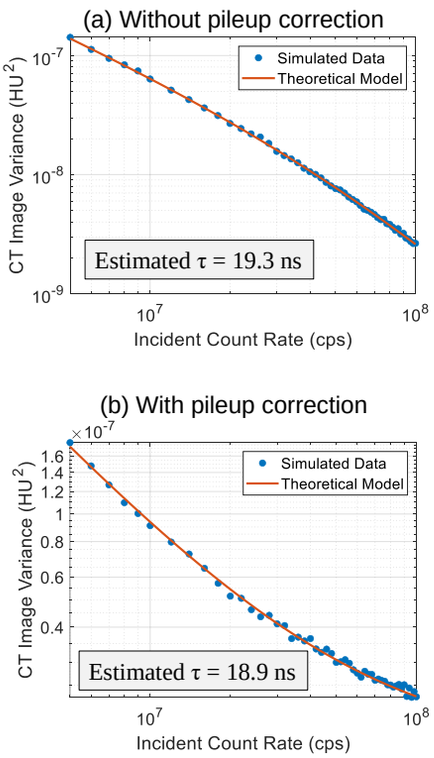


Fig. 2: Variances of PCD-CT images simulated at different input flux levels. (a) For systems without pileup correction, the simulation data points were compared with the theoretical model given in Equation (9). The estimated deadtime value is 19.3 ns. (b) For systems with pileup correction, the simulation data points were compared with the theoretical model given in Equation (15). The estimated deadtime value is 18.9 ns.

correction). In both cases, the estimated deadtime is close to the true value (20 ns). The magnitude of the relative estimation error is less than 5.5%.

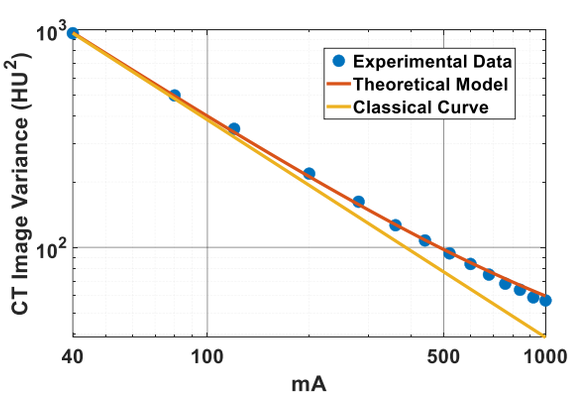


Fig. 3: Discrete data point: experimental PCD-CT noise variances measured from a Siemens NAEOTOM Alpha PCD-CT scanner. Red curve: proposed theoretical noise model (with pileup correction, Equation 15). Yellow line: the classical CT noise model, i.e., $\sigma^2 \propto \mathrm{mA}^{-1}$.

Experimental results from the clinical whole-body PCD-CT scanner are depicted in Fig. 3. The mean percent error of our noise model was only -1.4% (95% CI: [-8.0%, 5.1%]). In contrast, the classical noise model significantly underestimated PCD-CT noise with a mean percent error of 18.26% (95% CI: [-0.95%, 37.47%]). As shown in Fig. 4, the estimated deadtime of the clinical PCD-CT scanner (total energy bin, low threshold = 20 keV) is 19 ns, which is consistent with the value provided by the scanner manufacturer.

mA	PCD-CT image variance (HU^2)	Estimated n (cps)	Estimated % count loss	Estimated Deadtime (ns)
40	964.4	1.21+06	2.3%	19 ns
80	499.1	2.4E+06	4.4%	
120	350.5	3.6E+06	6.5%	
200	218.2	6.1E+06	10.3%	
280	162.1	8.5E+06	13.9%	
360	126.4	1.1E+07	17.2%	
440	107.6	1.3E+07	20.2%	
520	93.9	1.6E+07	23.0%	
600	83.9	1.8E+07	25.7%	
680	75.0	2.1E+07	28.1%	
760	68.1	2.3E+07	30.4%	
840	63.9	2.5E+07	32.6%	
920	58.9	2.8E+07	34.6%	
1000	57.1	3.0E+07	36.5%	

Fig. 4: This table lists the percent count loss at each mA level and the system deadtime (τ) estimated based on the NAEOTOM Alpha PCD-CT image variances shown in Figure 3 and the estimated true event rate (n).

V. Conclusions

The finite deadtime of PCD distorts the statistical distribution of the measured counts, leading to increased noise levels. Even when pileup correction is applied to stabilize the mean CT number across mA levels, the deadtime's influence persists in reconstructed images, causing alterations in the noise-mA relationship. Through a first principles analysis, two novel quantitative relationships between PCD-CT image noise and detector deadtime were derived—one applicable to systems without pileup correction and the other for systems with pileup correction.

Using these relationships, a new image-domain PCD deadtime estimation method was developed, eliminating the need for accessing raw counts. Validation through simulation and experimental results demonstrates the model's ability to accurately extract deadtime values from noisy PCD-CT images. This research opens a new avenue for end-users of clinical PCD-CTs seeking information on PCD deadtime, particularly for purposes like scanner comparison, quality control, and research. Additionally, the novel noise-mA relationships provide a foundation for protocol optimization in PCD-CT.

References

[1] S. Kappler, A. Henning, B. Kreisler, F. Schoeck, K. Stierstorfer, and T. Flohr, Proc. SPIE **9033**, 90331C (2014).
[2] F. Carloni, A. Corberi, M. Marseguerra, and C. Porceddu, Nuclear Instruments and Methods **78**, 70–76 (1970).
[3] J. W. Müller, Nuclear Instruments and Methods **112**, 47–57 (1973).
[4] K. M. Hanson, Medical Physics **6**, 441–451 (1979).

Application of the Calibration-based Statistical Framework for Material Decomposition to Digital Mammography: A Preliminary Study

Okkyun Lee

***Abstract*—This study proposes photon counting detector (PCD)-based digital mammography using a recently developed calibration-based statistical framework for material decomposition. The statistical framework uses calibration slab phantoms of two basis materials to build a pseudo-count generator that calculates the approximation of the mean photon counts for a given basis line integrals. Then, it performs the maximum likelihood-based material decomposition for adipose and glandular tissues to calculate breast density. Lastly, it performs low-pass filtering to reduce noise in the basis images (line integrals) and synthesizes low-energy images that mimic conventional mammogram images. We validate the method using a benchtop photon counting x-ray system and a CIRS mammo phantom. Preliminary results show that the proposed method accurately estimated the breast density. The synthesized image (at 35 keV; using 130kV) was slightly smoother than the conventional mammogram image using 40kV. However, it can potentially reduce exposure time (and mAs) and absorbed dose.**

***Index Terms*—Maximum likelihood, likelihood-based bilateral filter, material decomposition, digital mammography**

I. Introduction

DUAL-energy mammography can potentially calculate breast density more accurately than conventional mammography via material decomposition [1–3]. Furthermore, one can use the lower energy image for diagnosis as the traditional mammogram image. However, it may increase the dose level and the scan time. Photon-counting detector (PCD)-based mammography [4–8] may address the latter. Still, accurate material decomposition is challenging, and how to reduce the dose level needs to be addressed.

This study aims to reduce exposure time (and mAs) and dose level by using a high voltage in the PCD-based mammography, e.g., we used 130 kV. It performs the material decomposition of basis materials. Then, it synthesizes low-energy images to obtain high-contrast mammogram images, as in conventional digital mammography that uses a low voltage (we used 40 kV). However, it requires accurate material decomposition for estimating breast density and effective noise reduction in the synthesized low-energy images.

We propose a PCD-based mammography using a recently developed calibration-based statistical framework for material decomposition. The framework was introduced in our submitted companion abstract [9]. It aims for accurate and precise material decomposition and can be summarized as the following three steps: (1) It builds a pseudo-count generator (PCG) from a calibration process. PCG is a function of basis line integrals that calculates the approximation of the theoretical expectation of counts. We build it for each PCD pixel and energy bin. (2) It performs the maximum likelihood (ML)-based estimation for the basis line integrals using the scanned data (measured counts) and the PCGs. (3) It reduces the noise in the estimated basis sinograms using the likelihood-based bilateral filtering [10]. However, we modify this step by using simple low-pass filtering and synthesize low-energy images for diagnosis. Lastly, it converts the estimated basis line integrals to those of the adipose and glandular tissues to estimate the breast density.

We validated the method with experimental studies using a benchtop photon-counting X-ray system and a CIRS mammo phantom. The preliminary results show that the proposed method estimates the breast density close to the ground truth. The synthesized low-energy image from the proposed method is (qualitatively) slightly smoother than the conventional mammogram image using a low kV. However, we investigated the absorbed dose at different positions of the phantom and found a substantial reduction of exposure time (and mAs) and a decent decrease in the dose level.

Section II describes the proposed PCD-based mammography within the statistical framework and the experimental settings. Section III presents the results, followed by discussions and conclusions in Sec. IV.

II. Methods

A. Proposed PCD-based Mammography

The following subsections explain the calibration-based statistical frameworks for material decomposition. Then, we describe how to convert the estimated basis calibration materials to adipose and glandular tissues and how to synthesize them to low-energy images.

1) [1st Step] Pseudo-Counts Generator (PCG): We used PMMA slabs and aluminum (AL) foils as the two calibration materials, a total of 12 combinations of basis line integrals, as shown in Fig. 1(a). We scanned PMMA slabs and AL foils multiple times for each calibration point to approximate the mean photon counts for each PCD pixel and energy bin. We used the calibration data set (pairs of mean photon counts and corresponding thicknesses) to generate PCGs. The PCG is the

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2023-00249223).

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).

function of basis line integrals that approximates the mean photon counts. It interpolates or extrapolates the calibration data set to calculate the mean photon counts for arbitrary basis line integrals. More details are in our companion abstract [9]. We used linear interpolation and extrapolation.

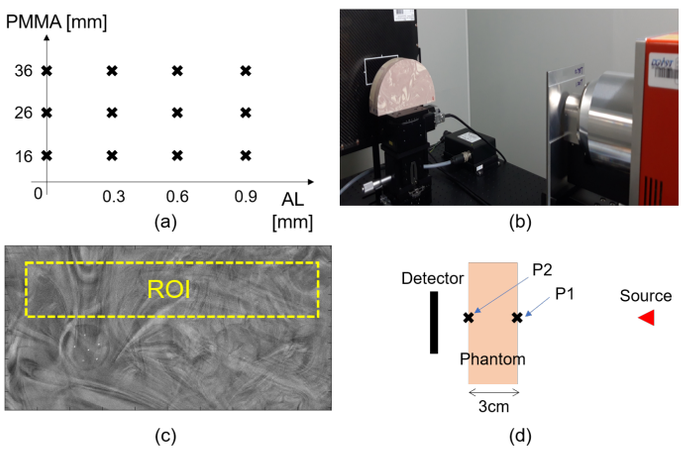


Fig. 1. (a) Basis calibration materials (PMMA and AL) used to build PCGs, (b) tabletop photon counting x-ray system for digital mammography, (c) ROI used to calculate breast density, and (d) an illustration describing the positions (P1 and P2) where we calculated the absorbed dose (air kerma for P1).

2) [2nd Step] Maximum Likelihood (ML)-based Estimator: The log-likelihood of the basis line integrals for a given PCD measurement can be described as follows:

$$\mathcal{L}(\mathbf{x}|\mathbf{y}) = \sum_{b=1}^{N_b} \left\{ -\bar{y}_b(\mathbf{x}) + y_b \ln\left(\bar{y}_b(\mathbf{x})\right) \right\}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{N_b}$ is the given PCD counts with N_b energy bins, $\mathbf{x} \in \mathbb{R}^{N_m}$ is the vector of basis line integrals with N_m materials, and $\bar{y}_b(\mathbf{x})$ is the vector of mean photon counts. The ML-based estimation requires the analytic system models to calculate the mean photon counts. However, we use the PCGs to approximate it to perform the ML-based estimation based on calibration data. We used an accelerated version of the ML in [9] due to the substantial size of the data for CBCT. Here, we used the derivative-free Nelder-Mead algorithm to search for the basis line integrals that maximize the likelihood.

3) [3rd Step] Low-pass Filtering (LPF): The third step of the original method was to apply statistical filters to the ML-based estimates of the basis line integrals. In short, the filter consists of the Gaussian kernel (location-dependent weights) and the statistical weights calculated from the neighborhood likelihoods [10]. More specifically, the noise-reduced basis line integrals can be described as follows:

$$\hat{\mathbf{x}}_j = C(j) \sum_{i \in \mathcal{N}_j} \mathcal{G}(i, j, \sigma_1) \mathcal{S}(i, j, \sigma_2) \hat{\mathbf{x}}_i^{ML}, \tag{2}$$

where $\hat{\mathbf{x}}_i^{ML}$ is the ML-based estimates for the i-th ray from the second step, $\mathcal{N}_j$ is the set of indices for the neighborhood rays (j is the center one), $C(j)$ is a normalization factor to make the sum of the weights to one, and $\mathcal{G}(i, j, \sigma_1)$ is the Gaussian kernel with a variance of σ_1^2. The statistical weight, $\mathcal{S}(i, j, \sigma_2)$, is given by,

$$\mathcal{S}(i, j, \sigma_2) = \exp\left\{ -\frac{\left(\mathcal{L}(\hat{\mathbf{x}}_i^{ML}|\mathbf{y}_j) - \mathcal{L}(\hat{\mathbf{x}}_j^{ML}|\mathbf{y}_j) \right)^2}{\sigma_2^2} \right\}, \tag{3}$$

where σ_2 controls the degree of smoothing of the filter.

However, PMMA is the dominant material used to model the adipose and glandular tissues while aluminum only slightly changes for different breast densities. In this case, the location-dependent weight (the Gaussian kernel) is dominant: for most of the portions in the scanned object (CIRS phantom), the ground truth would be the same with a very similar PMMA thickness with negligible AL thicknesses. Thus, the location-dependent weight was more effective than the statistical one in this study, and we changed the third step from the LBF to LPF by setting ($\sigma_2 \leftarrow \infty$).

4) Low-energy Image Synthesis and Data Conversion: The noise-reduced basis line integrals are the calibration materials, PMMA and AL. Let their basis line integrals be $x_{PMMA}(\mathbf{r})$ and $x_{AL}(\mathbf{r})$, respectively, where $\mathbf{r}$ indicates the specific position of the PCD pixel. We used them to synthesize the low-energy image at E [keV] as follows:

$$s^{prop}(\mathbf{r}, E, 130kV) = x_{PMMA}(\mathbf{r})\mu_{PMMA}(E) + x_{AL}(\mathbf{r})\mu_{AL}(E), \tag{4}$$

where $\mu_{PMMA}(E)$ and $\mu_{AL}(E)$ are the linear attenuation coefficients of the PMMA and AL, respectively. We use the notation for the conventional mammogram image as $s^{conv}(\mathbf{r}, 40kV)$.

Lastly, we converted the estimated basis line integrals to those of the adipose and glandular tissues, $x_a(\mathbf{r})$ and $x_g(\mathbf{r})$, to estimate the breast density by least squares fitting the linear attenuation coefficients of those materials [11].

B. Experimental Settings

We used a micro-focus X-ray source (L9181-02, Hamamatsu) with a collimator and AL prefilter (2 mm), as shown in Fig. 1(b). We used a PCD array (XC-THOR FX, Direct Conversion) with an active area of $50.6 \times 100.2\,\mathrm{mm}^2$, where each PCD has a pixel size of $100 \times 100\,\mu\mathrm{m}^2$. We used two energy thresholds (15 and 40 keV; $N_b = 2$) for the proposed method and a single energy threshold (15 keV) for the conventional mammography. Details for the mAs are given in Sec. III. The distance from the source to the detector was 67.84 cm, and the distance from the phantom (bottom) to the detector was about 9 cm.

C. Evaluations

We used breast equivalent slabs (CIRS; each slab is 1 cm thick; see Fig. 1(a)). We used three slabs (total 3 cm), with the center one including specks (CaCO3) and carcinomas. The phantom consists of the adipose and glandular equivalent materials swirled by a 50/50 ratio by weight. Thus, the ground truth of the breast density (average) in the ROI shown in Fig. 1(c) is known. We estimated the breast density from the second step of the proposed method (the ML-based estimates) and compared it to the ground truth. Then, we qualitatively compared the synthesized low-energy image to the conventional mammogram image.

We also calculated the absorbed dose at $P1$ and $P2$, as shown in Fig. 1(d), for the conventional and proposed mammography. The ratio of the absorbed doses was calculated as follows:

$$RoD(P_i) = \frac{D_{prop}(P_i, N_i^{prop}, 130kV)}{D_{conv}(P_i, N_i^{conv}, 40kV)}, \quad (5)$$

for $i = 1$ and 2, where D_{prop} and N_i^{prop} are the absorbed dose and the total number of photons measured at P_i, respectively, for the proposed method. More specifically, D_{prop} was calculated as follows [12, 13]:

$$D_{prop}(P_i, N_i^{prop}, 130kV) = \int N_i^{prop} \Phi_i(E, 130kV) E \frac{\mu_{en,i}}{\rho_i}(E) dE, \quad (6)$$

where $\Phi_i(E, 130kV)$ is the normalized photon flux density at P_i and $\frac{\mu_{en,i}}{\rho_i}(E)$ is the mass energy absorption coefficient at P_i. We assumed P_1 and P_2 were air and glandular, respectively. $\Phi_1(E, 130kV)$ was calculated as the normalized incident flux, and $\Phi_2(E, 130kV)$ was calculated as the normalized transmitted flux after passing 3 cm of adipose/glandular (50/50). D_{conv} was calculated in the same way. Note that we experimentally measured N_i^{prop} and N_i^{conv} (the distance between P_i and the detector was neglected) and simulated the normalized photon flux, Φ_i, using *SPEKTR* [14]. The mass energy absorption coefficients and linear attenuation coefficients for different materials were calculated using *PhotonAttenuation* [15].

III. Results

The calculated photon flux in (6), $N_i\Phi_i(E)$, is shown in Fig. 2 for the conventional and proposed mammography. The flux was used to calculate the absorbed dose.

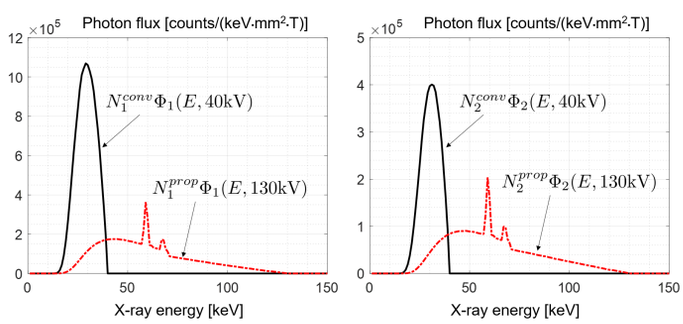


Fig. 2. Photon flux at $P1$ (left) and $P2$ (right) for the conventional and proposed mammography. We used the flux to calculate the absorbed dose. (T indicates the total exposure time that depends on the methods.)

Table I summarizes various scan conditions with measured photon flux and the calculated ratio of the absorbed dose at $P1$ and $P2$. We applied the same current (mA) for both methods. Hence, the exposure time of the proposed time is about 30 times reduced compared to the conventional method, which resulted in almost the same number of photons measured after passing the 3 cm phantom. We intentionally made a similar photon flux at $P2$ and compared the absorbed dose and the mammogram image quality. The calculated RoDs indicate that the air kerma of the proposed method is about 30%, and the absorbed dose at 3 cm depth of glandular is about 50% of the conventional method.

Table II presents the estimated breast density (defined in the first column) using the proposed method (we used the result of the second step) and the ground truth. We averaged the estimated glandular and adipose densities for 20 noise trials for each position. Then, we calculated the breast density since the 50/50 weight is not guaranteed (swirled) for each position but a large portion of the phantom. The estimated breast density is 53 [%], close to the ground truth of 50 [%].

TABLE II
Result of the breast density estimation. We considered the positions $\mathbf{r}_i$ in the ROI (see Fig. 1(c)). The estimated $\hat{x}_g$ and $\hat{x}_a$ for each position were averaged for 20 noise realizations and then calculated for the ROI. ρ_g and ρ_a are the densities of the glandular and adipose, respectively.

Breast density [%]	Ground truth	Proposed method (estimated)
$\frac{\sum_{i=1}^{M} \rho_g \hat{x}_g(\mathbf{r}_i)}{\sum_{i=1}^{M} \rho_g \hat{x}_g(\mathbf{r}_i) + \rho_a \hat{x}_a(\mathbf{r}_i)} \times 100$	50	53

We visually compare the mammogram images from the conventional and proposed methods in Fig. 3. The noise levels in the conventional and proposed methods (with noise reduction) look similar. Still, the conventional one presents sharp glandular changes, possibly due to the higher dose level. However, both lesions are easily recognizable for all methods.

IV. Discussions and Conclusions

This paper presented preliminary results for the proposed PCD-based mammography. It applies a previously developed calibration-based statistical framework for estimating basis line integrals of calibration materials, synthesizing noise-reduced low-energy mammogram images, and estimating the breast density by converting the estimated basis calibration materials. It showed a potential to use a high voltage in a digital mammography with much-reduced exposure time.

In this study, we matched the number of measured flux for the proposed and conventional methods, which caused more

TABLE I
Summary of the various scan conditions, measured photon flux, and the absorbed dose ratio (RoD). See Fig. 1(d) for the positions $P1$ and $P2$. (T indicates the total exposure time.)

	kV	mAs	N: # of measured photons ($\times 10^6$) / $T \cdot mm^2$		RoD	
			P1	P2	P1	P2
Conventional method	40	30	14.918	5.142	0.31	0.53
Proposed method	130	0.97	9.768	5.176		

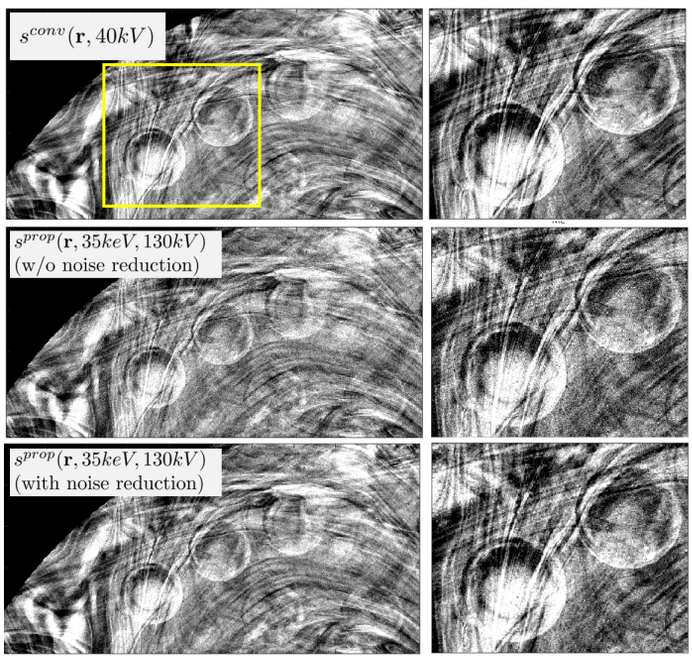


Fig. 3. Conventional mammogram image (top), synthesized low-energy image at 35 keV without noise reduction (middle; ML-based estimates), and the noise-reduced low-energy image (bottom). Zoomed images are shown on the right, focusing on the lesions.

severe noise in the proposed method. Further study is needed by changing the photon flux and different calibration materials suitable for applying the likelihood-based filtering in the third step of the proposed method. Quantitative analysis is also necessary.

References

[1] J. L. Ducote and S. Molloi, "Quantification of breast density with dual energy mammography: An experimental feasibility study," *Med. Phys.*, vol. 37, no. 2, pp. 793–801, 2010.

[2] A. R. Lam, H. Ding, and S. Molloi, "Quantification of breast density using dual-energy mammography with liquid phantom calibration," *Phys. Med. Biol.*, vol. 59, pp. 3985–4000, 2014.

[3] S. Molloi, H. Ding, H.-M. Cho, and D. Beidokhti, "Quantification of water and lipid density with dual-energy mammography: validation in postmortem breasts," *Eur Radiol.*, vol. 31, no. 2, pp. 938–946, 2021.

[4] H. Ding and S. Molloi, "Quantification of breast density with spectral mammography based on a scanned multi-slit photon-counting detector: a feasibility study," *Phys. Med. Biol.*, vol. 57, pp. 4719–4738, 2012.

[5] S. Han, "A quantification method for breast tissue thickness and iodine concentration using photon-counting detector," *J Digit Imaging*, vol. 28, pp. 594–603, 2015.

[6] K. Erhard *et al.*, "Characterization of cystic lesions by spectral mammography: Results of a clinical pilot study," *Invest Radiol.*, vol. 51, no. 5, pp. 340–347, 2016.

[7] E. Fredenberg *et al.*, "Measurement of breast-tissue x-ray attenuation by spectral mammography: solid lesions," *Phys. Med. Biol.*, vol. 61, pp. 2595–2612, 2016.

[8] B. Ghammraouia and S. J. Glick, "Investigating the feasibility of classifying breast microcalcifications using photon-counting spectral mammography: A simulation study," *Med. Phys.*, vol. 44, no. 6, pp. 2304–2311, 2017.

[9] O. Lee and J. Kim, "Calibration-based statistical framework for material decomposition in photon counting CT," *Proceedings of 8th International Conference on Image Formation in X-Ray Computed Tomography (CT meeting)*, 2024, *Submitted*.

[10] O. Lee, "Likelihood-based bilateral filters for pre-estimated basis sinograms using photon-counting CT," *Med. Phys.*, vol. 50, pp. 2733–2758, 2023.

[11] J. W. Byng, J. G. Mainprize, and M. J. Yaffe, "X-ray characterization of breast phantom materials," *Phys. Med. Biol.*, vol. 43, pp. 1367–1377, 1998.

[12] F. H. Attix, *Introduction to Radiological Physics and Radiation Dosimetry*. Weinheim: WILEY-VCH, 2004.

[13] D. R. Dance, S. Christofides, A. D. A. Maidment, I. D. McLean, and K. H. Ng, *Diagnostic Radiology Physics: A Handbook for Teachers and Students*. Vienna, Austria: IAEA, 2014.

[14] 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," *Med. Phys.*, vol. 43, no. 8, pp. 4711–4717, Aug. 2016.

[15] J. Tuszynski, "Photonattenuation," https://www.mathworks.com/matlabcentral/fileexchange/12092-photonattenuation, 2016.

Pileup-Proof Material Decomposition for Spectral CT Using a Simultaneous Photon-Counting and Energy-Integrating Readout ASIC

Yukang Song, Xuezhi Wang, Zhi Deng, and Hewei Gao*

***Abstract*—Photon-Counting Computed Tomography (PCCT) has demonstrated significant advantages and holds great promise in the field of spectral imaging. However, the inherent pulse pileup effects significantly affects counting measurements, leading to a degradation in the quality of spectral imaging. This study introduces a pileup-proof material decomposition (MD) approach for spectral CT using a simultaneous photon-counting and energy-integrating readout Application-Specific Integrated Circuit (ASIC) known as HPIX. The HPIX simultaneously record energy integration while counting photons by the introduction of charge pump circuits. Numerical calculations were conducted to explore the potential of energy integration in MD affected by pileup, followed by a comparison study to validate its efficacy. Among the MD methods utilized for comparison, three flux-independent lookup tables were generated to map measured projections to material thickness, one of which incorporates simultaneously recorded energy integration. Both the results of numerical calculations and comparison study indicate that our novel MD approach using HPIX demonstrates remarkable resilience against the adverse impact of pulse pileup. In comparison to MD using photon-counting information exclusively, the pileup-proof MD exhibits exceptional decomposition accuracy and enhanced robustness.**

***Index Terms*—spectral CT, photon-counting, material decomposition, pulse pileup, HPIX.**

I. Introduction

PULSE pileup occurs when two or more photons are detected by photon counting detector (PCD) within a short time window, resulting in a loss of recorded counts and distortion of the X-ray spectrum. Currently, two main approaches are employed for algorithmic compensation: measurement-based and model-based methods. The measurement-based approaches[1], [2], often require extensive measurement data for nonlinear interpolation correction. Deep learning is also applied to correct measured data[3], or learn the relationship between measurements and basis materials thicknesses as a MD method[4], [5]. On the other hand, model-based methods[6], [7] rely on the accuracy of the model and face high computational costs. It also require measurements for the evaluation of detector-specific properties such as dead time, energy response, etc.

This work was supported in part by the National Key R&D Program of China (No.2022YFE0131100), and the National Natural Science Foundation of China (No.U20A20169 and No.12075130).

Yukang Song, Xuezhi Wang, Zhi Deng, and Hewei Gao* are with the Key Laboratory of Particle and Radiation Imaging (Tsinghua University), Ministry of Education, China and Department of Engineering Physics, Tsinghua University, Beijing, 100084, China. (* indicates corresponding author: hwgao@tsinghua.edu.cn)

The aforementioned methods exclusively utilize photon-counting information, which is affected especially with a higher input photon flux, while traditional energy integration detection mode consistently remains faithful. Utilizing the energy integration to correct pileup-induced count losses has been shown to be feasible by introducing an readout circuit element to the circuits of paralyzable PCDs[8]. In this study, the HPIX is a simultaneous counts and charge integrating readout chip for semiconductors. The integration of both photon-counting and energy-integrating channels is achieved by the introduction of the charge pump circuits[9]. Diverging from CIX[10], this circuit design enables lower power consumption and smaller pixel dimensions. Leveraging the advantage of the ASIC, this paper proposes a pileup-proof MD method to estimate basis materials thicknesses.

II. Methods

A. HPIX-Detector and Pileup Model

We have assembled a detector comprising four modules, equipped with eight HPIX chips. Fig.1.(C) presents the physical depiction of the ASIC, illustrating its overall architecture. It is composed of readout channels and peripheral circuits. The chip's predominant portion is formed by 128 readout channels. Fig.1.(D) displays the schematic diagram of a single-channel circuit. The introduction of charge pump circuit extends the dynamic range of the integration channel without compromising the gain of the photon-counting channel.

Based on its signal readout mechanism, the HPIX-detector acquires data from both photon-counting channel and energy-

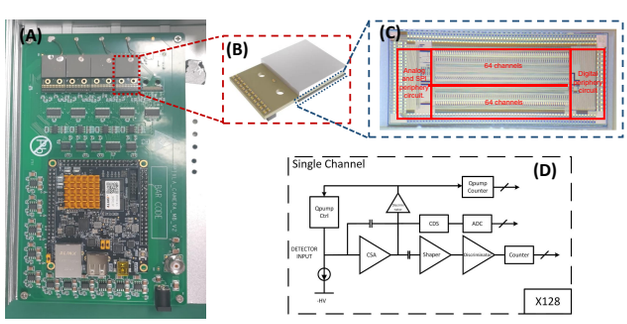


Fig. 1. HPIX-Detector: from overall physical structure to ASIC details. (A) Overall physical appearance of the detector equipped with HPIX. (B) Single detector module diagram. (C) The physical layout of the ASIC. (D) The schematic of the single channel circuit.

integrating channel, both containing identical photon information. The measurements, denoted as I_{EI} for energy-integrating channel and I_{bin_i} for photon-counting channel, are expressed as follows,

$$I_{EI} = \int_0^{E_{max}} S_0(E)D(E)W(E)e^{\int_L \mu(E,\vec{r})dl}dE. \quad (1)$$

$$I_{bin_i} = \int_{bin_i} M_{PPE}\left(S_0(E)D(E)W(E)e^{\int_L \mu(E,\vec{r})dl}\right)dE. \quad (2)$$

Where $S_0(E)$ represents the input spectrum of photons before transmission. $D(E)$ denotes the detection efficiency, which is consistent for both channels. $W(E)$ signifies the photon energy. For photon-counting channel, it is a constant of 1, indicating the data reflects the number of photon counts. $M_{PPE}(\cdot)$ is a pileup model that outputs a pileup-induced spectrum based on the input spectrum and other detector-related parameters.

This study adopted the pulse pileup probability model proposed by Taguchi et al[6]. Taking into account pulse pileup effects, the average photon count $N_{PPE}(E)$ for an energy E recorded by a PCD within a time interval Δt can be represented as follows,

$$\begin{aligned} N_{PPE}(\mathrm{E}) &= M_{PPE}(S_{in}(E)) \quad (3)\\ &= a\Delta t Pr(rec \mid a\tau)\sum_{m=0}^{\infty}(Pr(m \mid rec)Pr(E \mid m)). \end{aligned}$$

a represents the true counts rate of photons deposited in the detector, which can be obtained by integrating $S_{in}(E)$. τ denotes the detector dead time (resolution time), and the m-order pulse signifies that m+1 photons contribute to the pulse.

B. Pileup-Proof Material Decomposition

According to the MD theory, the attenuation properties of materials can be linearly combined by two basis functions.

$$\mu(E,\vec{r}) = a_1(\vec{r})\psi_1(E) + a_2(\vec{r})\psi_2(E). \quad (4)$$

Here, the basis functions are chosen to be the attenuation coefficients of the two basis materials. By substituting Eq.(4) into Eq.(1),(2), the spectral projection data acquired by the HPIX can be expressed,

$$P_{EI} = -\ln\left(\int S_{EI}(E)e^{A_1\mu_1(E)+A_2\mu_2(E)}dE\right). \quad (5)$$

$$P_{bin_i} = -\ln\left(\int S_{bin_i}(E)e^{A_1\mu_1(E)+A_2\mu_2(E)}dE\right). \quad (6)$$

with $A_1 = \int a_1(\vec{r})dl$ and $A_2 = \int a_2(\vec{r})dl$ being the corresponding material thickness. $S_{EI}(E)$ and $S_{bin_i}(E)$ are the effective energy spectrum representing different detection modes, according to Eq.(1), (2).

In the comparison study, pileup-proof MD, noted as MD_{PUP} was developed alongside two other approaches for comparison: MD_{2bins} and MD_{3bins}. The projection datasets $\boldsymbol{P}$ they utilized were $\{P_{bin_1}, P_{bin_2}, P_{EI}\}$, $\{P_{bin_1}, P_{bin_2}\}$ and $\{P_{bin_1}, P_{bin_2}, P_{bin_3}\}$, rspectively, as shown in Fig.2. Then, the pileup-proof MD in projection domain can be expressed as an inverse calculation,

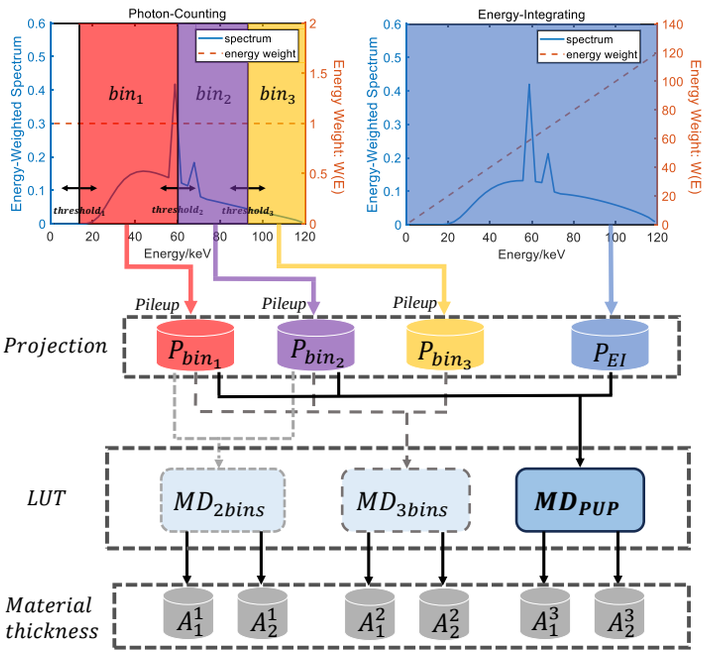


Fig. 2. An illustration of three MD approaches for the comparison study. MD_{2bins} uses photon-counting projections from two energy bins:$\{P_{bin_1}, P_{bin_2}\}$. MD_{3bins} uses photon-counting projections from three energy bins:$\{P_{bin_1}, P_{bin_2}, P_{bin_3}\}$. And MD_{PUP} uses photon-counting projections from two energy bins and energy-integrating projection acquired simultaneously:$\{P_{bin_1}, P_{bin_2}, P_{EI}\}$.

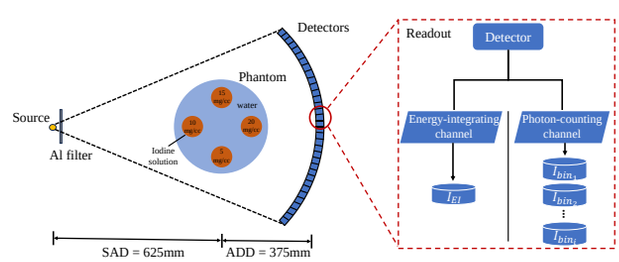


Fig. 3. Spectral imaging configurations for the comparison study. Both energy-integrating and photon-counting data are simultaneously acquired.

$$MD_{PUP}: \begin{bmatrix} A_1^3 \\ A_2^3 \end{bmatrix} = F_3 \begin{bmatrix} P_{\mathrm{bin}_1} \\ P_{\mathrm{bin}_2} \\ P_{EI} \end{bmatrix}. \quad (7)$$

The other two MDs for comparison can also be expressed,

$$\begin{aligned} MD_{2bins}: &\begin{bmatrix} A_1^1 \\ A_2^1 \end{bmatrix} = F_1 \begin{bmatrix} P_{\mathrm{bin}_1} \\ P_{\mathrm{bin}_2} \end{bmatrix}. \\ MD_{3bins}: &\begin{bmatrix} A_1^2 \\ A_2^2 \end{bmatrix} = F_2 \begin{bmatrix} P_{\mathrm{bin}_1} \\ P_{\mathrm{bin}_2} \\ P_{\mathrm{bin}_3} \end{bmatrix}. \end{aligned} \quad (8)$$

The lookup table serves as the MD matrix $\boldsymbol{F}$, employing various iodine-water combinations of distinct thicknesses for

calibration. The maximum thickness for water was capped at at 50 cm with 51 calibration points, while iodine's maximum thickness was set at 0.04 cm with 61 calibration points. In addition, we provide projection datasets of different initial flux intensities for the lookup table corresponding to each MD strategy, that is, the lookup table is flux-independent.

To evaluate the performance of these three MD approaches, they were applied to process data obtained under identical scanning configurations. A spectral CT system was simulated according to the configurations described in Fig.3. The X-ray source operated at 120 kV, followed by 4 mm aluminum filter. The detectors were set for a perfect energy response, with pixel dimensions of 1 x 1 mm². Three energy bins were used, with thresholds set at 20/60/90 keV. The scanned object is a cylindrical water phantom containing iodine solutions of different concentrations of 20 mg/cc, 15 mg/cc, 10 mg/cc, and 5mg/cc. Throughout the scanning process, both energy integration and photon counts data were simultaneously acquired.

III. RESULTS

A. Numerical Calculations of Projection Distribution

It's known that higher counts rate exacerbate the pileup effects, necessitating consideration of the impact of flux intensity when pileup effects occurs. As shown in the Fig.4.(A), the photon-counting projections:$\{P_{bin_1}, P_{bin_2}\}$ from two thresholds result in overlapping projection points of material combinations with different thicknesses as the flux intensity changes. This implies that MD is prone to errors, particularly in diagnostic CT applications with continuous tube current modulation.

To address this, we employ energy integration, which is unaffected by pulse pileup effects and performs well at high counts rate, as third-dimensional information to enhance MD. As calculation results illustrated in the Fig.4.(B), energy integration accurately identifies material thickness under varying flux intensities. Projection points at different flux intensities no longer interfere with each other. The combined use of photon counts and energy integration makes it possible to achieve precise MD. The Fig.4.(C) demonstrate that replacing energy integration with counts from the third threshold causes nearly all projection points to remain distributed on the same plane, resembling the 2D projection distribution. This may introduce errors in MD results, and the ill-conditioned nature of the mapping relationship renders the inverse solution process numerically unstable and susceptible to noise.

B. The Comparison Study

The detector model was configured as a paralysis type and scanned at a counts rate of $a\tau = 0.5$. Following the methodology outlined in Section II.B, three MD lookup tables were applied to process the acquired data. The datasets corresponding to the lookup tables for each MD are measured at the same mA with $a\tau = 0.05 \sim 1.0$. The water images, iodine images, and virtual mono-energetic images (VMIs) along with their corresponding quantitative results were displayed in Fig.5.

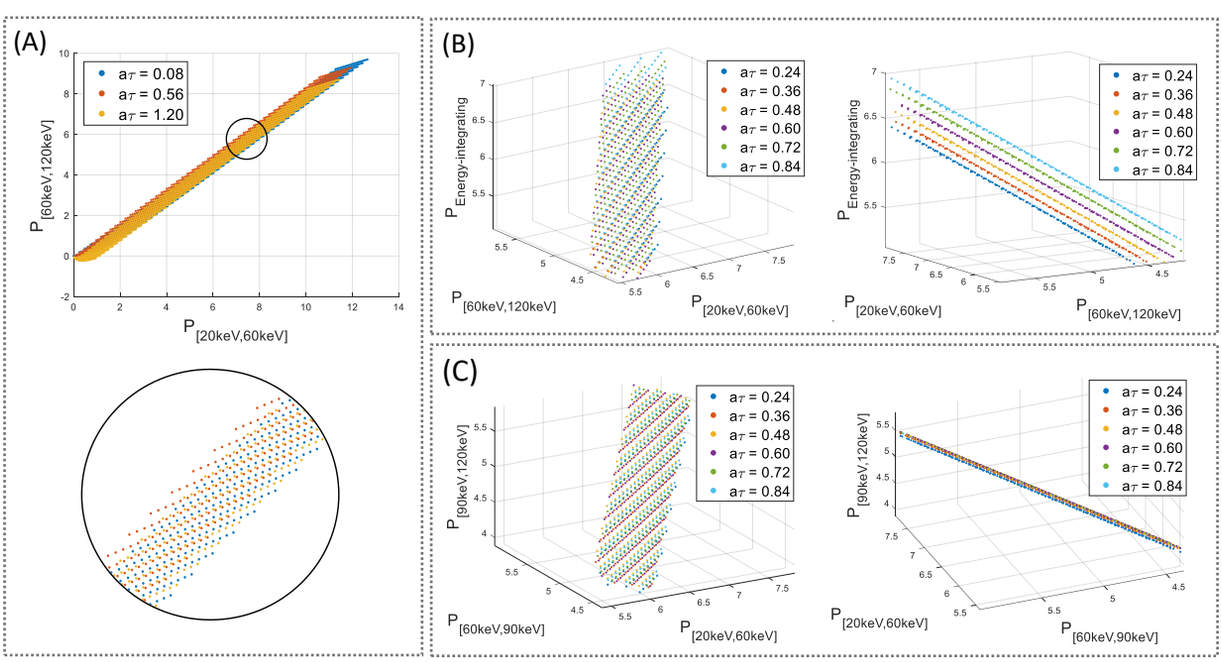


Fig. 4. Pileup-induced projection distribution for various thickness combinations of water and iodine. (A) 2D distribution:$\{P_{bin_1}, P_{bin_2}\}$, using two energy bins with $a\tau = 0.08, 0.56, 1.20$. Projection points of different thicknesses combinations blend together as the flux intensity changes. (B) Two views of 3D projection distribution:$\{P_{bin_1}, P_{bin_2}, P_{EI}\}$, using two energy bins and employing energy integration with $a\tau = 0.24 \sim 0.84$. Projection points at different flux intensities are separated into different planes, and no longer interfere with each other. (C) Two views of 3D projection distribution:$\{P_{bin_1}, P_{bin_2}, P_{bin_3}\}$, using three energy thresholds with $a\tau = 0.24 \sim 0.84$. All projection points almost remain distribute on the same plane, akin to the situation in figure (A).

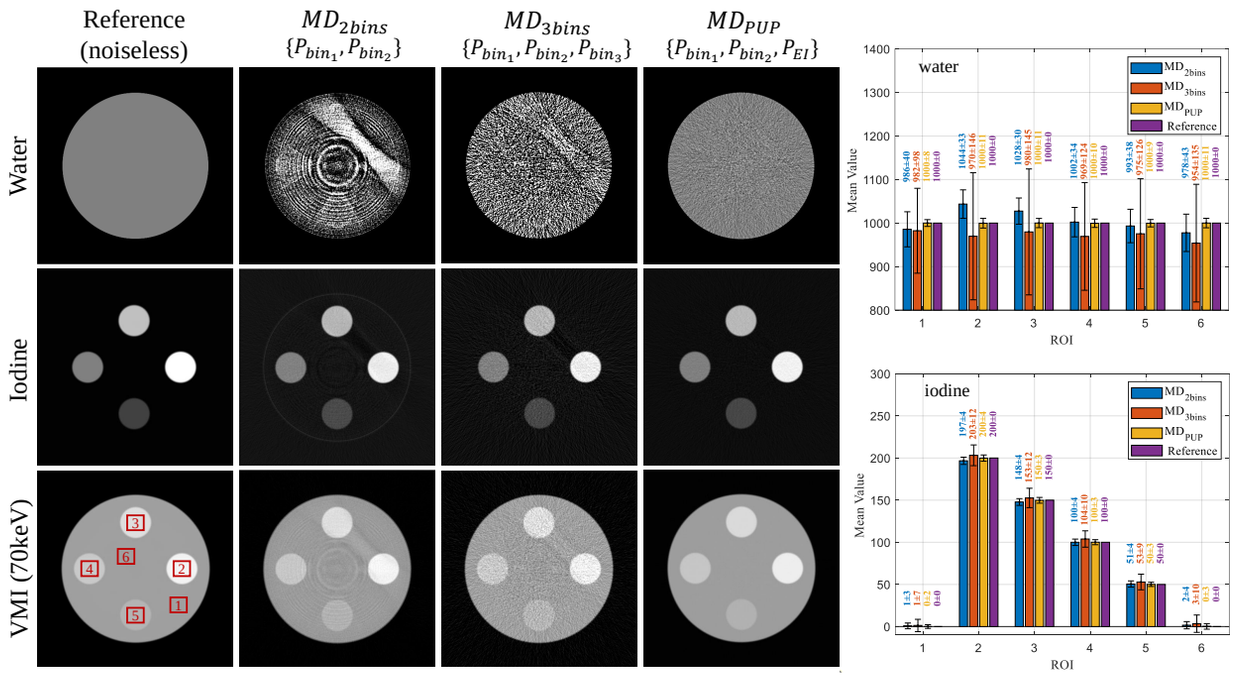


Fig. 5. Spectral images and quantitative results obtained through three MD approaches. Both images and quantitative results reveal that employing the pileup-proof MD shows better decomposition accuracy and a lower level of noise compared to the other two. First row: water images; second rows: iodine images; and third rows: VMIs at 70 keV. Display windows for the water, iodine and VM images are [950, 1050] mg/cc, [-1.5, 22] mg/cc and [0, 1600] HU, respectively. The histograms on the right are mean and standard deviation of the selected ROIs. Unit: mg/cc for water and 0.1mg/cc for iodine.

The comparison results reveal noticeable artifacts and higher levels of noise in the basis images obtained using the MD_{2bins} and MD_{3bins}. The noise in images generated by MD_{3bins} is particularly pronounced, consistent with the projection distribution analysis. In contrast, it is evident that the reconstructed images using the pileup-proof MD_{PUP} approach show better decomposition accuracy and less noise. Six regions of interest (ROIs) were selected as depicted in Fig.5. The mean values and standard deviation of ROIs were presented in bar charts on the right. Taking the water images as an example, for the six ROIs, the maximum relative errors of MD_{2bins} and MD_{3bins} are 4.36% and 4.60% respectively, while MD_{PUP} demonstrated a low maximum error of 0.04% and minimal noise.

IV. CONCLUSIONS

In conclusion, this study proposes a pileup-proof MD approach in projection domain using the HPIX for spectral CT. Theoretical calculations demonstrates that the incorporation of energy integration enables the differentiation of projection data with varying flux intensities, thereby resisting the impact of pulse pileup and enhancing the robustness of MD. In comparison to with traditional MD based on photon-counting data, the pileup-proof MD effectively resists the adverse impact of pulse pileup effects, resulting in exceptional decomposition accuracy and less noise.

Future research endeavors will focus on the validation of our findings using experimental data acquired through the HPIX-detector.

REFERENCES

[1] J. Cammin, S. Srivastava, Q. Tang, W. Barber, J. Iwanczyk, N. Hartsough, and K. Taguchi, "Compensation of nonlinear distortions in photon-counting spectral ct: deadtime loss, spectral response, and beam hardening effects," in *Medical Imaging 2012: Physics of Medical Imaging*, vol. 8313. SPIE, 2012, pp. 542–548.

[2] H. Ding and S. Molloi, "Image-based spectral distortion correction for photon-counting x-ray detectors," *Medical physics*, vol. 39, no. 4, pp. 1864–1876, 2012.

[3] R. Feng, D. Rundle, and G. Wang, "Neural-networks-based photon-counting data correction: pulse pileup effect," *arXiv preprint arXiv:1804.10980*, 2018.

[4] P. J. Jenkins and T. G. Schmidt, "Experimental study of photon-counting ct neural network material decomposition under conditions of pulse pileup," *Journal of Medical Imaging*, vol. 8, no. 1, pp. 013 502–013 502, 2021.

[5] J. D. Castiglioni, E. Y. Sidky, and T. G. Schmidt, "Material decomposition for photon-counting ct using a flux-independent neural network," in *Medical Imaging 2023: Physics of Medical Imaging*, vol. 12463. SPIE, 2023, pp. 902–906.

[6] K. Taguchi, M. Zhang, E. C. Frey, X. Wang, J. S. Iwanczyk, E. Nygard, N. E. Hartsough, B. M. Tsui, and W. C. Barber, "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.

[7] E. Roessl, H. Daerr, and R. Proksa, "A fourier approach to pulse pileup in photon-counting x-ray detectors," *Medical physics*, vol. 43, no. 3, pp. 1295–1298, 2016.

[8] K. Treb, J. Radtke, W. S. Culberson, and K. Li, "Simultaneous photon counting and charge integrating for pulse pileup correction in paralyzable photon counting detectors," *Physics in Medicine & Biology*, vol. 68, no. 15, p. 155003, 2023.

[9] X. Wang, Z. Deng, W. Liu, and Y. Liu, "Development of a pixel readout asic for czt detectors for spectral x-ray photon-counting imaging applications," *Journal of Instrumentation*, vol. 15, no. 01, p. C01030, 2020.

[10] J. Fink, "Characterization of the imaging performance of the simultaneously counting and integrating x-ray detector cix," Ph.D. dissertation, Universitäts-und Landesbibliothek Bonn, 2010.

Joint Material Decomposition and Scatter Estimation for Spectral CT

Altea Lorenzon, Stephen Z. Liu, Xiao Jiang, Grace J. Gang, J. Webster Stayman

***Abstract*—Accurate scatter correction is essential to obtain high-quality reconstructions in computed tomography. While many correction strategies for this longstanding issue have been developed, additional efforts may be required for spectral CT imaging - which is particularly sensitive to unmodeled biases. In this work we explore a joint estimation approach within a one-step model-based material decomposition framework to simultaneously estimate material densities and scatter profiles in spectral CT. The method is applied to simulated phantom data obtained using a parametric additive scatter mode, and compared to the unmodeled scatter scenario. In these preliminary experiments, We find that this joint estimation approach has the potential to significantly reduce artifacts associated with unmodeled scatter and to improve material density estimates.**

I. Introduction

X-RAY scatter can pose significant challenges to image quality in computed tomography (CT) and, particularly, cone-beam CT (CBCT), as the extensive amount of scatter generated by the large cone angle leads to loss of contrast, reduced contrast-to-noise ratio and the emergence of cupping and streak artifacts. The problem of scattered radiation has been addressed in several ways: for a comprehensive review of scatter correction methods refer to [1][2]. Hardware techniques for scatter rejection include anti-scatter grids and collimation approaches; whereas software-based scatter correction approaches typically involve the subtraction of a scatter estimate from the projection data - often achieved by experimental measurement or numerical estimation. Among the latter, common strategies utilize scatter-kernel superposition (SKS) and Monte Carlo (MC) simulations. SKS methods [3][4] involve the deconvolution of the projection data with kernels derived from scatter point-spread functions, obtained from MC simulations [5][6] or measurements [7]. MC simulations can provide very accurate scatter images, as they include complex stochastic processes avoiding the simplifications typical of other numerical methods. While SKS methods are known to provide good computational speed, MC simulations are in general more computationally demanding [8], although several acceleration techniques have been developed [9][10]. Both methods are subject to uncertainties with regard to modeling of the object, the imaging system, and/or other approximations (e.g. limited number of MC events, approximate kernel estimates, etc.).

While the problem of scattered radiation is well known with many proposed corrections methods, the challenges remain in the field of spectral CT [11], where basis material and density estimation can be more sensitive to unmodeled biases. In this work, we explore a joint estimation approach in the framework of one-step model-based material decomposition (MBMD), aiming to simultaneously estimate multiple quantities.

Altea Lorenzon, Stephen Z. Liu, Xiao Jiang, Grace J. Gang, and J. Webster Stayman are in Biomedical Engineering at Johns Hopkins University, Grace J. Gang is in the Department of Radiology at the University of Pennsylvania. Contact author e-mail: web.stayman@jhu.edu

In the context of spectral CT, the joint estimation method was proposed to directly estimate the basis material densities from spectral data [12][13], avoiding the complications that emerge from the separation between the reconstruction and the decomposition steps. We want to extend this approach to model and estimate an additional unknown component represented by the scattered radiation. Specifically, we focus on determining both the densities of the material basis and the scatter profile within the projection data simultaneously. This is achieved by incorporating the latter into the forward model for the spectral measurements and jointly estimating all parameters. Such simultaneous estimation is similar, in spirit, to related work by Ha et al.[14] where one-step material decomposition is conducted in conjunction with estimation of unknown spectral calibrations. In following sections, we detail the model used by our approach and the estimation strategy including different regularization schemes. The estimator is applied to simulated phantom data with different additive scatter profiles. Material density estimates for estimated and un-estimated scatter conditions are compared, illustrating the potential of the joint estimation technique to improve spectral CT imaging.

II. Methods

A. Forward Model

For the energy-dependent spectral measurements of an object made of N_m material basis we assume the following forward model:

$$\bar{y}_i = \int_E S_i(E) \exp\left[-\sum_{m=1}^{N_m} \mu^m(E) \sum_{j=1}^{N_v} A_{ij}\rho_j^m\right] + s_i \qquad (1)$$

where ρ_j^m is the density of the m-th material in the j-th voxel, with $1 \leq j \leq N_v$, being N_v the total number of voxels; A_{ij} represents the path length of the i-th projection through the j-th voxel; $\mu^m(E)$ denotes the energy-dependent mass attenuation coefficients and $S_i(E)$ indicates the energy-dependent system response, a general function that includes the model of the source spectra, filters, and the detector response. Finally, s_i is the term that accounts for scatter.

To facilitate digital estimation, it is convenient to introduce the following discretization in energy,

$$\bar{y}_i = \sum_{k=1}^{N_E} S_i(E_k)\Delta E \exp\left[-\sum_{m=1}^{N_m} \mu^m(E_k) \sum_{j=1}^{N_v} A_{ij}\rho_j^m\right] + s_i \qquad (2)$$

where N_E is the number of bins of the energy spectra sampling and ΔE the width of the energy bins: $E_k = k\Delta E$, $k = 1, ..., N_E$.

The forward model (2) can be represented compactly using matrix notation:

$$\bar{\mathbf{y}} = \mathbf{S} \exp\left[-\mathbf{Q}\mathbf{A}\boldsymbol{\rho}\right] + \mathbf{s} \qquad (3)$$

where $\bar{\mathbf{y}}$ is the N-dimensional vectorized stack of all projections and all the spectral channels, $\boldsymbol{\rho}$ is the $N_m N_v$ vectorized stack of all the density images with N_v voxels in N_m material bases, $\mathbf{A}$ the $NN_m \times N_m N_v$ system matrix and $\mathbf{Q}$ is the $NN_E \times NN_m$ energy-dependent mass attenuation coefficient matrix. Here, $\mathbf{S}$ represents the $N \times NN_E$ matrix with spectral sensitivities. Finally, $\mathbf{s}$ is the vectorized stack of the scatter contribution for each measurement. In this work, two models have been considered for $\mathbf{s}$, illustrated in Section II-C.

B. Model Based Material Decomposition

Assuming the spectral measurements follow a multivariate Gaussian distribution with mean given by Equation (3) and covariance $\boldsymbol{\Sigma}$

$$y_i \sim N(\bar{y}_i, \boldsymbol{\Sigma})$$

the corresponding penalized weighted least-square objective function can be written as:

$$\psi = \frac{1}{2}\left(\mathbf{y} - \mathbf{S}e^{-\mathbf{QA}\boldsymbol{\rho}} - \mathbf{s}\right)^T \mathbf{W}\left(\mathbf{y} - \mathbf{S}e^{-\mathbf{QA}\boldsymbol{\rho}} - \mathbf{s}\right) + R(\boldsymbol{\rho}) \tag{4}$$

Here, $\mathbf{W} = \boldsymbol{\Sigma}^{-1}$. We further presume uncorrelated measurement noise, so the covariance matrix is diagonal, and we approximate the variance with the measurements. Hence $\mathbf{W} = \mathrm{D}\{1/\mathbf{y}\}$, with $\mathrm{D}\{\cdot\}$ representing a diagonal matrix. The term $R(\boldsymbol{\rho})$ is a regularizer controlling the smoothness of the estimated images, independently for each material basis. In this work, we utilize both a penalty function based on the L1 norm of the difference in density of neighboring pixels, or:

$$R(\boldsymbol{\rho}) = \frac{1}{4}\sum_{m=1}^{M_m}\beta_m \sum_{j=1}^{N_v}\sum_{j'\in\mathcal{N}_j}|\rho_j^m - \rho_{j'}^m| \tag{5}$$

and a penalty function based on L2 norm:

$$R(\boldsymbol{\rho}) = \frac{1}{4}\sum_{m=1}^{M_m}\beta_m \sum_{j=1}^{N_v}\sum_{j'\in\mathcal{N}_j}\left(\rho_j^m - \rho_{j'}^m\right)^2 \tag{6}$$

where β_m is the regularization strength for the m-th material basis and $\mathcal{N}_j$ identifies the set of four pixels in the neighborhood of pixel j.

The minimization of the objective function (4) yields the material decomposition estimates $\hat{\boldsymbol{\rho}}$ and it is achieved iteratively with a stochastic gradient descent algorithm.

C. Simulation Study

The 2D digital phantom presented in Figure 1 was used to test the joint estimation of scatter and material densities with the model illustrated in Section II-A. This phantom was created with $0.5 \times 0.5\,\mathrm{mm}^2$ voxels. Mass attenuation coefficients shown in Figure 2a were obtained from Spektr. We simulated an imaging system including a 1024 pixel triple-layer indirect flat-panel detector (FPD) with a 0.3 mm pitch, assuming a source-to-axis distance of 80 cm and a source-to-detector distance of 120 mm. CsI scintillation layers were 0.20, 0.55 and 0.55 mm thick, for the top, middle, and bottom layers, respectively. A 120 kVp X-ray spectrum was generated with Spektr, including a 10 mm-thick aluminum filter. The bare-beam intensity is set to $4 \cdot 10^5$ photons/pixel. The spectral response of the detector is shown in Figure 2b. The imaging geometry used 600 projections with a 0.6 degree sampling, and Poisson noise was added to the data.

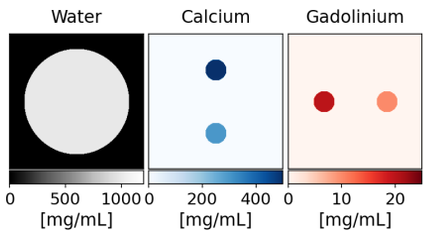
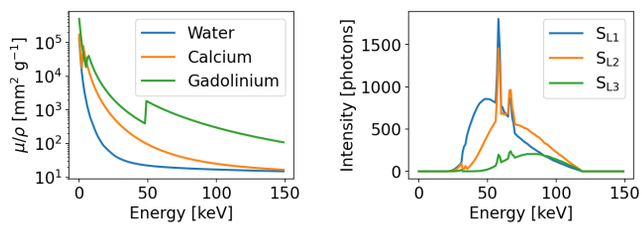


Fig. 1. Digital phantom with material bases: water, Calcium and Gadolinium.



(a) (b)

Fig. 2. (a) Mass attenuation coefficients for water, Calcium and Gadolinium. (b) Spectral response of the simulated triple-layer FPD imaging system.

Two scenarios were considered for scatter simulation:

(A) *Idealized Quadratic Case*, where the profile s is given by quadratic functions of the lateral detector position u:

$$s_l(u) = -\frac{2}{r_l N_u}u^2 + \frac{2}{r_l}u + m_l - \frac{N_u}{2r_l} \;\forall l = LT, LM, LB \tag{7}$$

where N_u is the total number of pixels, and the labels LB, LM, LT refer to the top, middle and bottom layer of the detector, respectively. The functions are parameterized in terms of m_l and r_l, controlling the maximum and the curvature of the profile, respectively. In particular, the parameters m_l are chosen such that the maximum of the air normalized scatter is 10%.

(B) *Monte Carlo Case*, where the profile s is generated for each energy channel with a fast MC simulation approach previously developed [15] and studied for dual energy applications [16]. The air normalized scatter profiles are shown in Figure 3.

Given the symmetry of the simulated object, in both cases the scatter profile is assumed to be the same in each angular projection of the same energy channel.

In the Quadratic case, data were fitted with the forward model of Equation (3) which incorporates a quadratic scatter profile

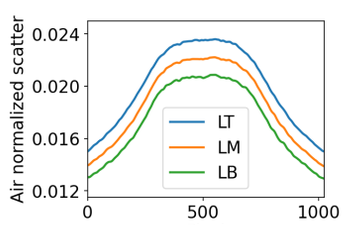


Fig. 3. Air normalized scatter profile obtained by fast MC simulation of a water phantom, in each layer of the simulated imaging system.

expressed in Equation (7), so that the estimation model matches the data. For comparison, MBMD was performed also with a forward model that does not include the scatter term. In the Monte Carlo case, the material densities and scatter are also estimated assuming a quadratic profile for the scatter term, as in Equation (7). It's important to note that in this instance, the assumption of a quadratic profile represents a simplification, considering that the scatter exhibits a more complex shape.

The regularization strength parameters β_m were defined as follows:

- $\beta_w = 10$, $\beta_{Ca} = 10$, $\beta_{Gd} = 200$ for water, Calcium and Gadolinium, respectively, when the L1 penalty of Equation (5) is used;
- $\beta_w = 0.01$, $\beta_{Ca} = 0.01$, $\beta_{Gd} = 1$ for water, Calcium and Gadolinium, respectively, when the L2 penalty of Equation (6) is used.

In each decomposition, the water density image is initialized to its ground truth value, while all the other parameters are initialized to zero; 20000 iterations are used to achieve a good convergence.

III. Results

A. Quadratic Scatter Profile

Figure 4 summarizes the results of the MBMD of data simulated in the hypothesis of a quadratic scatter profile. These results were obtained using a forward model that does not include the scatter component. Specifically, the estimator is formulated as Equation (3) but without the term $\mathbf{s}$. As can be seen from the figure, the reconstructions are dominated by artifacts that can be traced back to the un-modeled scatter component. Figure 5 shows that better results can be achieved in the three-material decomposition images when the scatter is included in the forward model: these results were obtained with the forward model of Equation (3) where the scatter has a quadratic profile as in Equation (7). Here the unregularized case appears noisy with some streaking artifacts. We observe that incorporating regularization with the L1 penalty enhances image quality compared to the unregularized scenario. However, there is a minor bias observed in the water and Gd density images corresponding to positions of other materials. Conversely, the reconstruction with the L2 penalty shows limited improvement compared to the unregularized case (though a more comprehensive parameter sweep may be used to fine-tune this reconstruction).

In Figure 6 we show the scatter profiles that were jointly estimated with the material densities. We observe that the scatter estimates do not match the true profiles exactly and in most cases the maximum in underestimated. This likely leads to the residual artifact in Figure 5. However the estimated scatter profiles reproduce the scatter level in the detector region outside the phantom shadow well enough to allow for a satisfactory material decomposition.

We have observed some variability in these scatter estimation results. Consider the results shown in Figure 7. In constrast to the previous case, here the scatter profile parameters m_l were not initialized with zeros but with an arbitrary value, higher than the ground truth. The scatter profiles are now in most cases overestimated. Whether this phenomenon is due to the initialization, the stochastic nature of the optimization process, or is inherent to the estimation problem is the subject of ongoing studies.

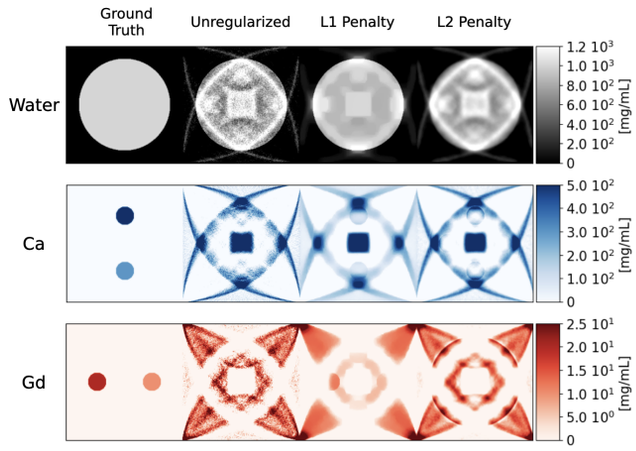


Fig. 4. From top to bottom, the reconstructions of water, Calcium and Gadolinium concentrations from projection data incorporating a quadratic scatter profile, obtained with a forward model that neglects scatter, are directly compared with the phantom concentrations (on the leftmost column). Then, from left to right, the reconstructions are obtained without any regularization function in the objective function, with a penalty based on L1 norm (5) and with a penalty function based on L2 norm (6).

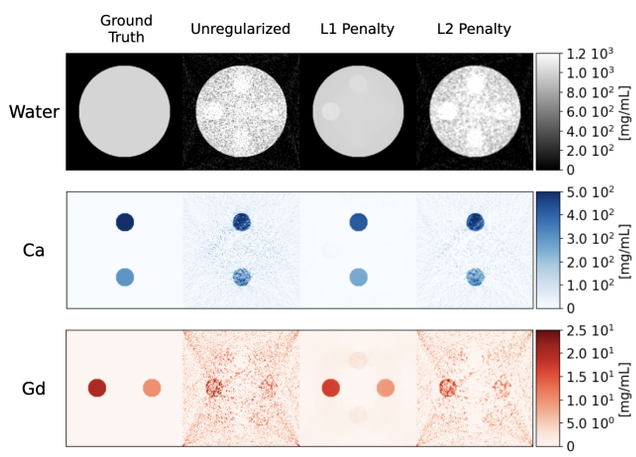


Fig. 5. From top to bottom, the reconstructions of water, Calcium and Gadolinium concentrations from projection data incorporating a quadratic scatter profile, obtained with the forward model (3) that includes scatter with a quadratic profile (7), are directly compared with the phantom concentrations (on the leftmost column). Then, from left to right, the reconstructions are obtained without any regularization function in the objective function, with a penalty based on L1 norm (5) and with a penalty function based on L2 norm (6).

B. MC Scatter Profile

The projection data with the addition of the Monte Carlo scatter profile were jointly processed with a quadratic scatter profile estimation model. The results of the MBMD are shown in Figure 8. As in the results discussed in the previous Section, the utilization of the L1 penalty brings clear advantages in the noise reduction over the unregularized case, while further fine-tuning may be required for the L2 penalty regularization strength. Unlike the other cases, where the density images appears substantially biased, the reconstruction with L1 penalty presents only a slight underestimation of the water and Calcium concentrations. The Gadolinium concentration is instead slightly overestimated, but in the latter image the shape of the water and Calcium inserts are also visible.

Figure 9 shows the scatter profile that is jointly estimated

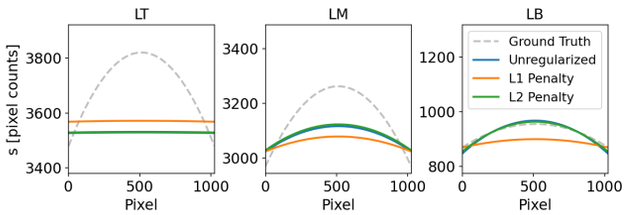


Fig. 6. From left to right, estimates of the scatter quadratic profile in layer LT, LM and LB. The profile maximum is initialized to zero.

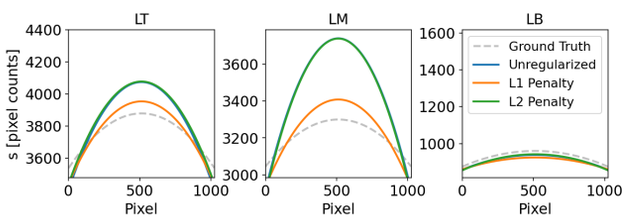


Fig. 7. From left to right, estimates of the scatter quadratic profile in layer LT, LM and LB. The profile maximum is initialized to an arbitrary higher value.

with material densities for each energy channel. While there is an evident data-model mismatch, the reconstruction of the profile appears to be more accurate in the air projection region, i.e. towards the extremities of the detector, as observed also in Figure 6 of the previous Section.

IV. Conclusion

In this work, we proposed a model-based method to jointly estimates material densities and scatter in spectral CT data. We simulated spectral CBCT data of a three-material basis phantom assuming a triple layer FPD imaging system. Specifically, two scenarios were explored for the simulation of scatter. In the first case, a simplified model assuming a quadratic function for the scatter profile was adopted. The material density estimates were compared when obtained from a forward model that neglects scatter and from our proposed forward model, showing the advantages of jointly estimating the scatter from the projection data. Additionally, we compared reconstructions using an objective function with various regularization strategies, including no regularization, L1 norm, and L2 norm penalty functions. In the second case, the scatter profiles of each spectral channel are obtained with a fast scatter MC simulation. The material density estimates are obtained using a forward model incorporating a quadratic scatter profile.

Preliminary findings reveal the need for fine-tuning some geometry parameters (e.g., FOV) and estimator parameters (e.g., L2 regularization strengths). We observed an instability in the scatter estimation with different initialization values, prompting further investigation into whether this is initialization-related or stems from the stochastic nature of the optimization algorithm. However, these preliminary results suggest that the joint estimation technique has the potential to significantly enhance density estimates in spectral CT reconstructions.

Acknowledgments

This work is supported, in part, by NIH grant R01EB030494.

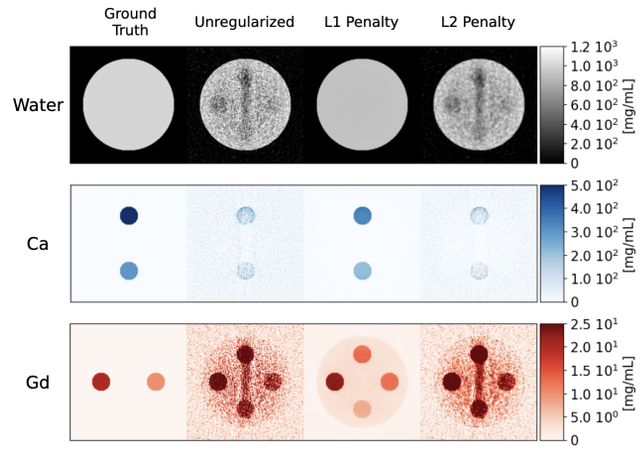


Fig. 8. From top to bottom, the reconstructions of water, Calcium and Gadolinium concentrations from projection data incorporating a MC simulated scatter profile, obtained with the forward model (3) that includes scatter with the quadratic profile (7), are directly compared with the phantom concentrations (on the leftmost column). Then, from left to right, the reconstructions are obtained without any regularization function in the objective function, with a penalty based on L1 norm (5) and with a penalty function based on L2 norm (6).

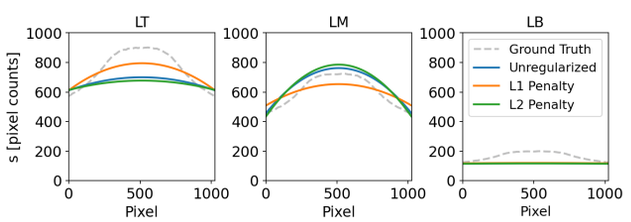


Fig. 9. From left to right, the MC simulated scatter profile of is estimated as a quadratic function (7) in layer LT, LM and LB.

References

[1] Rührnschopf, E.-P and Klingenbeck, K, *Medical physics*, vol. 38, no. 7, pp. 4296–4311, 2011.
[2] ——, *Medical physics*, vol. 38, no. 9, pp. 5186–5199, 2011.
[3] Love, L. A and Kruger, R. A, *Medical physics*, vol. 14, no. 2, pp. 178–185, 1987.
[4] Seibert, J. A and Boone, J. M, *Medical physics*, vol. 15, no. 4, pp. 567–575, 1988.
[5] Maltz, J. S *et al.*, *IEEE transactions on medical imaging*, vol. 27, no. 12, pp. 1791–1810, 2008.
[6] Sun, M and Star-Lack, J, *Physics in Medicine & Biology*, vol. 55, no. 22, p. 6695, 2010.
[7] Rinkel, J *et al.*, *Physics in Medicine & Biology*, vol. 52, no. 15, p. 4633, 2007.
[8] Jarry, G *et al.*, in *Medical Imaging 2006: Physics of Medical Imaging*, vol. 6142. SPIE, 2006, pp. 1634–1643.
[9] Poludniowski, G *et al.*, *Physics in Medicine & Biology*, vol. 54, no. 12, p. 3847, 2009.
[10] Sisniega, A *et al.*, *Medical physics*, vol. 40, no. 5, p. 051915, 2013.
[11] Pivot, O *et al.*, *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 2267–2276, 2020.
[12] Schmidt, T. G *et al.*, *IEEE transactions on medical imaging*, vol. 36, no. 9, pp. 1808–1819, 2017.
[13] Tilley, S *et al.*, *Physics in Medicine & Biology*, vol. 64, no. 3, p. 035005, 2019.
[14] Ha, W *et al.*, in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*, 2018, pp. 1–5.
[15] Sisniega, A *et al.*, *Physics in Medicine & Biology*, vol. 60, no. 4, p. 1415, 2015.
[16] Liu, S. Z *et al.*, in *7th International Conference on Image Formation in X-Ray Computed Tomography*, vol. 12304. SPIE, 2022, pp. 209–215.

Spectral Orbits: Combining Spectral Imaging and Non-Circular Orbits for Interventional CBCT

Grace J. Gang[1]

***Abstract*—Cone-beam CT imaging using non-circular orbits has been demonstrated to be effective in reducing artifacts around metal. With the increasing interest in spectral imaging in the interventional suite, there are potential advantages to combine both technologies to yield further image quality benefits. We simulated a neuro-interventional application where imaging around the embolization is challenged by metal artifacts and the differentiation of bleeds and contrast extravasation is difficult with single-energy imaging. The imaging system was simulated with a dual-layer detector and different sinusoidal orbits. Material decomposition used a projection-domain approach followed by a model-based reconstruction of the density line integrals of each basis. The spectral non-circular orbits acquisitions were compared with single-energy circular, single-energy non-circular, and spectral circular orbits. Results using spectral non-circular orbit contain minimal metal artifacts and allow the differentiation of bleeds and contrast extravasation, demonstrating the potential of the combined technologies.**

I. Introduction

Metal artifacts are a persistent problem in CT and cone-beam CT (CBCT), especially in interventional imaging scenarios where implants and surgical tools are frequently present in the field of view. Many previous studies have demonstrated the benefit of non-circular orbits in reducing artifacts around metal objects.[1]–[4] Various orbits using different design methodologies have been proposed. Some are patient-specific, leveraging prior information of the metal and/or the patient anatomy; some intend to be general purpose, relying on the intrinsic sampling patterns of the orbit itself. We have previously demonstrated a design framework that treats projection data with metal as missing data, and designed orbits that maximize the sampling completeness for arbitrary metal shapes and locations. We found that a simple class of sinusoidal orbits are fairly robust against metal artifacts.

Spectral imaging is being increasingly explored for interventional applications. Advantages of spectral imaging include beam hardening corrections, reduced metal artifacts, the ability to quantify (iodine) concentrations[5], virtual non-contrast imaging[6], and advanced applications like thermometry[7].

In this work, we investigate the imaging capability of non-circular orbit combined with spectral imaging. We focus on an example application to maximize visibility of bleeds and/or contrast extravasation following an embolization procedure. Both types of signals present as hyperdense foci and are therefore challenging to distinguish using single energy imaging. While diagnostic spectral CT has been shown to be effective in distinguishing the two in multiple studies,[8], [9] moving the patient from the interventional suite to a diagnostic spectral CT system can be sub-optimal in terms of workflow. Furthermore, visualization of bleed or contrast extravasation close to the embolization coil is still challenged by metal artifacts. We thus hypothesize that combined CBCT imaging using spectral and non-circular orbits technology can improve imaging quality around metal, helping to distinguish bleeds and contrast extravasation, while enabling post-operative assessment at the point of care in the operating room.

In this paper, we present the underlying methodology for model-based material decomposition using spectral CBCT and illustrate its application in an emulated intracranial embolization. Circular and noncircular orbits are compared including differences among different sinusoidal orbits. The improved visualization and density estimates of noncircular orbits is quantified illustrating the potential of this technique to improve intraoperative assessments.

[1] University of Pennsylvania, Philadelphia, PA, USA

II. Methods

A. Forward Model

A general forward model for mean measurements, $\overline{\mathbf{y}}$, in spectral CT/CBCT is given by:

$$\overline{\mathbf{y}} = \mathbf{BS}\exp(-\mathbf{QAx}) \tag{1}$$

where $\mathbf{x}$ are the physical densities for material basis volumes, $\mathbf{Q}$ are mass attenuation coefficients for each basis, and $\mathbf{A}$ is the forward projector. The channel-specific spectral response and system blur is given by $\mathbf{S}$ and $\mathbf{B}$, respectively. Non-circular orbit acquisitions are accounted for through the forward projector $\mathbf{A}$ using the appropriate geometry.

B. Spectral Imaging System

We simulated an imaging system capable of both spectral imaging and non-circular orbit acquisition. Practically, such a system can consist of a robotic c-arm with a spectral detector, or an x-ray source capable of kV-switching, or a pre-patient split-filter, or combinations of the above. In this work, we focus on a system with a dual-layer flat-panel detector. The simulated detector configuration follows that of a prototype from Varex Imaging Corporation, which consists of a top layer of 200 μm CsI scintillator, a 1 mm Cu interstitial filter, and a bottom layer of 500 μm CsI scintillator.[10] The imaging technique was set to 100 kV and 0.5 mAs per projection. The overall spectral response

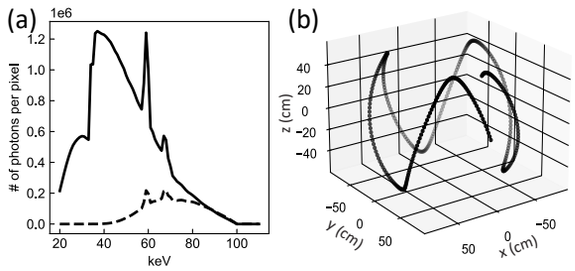


Fig. 1: ((a) The overall spectral response of the two energy channels. (b) An example sinusoidal orbit ($f = 3.3$, $\phi_{max} = 40°$) shown as source positions in the world coordinate.

of the two energy channels are plotted in Fig.1(a). Geometry mismatch (magnification and pixel grid misalignment) between the two layers were not simulated. Detector blur was not modeled, e.g. $\mathbf{B} = \mathbf{I}$.

C. Orbit Design

We have demonstrated in previous work that a simple class of sinusoidal orbits are robust against metal artifacts in single-energy imaging. The relative performance of different sinusoids are driven by sampling completeness, with orbits with minimal redundancy in projection views (e.g., even frequencies, non-integer frequencies) outperforming the rest (e.g., odd frequencies).

In this work, we investigate the performance of sinusoidal orbits with spectral imaging. Similar to previous work, we simulated sinusoidal orbits where the elevation angle ϕ as a function of the azimuthal angle θ is given by:

$$\phi = \phi_{max} \sin(f\theta) \tag{2}$$

The angles, θ, cover 360 projections over 360° at an angular increment of 1°. We simulated sinusoids of frequencies ranging from 1.0 to 3.3 and the maximum tilt angle, ϕ_{max}, ranging from 10° to 50°. The source-to-detector and source-to-axis distances were set to 112.6 and 82.8 cm, respectively. The source positions of the orbits are illustrated in Fig.1(b).

D. Material Decomposition Algorithm

Considerations in choosing the type of material decomposition algorithm are similar in acquisitions using circular and non-circular orbits. One-step model-based material decomposition algorithms[11], [12] are versatile and can accommodate all spectral systems. Alternatively, image domain decomposition can be performed using reconstructions of data from each spectral channel. A reconstruction algorithm, either model-based or analytic, that can properly reconstruct data from non-circular geometry should be adopted. If projection data from each spectral channel have consistent geometry, i.e., in the case of spectral detectors,[1] one can also perform projection domain decomposition and reconstruct the resulting density line integrals for each basis to obtain its respective density volumes.

There is a choice of the material types and the number of basis images used in material decomposition. For the imaging task investigated in this work, the exact density of the metal object is not important and often impossible to obtain accurately due to photon starvation. It is therefore sufficient to allow the metal to be projected onto other the basis materials. For visualization of bleeds and contrast, calcium is typically not a material of interest. Furthermore, three material decomposition is significantly noisier than two material decomposition. We therefore chose to perform a two material decomposition into water and iodine bases.

Based on the above considerations, we performed a projection domain material decomposition into water and iodine density line integrals, which were then reconstructed separately using a penalized weighted least-squares (PWLS) algorithm. A lookup table approach was adopted for fast computation. Denoting density line integrals as $\mathbf{l}$ ($= \mathbf{Ax}$), the density line integrals can be obtained by mapping the relationship between l and y using a lookup table $\mathcal{L}$, i.e.:

$$\{\hat{\mathbf{l}}^{H_2O}, \hat{\mathbf{l}}^{I}\} = \mathcal{L}_{H_2O,I}(\mathbf{y}_1, \mathbf{y}_2) \tag{3}$$

where subscripts $_1$ and $_2$ denotes the two spectral channels. The PWLS algorithm for $\hat{\mathbf{l}}$ corresponding to a single basis can be expressed as:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \Phi(\mathbf{x}; \hat{\mathbf{l}}) \quad \text{where}$$
$$\Phi(\mathbf{x}; \hat{\mathbf{l}}) = \frac{1}{2}(\hat{\mathbf{l}} - \mathbf{l})^T \mathbf{K}_{\mathbf{l}}^{-1} (\hat{\mathbf{l}} - \mathbf{l}) + R(\mathbf{l}) \tag{4}$$

Here, the inverse covariance of $\hat{\mathbf{l}}$, $\mathbf{K}_{\mathbf{l}}^{-1}$, is approximated by the measurements $\mathbf{y_1}$. We adopted a simple quadratic regularization that penalizes voxel differences in the first-order neighborhood in 3D.

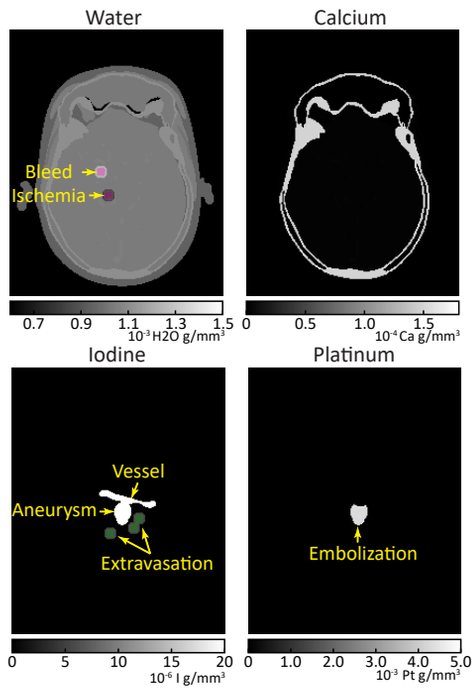


Fig. 2: Ground truth basis material density maps consisting of water, calcium, iodine, and a platinum embolization coil. 3D volumes of interests indicated by color masks were used to evaluate image quality in terms of the RMSE in the water and iodine basis.

E. Phantom

We simulated a head phantom as shown in Fig.2. Soft tissue and bone density maps were obtained using an image domain decomposition of XCAT head phantoms simulated at 60 and 100 kV. A vessel with an aneurysm was inserted and filled with iodine at 75 mg/mL. An embolization coil made of platinum was simulated to fill the aneurysm. Several imaging targets were placed around the coil to evaluate imaging performance around metal objects, including: 1) a region simulating bleed as dense water at a contrast of $+2 \times 10^{-5}$ g/mm^{-3}, 2) a region simulating ischemia as hypodense water at a contrast of -2×10^{-5} g/mm^{-3}, 3) three regions of extravasation of iodine at a contrast of 5×10^{-6} g/mm^{-3}. All targets are circular with a diameter

[1] Interpolation can be performed for multi-layer detectors to account for slight mismatch in magnification and pixel grid.

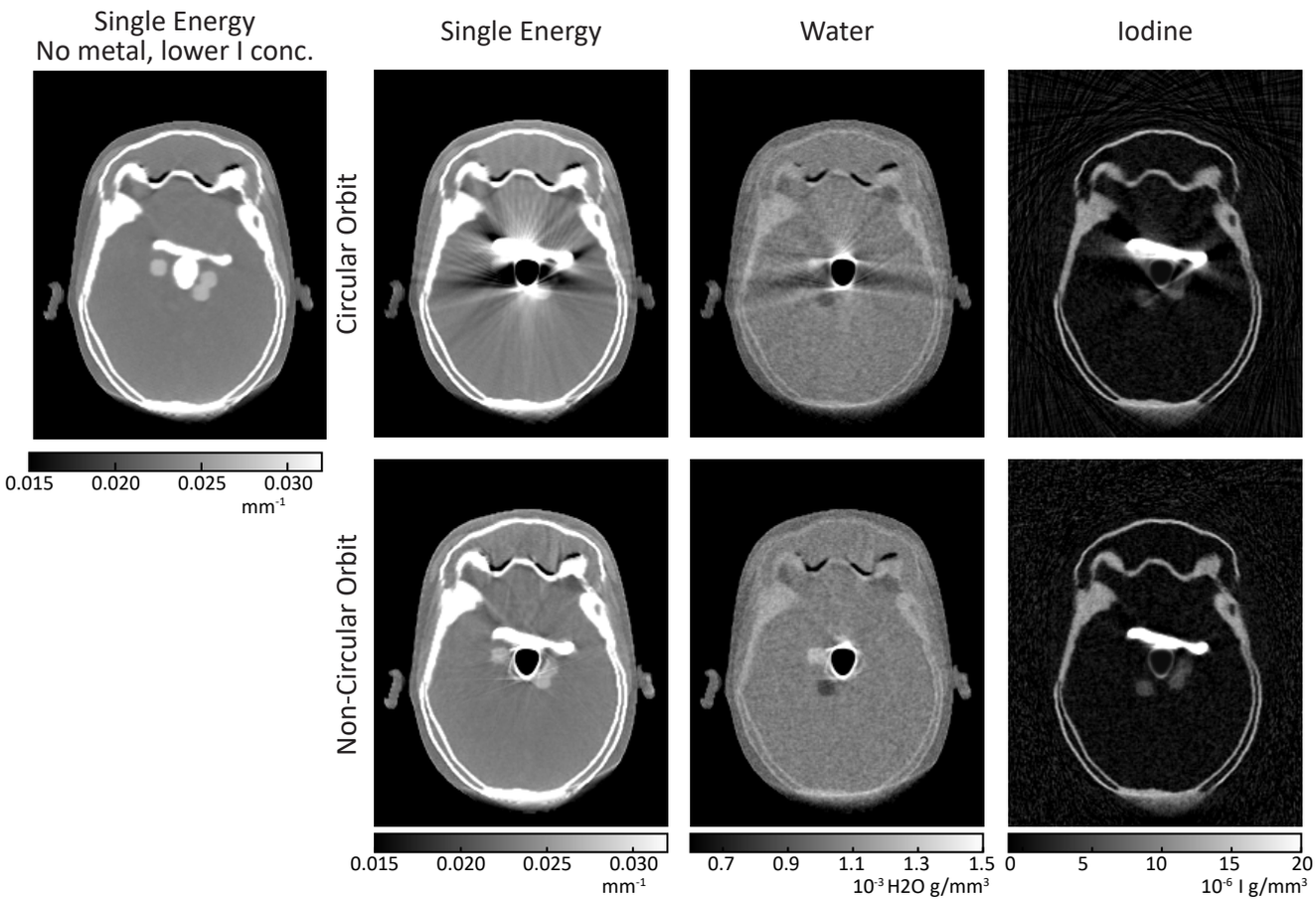


Fig. 3: Comparison of reconstruction results from circular and non-circular orbits using single energy and spectral measurements.

of ~ 1 cm. Contrast was specified relative to the soft tissue background the targets are superimposed over. The ischemic region and one of the extravasation targets completely overlap, resulting in a net zero contrast to the background.

F. Evaluations

The performance of different sinusoidal orbits were evaluated by visual inspection of the imaging targets in the water and iodine bases. We further quantified the root mean square error (RMSE) between the reconstructed and the ground truth densities. Volumes of interest (VOI) used for this purpose reside in the interior of the imaging targets and are indicated in Fig.2 by color masks. Spectral non-circular orbit performance was compared with single-energy circular, single-energy non-circular, and spectral circular acquisitions.

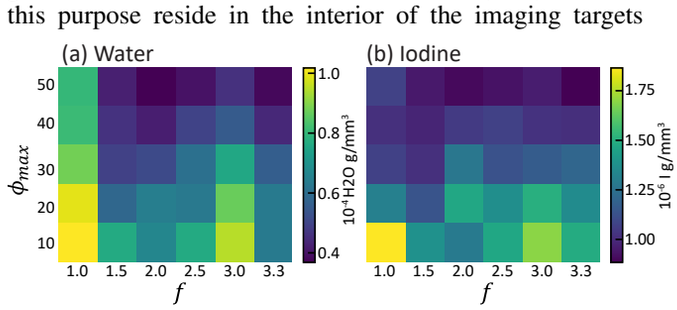


Fig. 4: The RMSE betweem the spectral results and ground truth within water and iodine imaging targets as a function of the maximum tilt (ϕ_{max}) and frequency (f) of the sinusoidal orbits.

III. Results

Comparisons amongst different imaging strategies are shown in an axial slice in Fig.3. The spectral orbit reconstructions uses $f = 3.3$ and $\phi_{max} = 40°$. A single energy reconstruction without metal and with a lower concentration of iodine (20 mg/mL) is also shown for comparison on the left. Significant streaking and shading artifact is observed in single energy circular orbit as a result of both the metal and the high density iodinated vessel. Such artifacts are reduced for spectral circular acquisition, where imaging targets are somewhat visible in both the water and iodine bases. Using the non-circular orbit, artifacts are significantly reduced, clearly showing the bleed and two of the extravasation targets. It is however impossible to distinguish them due to similar attenuation. The overlapping ischemia and extravasation are not visible as expected. Spectral orbit acquisitions demonstrate the best performance, reducing artifacts resulting from the metal and the high intensity iodinated vessel. The bleed appears in the water basis while extravasation appears in the iodine basis. The overlapping ischemia and extravasation are both visible in their respective basis. However, the increase in image noise in the

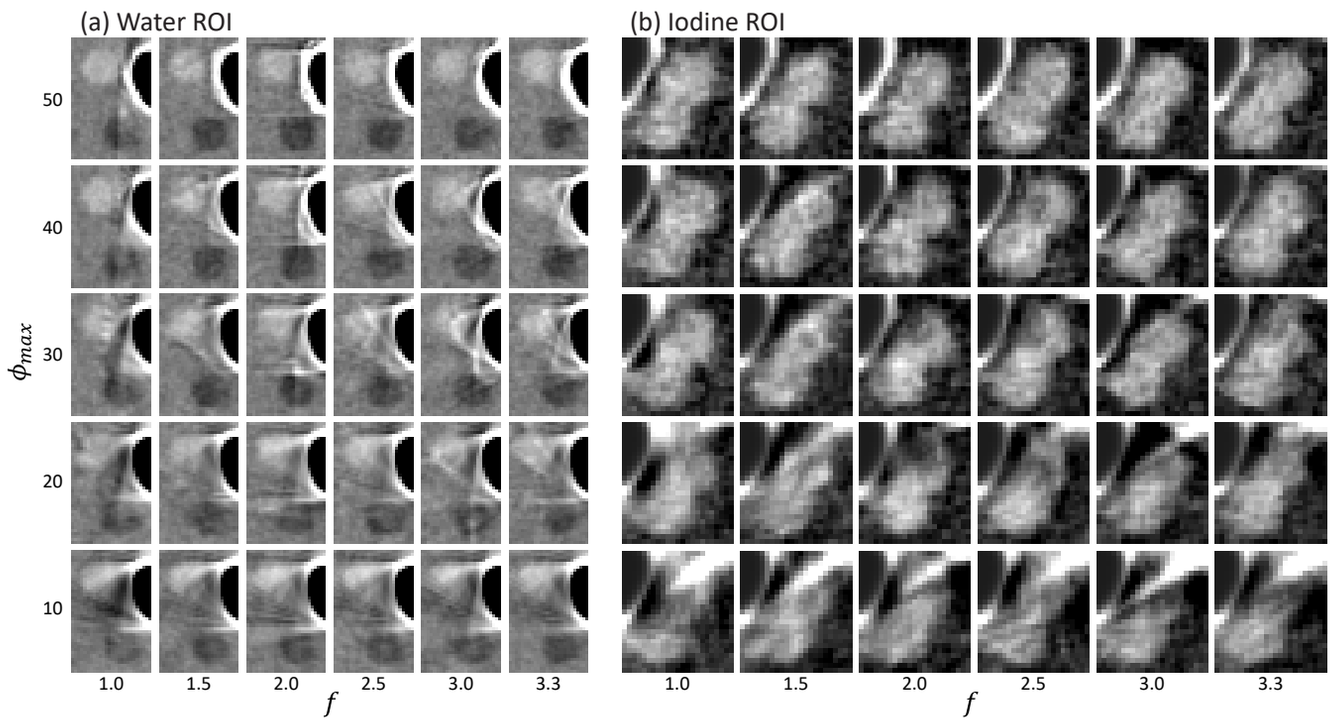


Fig. 5: ROIs from water and iodine basis as a function of the maximum tilt (ϕ_{max}) and frequency (f) of the sinusoidal orbits.

spectral results is evident due to noise amplification from material decomposition.

To illustrate the relative performance of different sinusoidal orbits, two VOIs containing imaging targets around the embolization were extracted from the water and iodine basis and displayed in Fig.5 as a function of the frequency, f, and the maximum tilt, ϕ_{max}, of the sinusoidal orbit. The RMSE within the VOIs indicated in Fig.2 are shown in Fig.4. The trend in imaging performance is consistent with single-energy imaging, where larger tilts, even, and non-integer frequencies generally outperforms the rest.

IV. Discussions and Conclusions

In this work, we demonstrated the potential advantages of combining spectral imaging and non-circular orbits in an example neuro-interventional application. Such a method is effective in reducing streaking and shading artifacts that are the result of highly attenuating metal and iodinated structures in the phantom. Material decomposition enabled by spectral imaging allows the differentiation of imaging targets that appear similar in single-energy imaging, e.g., bleeds and contrast extravasation.

Future work include the following main areas: (1) Joint optimization of orbit and spectral imaging techniques: The sinusoidal orbits considered in this initial investigation are driven by sampling completeness. Potential improvements in imaging performance may be obtained with additional consideration of factors like spectral separation; (2) Realizing spectral orbits in physical experiments; and (3) Developing noise mitigation strategies in material decomposition beyond a simple quadratic regularizer.

Acknowledgments

The author is thankful for helpful discussions with Dr. J. Webster Stayman. This work is supported, in part, by NIH Grant R01EB027127.

References

[1] J. W. Stayman *et al.*, *Journal of Medical Imaging*, vol. 6, no. 2, pp. 025 002–025 002, 2019.

[2] G. J. Gang and J. W. Stayman, *Physics in Medicine & Biology*, vol. 67, no. 11, p. 115008, 2022.

[3] P. Wu *et al.*, *Physics in Medicine & Biology*, vol. 65, no. 16, p. 165012, 2020.

[4] A. M. Kingston *et al.*, *IEEE Transactions on Computational Imaging*, 2018.

[5] Z. Wang *et al.*, *Medical Physics*, vol. 50, no. 11, pp. 6762–6778, 2023. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.16711

[6] L. Shi *et al.*, in *Medical Imaging 2020: Physics of Medical Imaging*, G.-H. Chen and H. Bosmans, Eds., vol. 11312, International Society for Optics and Photonics. SPIE, 2020, p. 1131220. [Online]. Available: https://doi.org/10.1117/12.2549781

[7] L. P. Liu *et al.*, *Scientific Reports*, vol. 13, no. 1, 2023.

[8] Y. Zaouak *et al.*, *Journal of the Belgian Society of Radiology*, vol. 104, no. 1, 2020.

[9] I. Riederer *et al.*, *Clinical Imaging*, vol. 79, pp. 158–164, 2021.

[10] L. Shi *et al.*, *Medical physics*, vol. 47, no. 8, pp. 3332–3343, 2020.

[11] S. Tilley *et al.*, *Physics in Medicine & Biology*, vol. 64, no. 3, p. 035005, 2019.

[12] W. Wang *et al.*, *Medical physics*, vol. 48, no. 10, pp. 6375–6387, 2021.

Non-unique water and contrast agent solutions in dual-energy CT

JP Phillips, Emil Y. Sidky, Fatma Terzioglu, Ingrid S. Reiser, Guillaume Bal, Xiaochuan Pan

***Abstract*—The goal of this work is to study occurrences of non-unique solutions in dual-energy CT (DECT) for objects containing water and a contrast agent. Previous studies of the Jacobian of nonlinear systems identified that a vanishing Jacobian determinant indicates the existence of multiple solutions to the system. Vanishing Jacobian determinants are identified for DECT setups by simulating intensity data for practical thickness ranges of water and contrast agent. Once existence is identified, non-unique solutions are found by simulating scan data and finding intensity contours with that intersect multiple times. With this process non-unique solutions are found for DECT setups scanning iodine and gadolinium, including setups using tube potentials in practical ranges. Non-unique solutions demonstrate a large range of differences and can result in significant discrepancies between recovered and true material mapping.**

***Index Terms*—Dual-energy CT image reconstruction, contrast agents, non-uniqueness**

I. Introduction

DUAL-energy CT (DECT) systems are able to identify two different materials in a scanned object by simultaneously collecting two different spectral measurements. Hounsfield introduced the concept of DECT in the 1970s with the mathematical framework being developed in 1976 by Alvarez and Macovski [1][2]. Their approach for material decomposition depends on the energy dependence of a material's linear attenuation coefficients, a property that can be exploited by making measurements with two different x-ray spectra.

This work focuses on projection-based material decomposition in DECT which requires solving a nonlinear system of equations for x-ray paths. When determining the value of projection line integrals, the nonlinearity of the mapping from integral to DECT measurements results in the possibility of non-unique line integral values. Levine first demonstrated an example of non-unique solutions for a water and bone material basis using a spectra of three discrete photon values [3]. Alvarez later investigated uniqueness in relation to a nonvanishing Jacobian and found in general that only local uniqueness is guaranteed by a nonvanishing Jacobian [4]. This work seeks to demonstrate non-uniqueness when using basis materials of water and CT contrast agents iodine and gadolinium. In addition to using different basis materials from past studies, this work will also look for non-unique solutions using full x-ray spectra rather than discrete photon energies.

J. P. Phillips, E.Y Sidky, I. S. Reiser, and X. Pan are with the Department of Radiology MC-2026, The University of Chicago, 5841. S. Maryland Ave., Chicago, IL, 60637.

F. Terzioglu is with the Department of Mathematics, North Carolina State University, 2311 Stinson Dr., Raleigh, NC, 27695

G. Bal is with the Departments of Statistics and Mathematics, The University of Chicago, 5747 S. Ellis Ave., Chicago, IL, 60637

II. Methods

For contrast imaging with a DECT system, we assume that only water and the contrast agent are present in the scanned object. Such an object is scanned using x-ray spectra $S_1(E)$ and $S_2(E)$ which correspond to a low and high x-ray tube potentials respectively. Using the publicly available Python package SpekPy, x-ray spectra can be modeled for a given tube potential [5]. For such sources, the total number of photons transmitted through an object is

$$I_i(t_w, t_c) = \int_0^{E_{max}} S_i(E) e^{-\mu_w(E)t_w - \mu_c(E)t_c} dE,\ i = 1, 2 \tag{1}$$

where $\mu_w(E)$ and $\mu_c(E)$ represent the energy dependant attenuation coefficients of water and the contrast agent respectively while t_w and t_c are the thickness of water and contrast agent respectively in centimeters. X-ray spectra are normalized so that intensity is measured relatively and will only have values between 0 and 1.

The nonlinearity of the system in Eq. 1 gives rise to non-uniqueness, resulting in different pairs of t_w and t_c values producing the same intensity values. Non-uniqueness can be prevented if parameters are selected that ensure the measured intensity values can only be obtained with one set of thicknesses. The Jacobian of the system can be used to guide parameter selection.

For notation, we will define

$$\mu = \begin{bmatrix} \mu_w(E) \\ \mu_c(E) \end{bmatrix},\ \ t = \begin{bmatrix} t_w \\ t_c \end{bmatrix}$$

When looking at the Jacobian, we will consider the negative logarithm of the intensity measurements defined as

$$g_i(t) = -\ln I_i(t),\ i = 1, 2 \tag{2}$$

For the system in Eq. 1, the entries of Jacobian matrix $J(t)$ are given by

$$J_{ij}(t) = \frac{\partial g_i}{\partial t_j}(t) = \frac{\int_0^{E_{max}} S_i(E)\mu_j(E)e^{-\mu\cdot t} dE}{\int_0^{E_{max}} S_i(E)e^{-\mu\cdot t} dE},\ i, j = 1, 2 \tag{3}$$

The matrix from Eq. 3 is then used to calculate the Jacobian determinant

$$det(J) = J_{11}(t)J_{22}(t) - J_{12}(t)J_{21}(t) \tag{4}$$

It has previously been proved that should Eq. 4 not equal zero for all $t \in \mathbb{R}$, then the estimation of t from $I(t)$ is unique

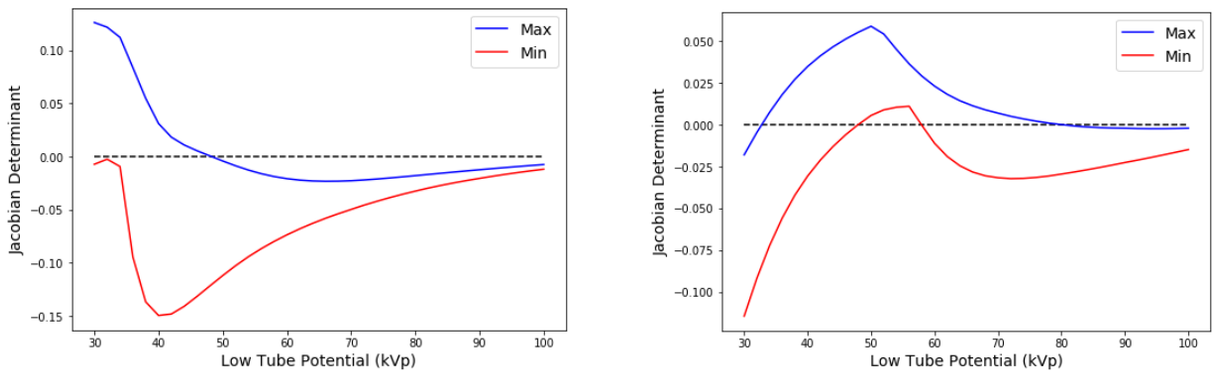


Fig. 1. Plots of Jacobian determinant for thickness sweep. The blue and red curves correspond to the maximum and minimum values of the determinant among all thickness combinations. High tube potential was held at 120 kVp and low tube potential was varied from 30 to 100 kVp. Materials were water and iodine (left) or gadolinium (right).

[6][7]. For this work we demonstrate the dependence of the Jacobian determinant on material basis pair and x-ray spectra. For high tube potentials of 120 kVp and 140 kVp, values of $I_i(t_w, t_c)$ were simulated for t_w varied between 0 and 30 cm and t_c varied between 0 and 3 cm for iodine and gadolinium with a 1% concentration. Low tube potential was varied from 30 to 100 kVp for the simulation. Eq. 3 and Eq. 4 were used to calculate the Jacobian determinant for every pair of intensities measured for a low and high tube potential combination. If zero is contained between the minimum and maximum values of the Jacobian determinants for a tube potential pair, then there exists a non-unique solution.

To search for the values of non-unique solutions, a grid search of solutions to Eq. 1 for values of $I_i(t_w, t_c)$ ranging from 0 to 1 was completed. For a given pair of $I_i(t_w, t_c)$ values, contour curves specified by t_w and t_c were drawn. t_w and t_c were restricted to ranges of 0 to 30 cm and 0 to 3 cm respectively. If the contours of I_1 and I_2 intersected once, then there was a set of thicknesses that resulted in measuring the specified intensity values. If the contours intersected twice, than there were two pairs of thicknesses that would solve Eq. 1 for the values of I_1 and I_2 and the solutions would not be unique.

From the study of the Jacobian determinant, we know that changing the x-ray spectra and materials will effect the existence of non-unique solutions. The results of examining the Jacobian determinant are useful for parameter selection as they highlight where non-unique solutions are likely to appear. Based on the Jacobian, for iodine the low tube potential was varied from 30 to 50 kVp and the high tube potential from 70 to 140 kVp. For gadolinium, the low tube potential was varied from 40 to 80 kVp and the high tube potential from 100 to 140 kVp.

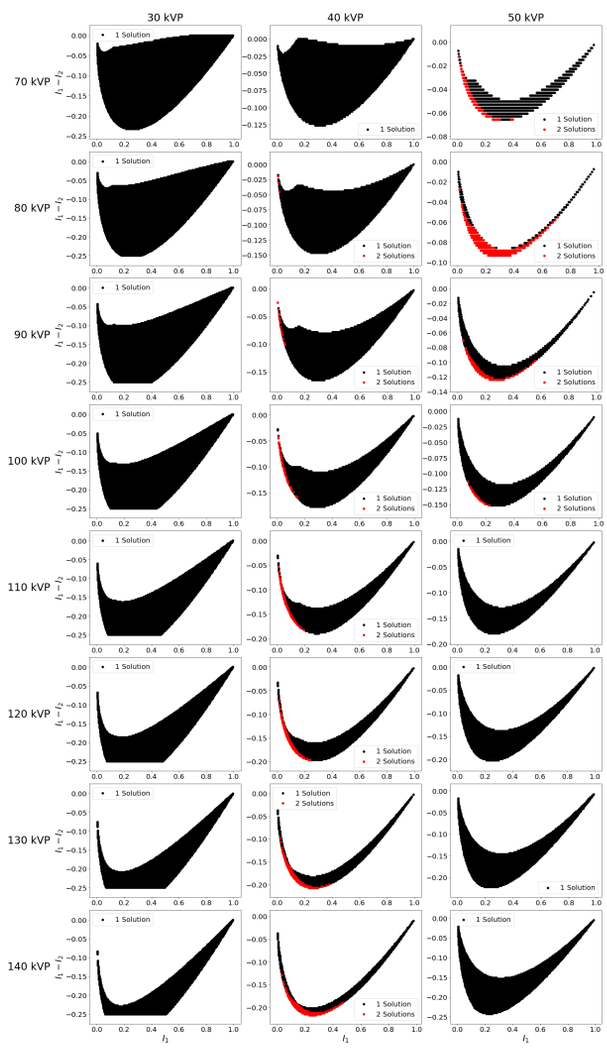


Fig. 2. Intensity sweep plots for water and iodine. Low tube potential increases from 30 kVp to 50 kVp moving left to right and high tube potential increases from 70 kVp to 140 kVp moving top to bottom. Black areas indicate intensity combinations that only have one solution to Eq. 1 and red areas indicate where there are two solutions.

III. Results

Using Eq. 3 and Eq. 4, values of the Jacobian determinant were calculated. Figure 1 shows how the Jacobian determinant changes for water-iodine and water-gadolinium material pairs as low tube potential is varied while high tube potential is held constant at 120 kVp. Since non-unique solutions occur

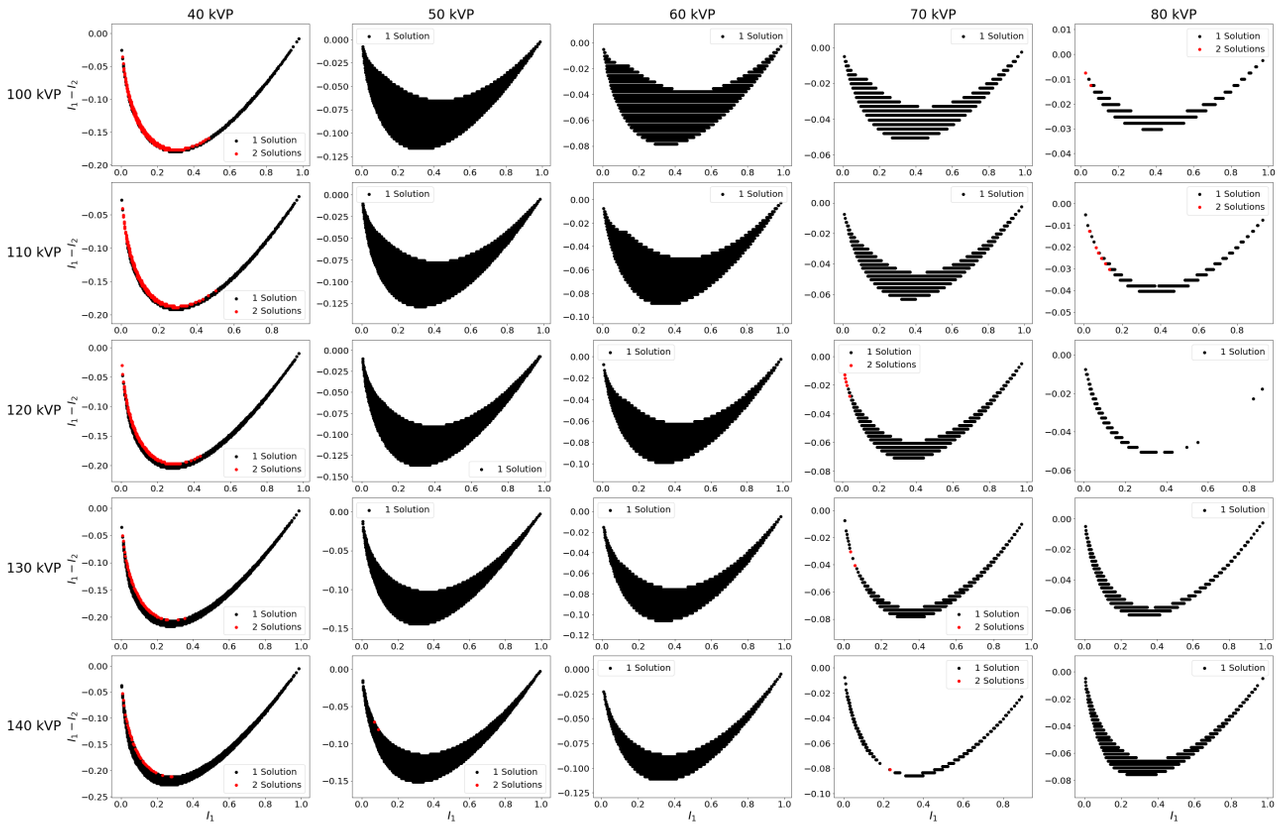


Fig. 3. Intensity sweep plots for water and gadolinium. Low tube potential increases from 40 kVp to 80 kVp moving left to right and high tube potential increases from 100 kVp to 140 kVp moving top to bottom. Black areas indicate intensity combinations that only have one solution to Eq. 1 and red areas indicate where there are two solutions.

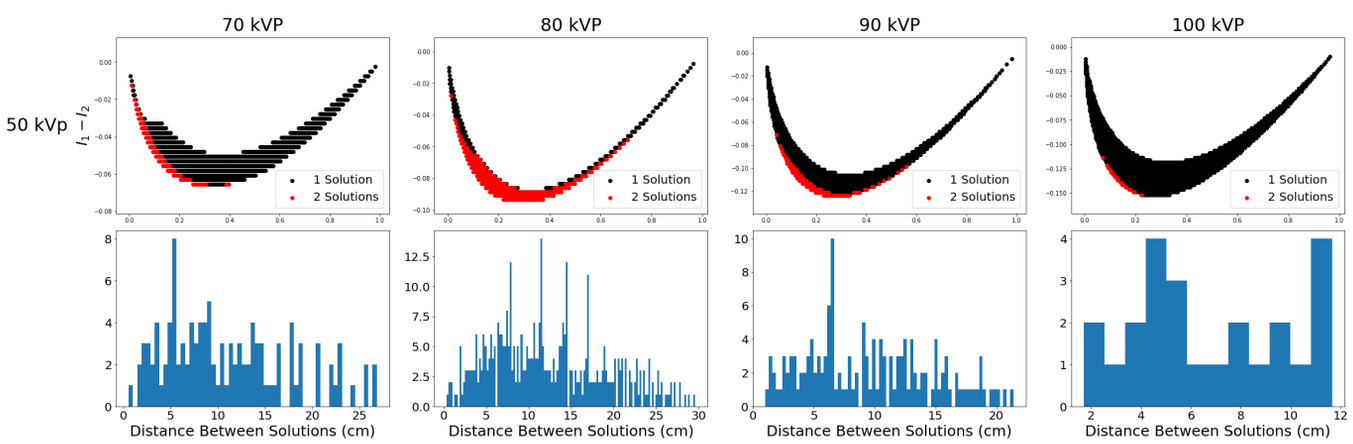


Fig. 4. Selected solution plots for water and iodine. Low tube potential is held at 50 kVp and high tube potential increases from 70 to 100 kVp from left to right. The same solution plots from Fig. 2 are shown (top) along with corresponding histograms of the Euclidean distance between each pair of solutions for a given intensity pair (bottom).

when the Jacobian determinant vanishes, the plots in Figure 1 suggest that multiple solutions are likely to be found for low tube potentials in the 30-50 kVp range for iodine and between the 35-50 kVp and 60-80 kVp ranges for gadolinium.

All possible solutions to Eq. 1 for given sets of tube potentials when measuring 0 to 30 cm of water and 0 to 3 cm of iodine are shown in Figure 2. Similar results for gadolinium are displayed in Figure 3. For both figures sparsity of solutions in some plots is due to discrete sampling of intensities. Multiple tube combinations resulted in non-unique solutions as denoted by red regions on the plots. For iodine non-uniqueness occurred for a low tube potential of 40 kVp and high tube potential between 100 and 140 kVp. Non-uniqueness also occurred for a low tube potential of 50 kVp and high tube potential between 70 and 100 kVp. When using gadolinium as the contrast agent, non-uniqueness occurred when using a low tube potential of 40 kVp and all high tube potentials studied. For low tube potentials of 70 and 80 kVp,

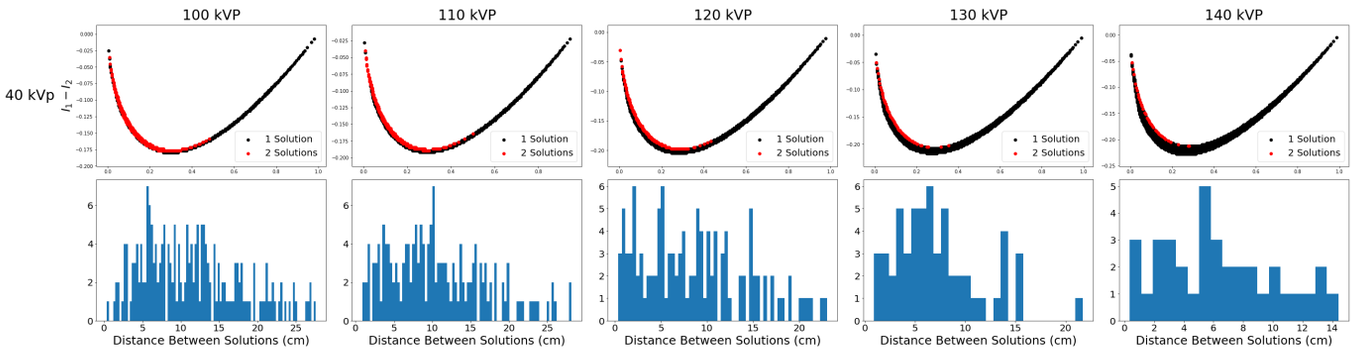


Fig. 5. Selected solution plots for water and gadolinium. Low tube potential is held at 40 kVp and high tube potential increases from 100 to 140 kVp from left to right. The same solution plots from Fig. 3 are shown (top) along with corresponding histograms of the Euclidean distance between each pair of solutions for a given intensity pair (bottom).

a small number of non-unique solutions were found. While small in number, these points are significant as they occur in the practical range of DECT which typically use 70 to 100 kVp and 135 to 150 kVp [8].

The red points in Figure 2 and Figure 3 can be plotted on a plane defined by t_w and t_c. The Euclidean distance between the points d can then be calculated with

$$d = \sqrt{(t_w^{I_1} - t_w^{I_2})^2 + (t_c^{I_1} - t_c^{I_2})^2} \tag{5}$$

The farther away the points are, the greater the difference between the values of the solutions. Figure 4 displays the solutions plots for iodine using a 50 kVp low tube potential and corresponding histograms showing d for all two solution points. Figure 5 shows similar results for gadolinium using a 40 kVp low tube potential. The histograms for iodine and gadolinium indicate that some intensities will result in two solutions that are significantly different.

IV. Conclusion

The main results from this work identify DECT parameters that can result in non-unique material mapping when scanning an object containing water and iodine or gadolinium. Most significant is the identification of multiple solutions when using practical tube potentials that are used in a clinical setting. Furthermore, the discrepancy between possible solutions can be large and result in a significant error in material mapping.

This work will be extended to examine non-uniqueness in the practical range more closely and look for non-uniqueness for more material combinations. Results using a finer sampling of some parameter combinations will also be presented at the conference.

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 R01EB023968 and R21CA263660.

References

[1] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. Description of system," *The British Journal of Radiology*, vol. 46, pp. 1016–1022, 1973.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Physics in Medicine & Biology*, vol. 21, p. 733, 1976.

[3] Z. H. Levine, "Nonuniqueness in dual-energy CT," *Medical Physics*, vol. 44, pp. e202–e206, 2017.

[4] R. E. Alvarez, "Invertibility of the dual energy x-ray data transform," *Medical Physics*, pp. 93–103, 2019, _eprint: https://aapm.onlinelibrary.wiley.com/doi/pdf/10.1002/mp.13255.

[5] 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*, pp. 3630–3637, 2021.

[6] D. Gale and H. Nikaido, "The Jacobian matrix and global univalence of mappings," *Mathematische Annalen*, vol. 159, pp. 81–93, 1965.

[7] G. Bal and F. Terzioglu, "Uniqueness criteria in multi-energy CT," *Inverse Problems*, vol. 36, no. 6, p. 065006, Jun. 2020.

[8] F. Tatsugami, T. Higaki, Y. Nakamura, Y. Honda, and K. Awai, "Dual-energy CT: minimal essentials for radiologists." *Japanese journal of radiology*, vol. 40, pp. 547–559.

Tube Voltage Modulation for Contrast-Enhanced CT Scans

Edith Baader and Marc Kachelrieß

***Abstract*—While tube current modulation (TCM), which modulates the tube current as a function of the tube's angular position, is available at clinical CT scanners, automatic tube voltage selection selects a constant tube voltage. Here, we investigate whether an additional modulation of the voltage, i.e. an adjustment of the voltage to the view angle, would yield better image quality for iodine-enhanced scans at constant effective dose. We compare the combined tube voltage and tube current modulation to conventional mAs-minimizing TCM, TCM minimizing the effective dose, and a combined tube voltage and tube current modulation which is optimized for noise for scans without iodine. Our results confirm that the voltage should be decreased when iodine is present and suggest that there might be a slight advantage of a tube voltage modulation for anatomical regions with a high variation of attenuation for different view angles.**

***Index Terms*—Automatic exposure control, tube current modulation, dose reduction, CT.**

I. Introduction

AUTOMATIC exposure control techniques aim at keeping the radiation dose as low as reasonably achievable. Among these, automatic tube voltage selection (ATVS) and tube current modulation (TCM) [1], [2] are fundamental concepts. In TCM, the tube current becomes a function of the angular and longitudinal position of the X-ray tube. In conventional TCM, the tube current is optimized so that the tube output (mAs-product) is minimal for a certain noise level. Therefore, we call the conventional TCM mAsTCM in the following. So-called organ-based tube current modulations instead aim at reducing the dose to a certain organ by lowering the tube current anterior to this organ. Recently, riskTCM, a TCM which minimizes patient dose in terms of effective dose, was proposed [3]. ATVS on the other hand selects the optimal voltage for a scan which depends on the attenuation of the patient. In general, the optimal voltage increases with the thickness of the patient. Furthermore, the optimal voltage decreases if the scan is iodine-enhanced which is a common contrast medium. This is due to the K-edge of iodine at $33\,\mathrm{keV}$ which results in a higher contrast between water and iodine at lower energies. In contrast to clinical TCM techniques which modulate the tube current, current ATVS approaches in clinical CT select a constant voltage. We here want to study whether modulating the tube voltage gives any benefit in patient dose or image quality. We have previously developed riskTCTVM, a combined tube current and tube voltage modulation that minimizes the effective dose for non-enhanced scans [4]. For riskTCTVM, we found only slight improvements in effective dose over a sole TCM for circular scans. Here, we extend riskTCTVM to iodine-enhanced scans for which we consider the iodine contrast in addition to the image noise in our optimization process.

Edith Baader and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls-University, Heidelberg, Germany.

Corresponding author: edith.baader@dkfz.de

II. Methods

A. Patient model

We perform our simulations based on clinical whole-body CT scans from the Visual Concept Extraction Challenge in Radiology (Visceral) [5] with the organ segmentations that were created for the study of riskTCM [3].

B. Dose estimation

For circular scans, we obtained dose distributions $D(\alpha, \boldsymbol{r})$ for different angular positions α of the X-ray tube using our in-house Monte Carlo dose simulation software. From these dose distributions, organ doses and effective dose values $D_{\text{eff}}(\alpha)$ with the weighting factors as proposed by the International Commission on Radiological Protection (ICRP) [6] were calculated. To calculate dose values for arbitrary tube spectra, we performed the dose simulations for photon energies in uniform intervals of $5\,\mathrm{keV}$ between $20\,\mathrm{keV}$ and $150\,\mathrm{keV}$. In the simulations, the photons were attenuated by $6\,\mathrm{mm}$ aluminum and a bowtie filter. For our optimization, we consider tube voltages U between $70\,\mathrm{kV}$ and $150\,\mathrm{kV}$. Using Tucker spectra [7] for the respective voltages, we calculated the proportion of photons in every simulated energy interval and weighted the dose values respectively.

C. Optimization objective

The scans from the Visceral dataset were performed without contrast material. To study the effect of contrast on a combined tube current and tube voltage modulation, we overlay the CT volume with a circular cylinder of diameter $l_{\mathrm{I}} = 2\,\mathrm{cm}$ filled with iodine of density $4.93\,\mathrm{mg\,cm^{-3}}$ which corresponds to a CT value of around 250 HU at $70\,\mathrm{kV}$.

We aim to find the optimal tube current I and tube voltage U as a function of the projection angle α for circular scans. With optimal, we denote here the tube current $I(\alpha)$ and tube voltage modulation $U(\alpha)$ which maximize the contrast to noise ratio at unit dose (CNRD):

$$\mathrm{CNRD} = \frac{C}{\sqrt{D/D_{\text{ref}}} \cdot \sigma}. \tag{1}$$

Here, C denotes the contrast, σ the image noise, D the dose (in our case the effective dose), and D_{ref} a reference

dose value. With I, we denote in this work not the electric current, but the photon current leaving the X-ray tube which is proportional to the electric current of the tube. However, the proportionality depends on the voltage.For the optimization process, we use the simplex algorithm [8] to find the tube current and tube voltage curves yielding the maximum CNRD. Note, that the CNRD as defined in equation (1) is invariant to a constant scale factor λ of the tube current, i.e. $\mathrm{CNRD}(\lambda I(\alpha)) = \mathrm{CNRD}(I(\alpha))$, as $D \propto I$ and $\sigma \propto 1/\sqrt{I}$. To shorten computation time, our optimization relies on estimates of the noise and the contrast for the central pixel.

D. Noise estimation

We divide every intersection through the patient into a water-equivalent pathlength l_W for water of density $\rho_\mathrm{W} = 1\,\mathrm{g\,cm^{-3}}$ and an intersection through iodine with density $\rho_\mathrm{I} = 4.93\,\mathrm{mg\,cm^{-3}}$.We can therefore calculate the measured rawdata value q by

$$q = -\ln \int dE\, w(U,E) \exp^{-l_\mathrm{W}\psi_\mathrm{W}(E) - l_\mathrm{I}\psi_\mathrm{I}(E)}. \quad (2)$$

Here, $w(U,E)$ denotes the detected spectrum which would occur for a Tucker spectrum [7] for a voltage U with $6\,\mathrm{mm}$ aluminum prefiltration and a $1.4\,\mathrm{mm}$ gadolinium oxysulfide scintillator with density $7.32\,\mathrm{g\,cm^{-3}}$. $\psi_\mathrm{W}(E)$ and $\psi_\mathrm{I}(E)$ denote the attenuation coefficients for water and iodine with density ρ_W and ρ_I respectively. As Klein et al. [3], we assume the measured signal Ie^{-q} to be Poisson-distributed. Then, error propagation yields for the water-precorrected projection value $p = P(q,U)$:

$$\mathrm{Var}\; p \propto \frac{e(p,U)}{I} \;\text{ with }\; e(q,U) = (P'(q,U))^2 e^q.$$

Let us now consider two complementary rays at source positions α_C and α_D. For a reconstruction algorithm with statistically optimally weighting, i.e. inversely to their variance, the resulting variance becomes

$$V_\mathrm{opt} = \frac{1}{1/V_\mathrm{D} + 1/V_\mathrm{C}} \quad (3)$$

$$= \left(I_\mathrm{D}/e(q(\alpha_\mathrm{D}, +\beta), U_\mathrm{D}) + I_\mathrm{C}/e(q(\alpha_\mathrm{C}, -\beta), U_\mathrm{C})\right)^{-1} \quad (4)$$

In the optimization process, we estimate the noise solely for the central pixel which results in a simple summation of V_opt over $180°$ with projection values of the central ray with $\beta = 0$.

E. Contrast estimation

The water precorrection converts us the intersection lengths through water and iodine into a single intersection length through water $\hat{p}$. From this intersection length $\hat{p}$, we can extract the contrast value C_I of our iodine which corresponds to the density that water would need to produce the same attenuation. We can therefore rewrite equation (2)

$$q = -\ln \int dE\, w(U,E) \exp^{-\hat{p}\psi_\mathrm{W}(E)}$$

$$= -\ln \int dE\, w(U,E) \exp^{-l_\mathrm{W}\psi_\mathrm{W}(E) - l_\mathrm{I}C_\mathrm{I}\psi_\mathrm{W}(E)}.$$

We can therefore identify

$$\hat{p} = l_\mathrm{W} + l_\mathrm{I} C_\mathrm{I}$$

and calculate the contrast via

$$C_\mathrm{I} = \frac{\hat{p} - l_\mathrm{W}}{l_\mathrm{I}}.$$

As we put the contrast in the isocenter, we estimate the contrast for the central pixel. Thus, we calculate C_I for all α for the central ray ($\beta = 0$) and build the sum over all α with the weighting of complementary rays by their variance (see equation (3)).

F. Image simulation

For the image simulation, we calculate the polychromatic projection values as defined in equation (2). We then add noise, perform a water precorrection, the weighting of complementary rays (see equation (3)), and a filtered backprojection.

G. Baseline TCM modulations

We compare our obtained riskTCTVM modulations to conventional mAsTCM and the previously proposed riskTCM [3]. Here, we modified the original riskTCM optimization process which optimizes the noise throughout the whole patient to optimize only the noise in the central pixel for better comparison with riskTCTVM optimizing the CNRD. For mAsTCM we considered also only the projection values of the central rays. We tested which (constant) voltage between $70\,\mathrm{kV}$ and $150\,\mathrm{kV}$ in steps of $10\,\mathrm{kV}$ yields the best CNRD for both mAsTCM and riskTCM and selected this voltage.

III. Results

We tested our optimization approach first for one patient for a slice in the pelvis region. We chose the pelvis region as we expect the highest impact of a tube voltage modulation for scans with a high variation of projection values for different view angles as it is especially the case for the shoulder and the pelvis region.

Figure 1 shows the resulting modulation curves. For mAsTCM and riskTCM, we found the best CNRD values for a voltage of $U = 70\,\mathrm{kV}$. Compared to mAsTCM, riskTCM reduces the noise slightly at the same effective dose. riskTCTVM reduces the noise further at the cost of decreasing also the contrast. However, the resulting CNRD value is slightly higher than for riskTCM.

IV. Discussion

Currently, we use a noise and contrast estimate of the central pixel where we also artificially added iodine to the CT data. In reality, however, the contrast distribution is neither known a priori nor only concentrated in the isocenter. Therefore, we will investigate if the exact distribution of the contrast is important for the optimization or if it is just important whether contrast is present. Also, we might extend the noise estimate to a noise estimation inside the whole patient as proposed for riskTCM [3]. Furthermore, we currently weigh complementary

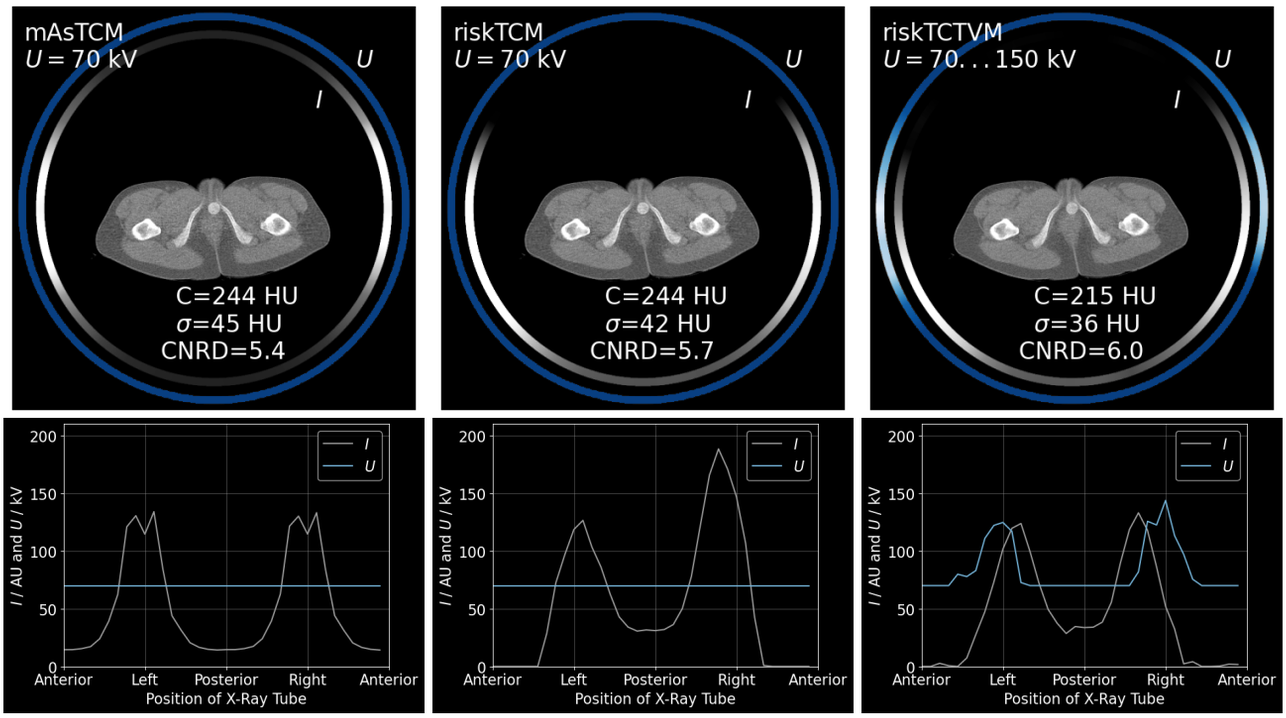


Fig. 1. Image Simulations for mAsTCM at 70 kV (left), riskTCM at 70 kV (center), and riskTCTVM (right). All modulations were normalized to the same effective dose value. The values for the contrast C and the image noise σ were calculated using the contrast and noise estimation used in our optimization process.

rays, so that the variance is optimal. However, we want to optimize the CNRD and will therefore investigate different weighting strategies.

In the shown example (see figure 1), the optimal voltage seems to be a constant voltage of 70 kV for TCM without tube voltage modulation. However, contrast-enhanced scans are often performed at a higher voltage in clinical practice if the patient is not sufficiently thin. Therefore, we will test our approach also on patients who would be scanned at a higher voltage than 70 kV at a clinical scanner.

V. Conclusion

We developed a combined tube current and tube voltage modulation for scans with contrast material. Our results suggest that there might be a slight advantage of a tube voltage modulation for anatomical regions with a high variation of attenuation for different view angles over a plain tube current modulation at 70 kV. We will test our method on more patients, especially patients who would be scanned at a higher voltage than 70 kV on a clinical scanner.

Acknowledgment

This work was supported in part by a DKFZ International PhD Program Fellowship. The Monte Carlo dose estimation method RayConStruct-MC and the image reconstruction RayConStruct-IR was provided by RayConStruct® GmbH, Nürnberg, Germany.

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," *Med. Phys.*, vol. 26, no. 11, pp. 2248–2253, Nov. 1999.

[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, 2022.

[4] E. Baader, L. Klein, S. Sawall, J. Maier, and M. Kachelrieß, "CT tube current and tube voltage modulation to minimize the patient's radiation risk," *Proceedings of the European Congress of Radiology (ECR)*, 2023.

[5] O. Jimenez-del Toro, H. Müller, M. Krenn, K. Gruenberg, A. A. Taha, M. Winterstein, I. Eggel, A. Foncubierta-Rodríguez, O. Goksel, A. Jakab, G. Kontokotsios, G. Langs, B. H. Menze, T. Salas Fernandez, R. Schaer, A. Walleyo, M.-A. Weber, Y. Dicente Cid, T. Gass, M. Heinrich, F. Jia, F. Kahl, R. Kechichian, D. Mai, A. B. Spanier, G. Vincent, C. Wang, D. Wyeth, and A. Hanbury, "Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: VISCERAL anatomy benchmarks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 11, pp. 2459–2475, 2016.

[6] "ICRP, 2007. the 2007 recommendations of the international commission on radiological protection," 2007.

[7] D. M. Tucker, G. T. Barnes, and D. P. Chakraborty, "Semiempirical model for generating tungsten target x–ray spectra," *Med. Phys.*, vol. 18, no. 3, pp. 211–218, Mar./Apr. 1991.

[8] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

A basis-region data model for reconstructing multiple basis materials from DECT data

Buxin Chen, Zheng Zhang, Dan Xia, Emil Y. Sidky, and Xiaochuan Pan

***Abstract*—In this work, we report a basis-region data model and optimization-based algorithms for reconstructing multiple (>2) basis images only from dual-energy-CT (DECT) data. In CT imaging applications, it is not uncommon that some basis materials such as bone, metals, and contrast materials involved are confined spatially within specific regions in the image. Basing upon this observation, we devise a data model based on basis regions and an optimization-based algorithm to reconstruct basis-region images that can be used for obtaining multiple (>2) basis images and virtual monochromatic images (VMIs) from DECT data. Using simulated and real DECT data, we reconstruct basis-region images and subsequently the basis images and VMIs. The study results reveal that the algorithm developed can accurately and robustly reconstruct multiple (>2) basis images as well as VMIs over the energy range of interest directly from DECT data. The work may provide insights into the development of practical procedures for reconstructing multiple basis images, VMIs, and physical quantities from DECT data in applications.**

***Index Terms*—multiple basis material, DECT reconstruction, basis region**

I. Introduction

In current dual-energy CT (DECT), data are collected with two distinct low- and high-kV spectra, and basis images of two basis materials, often water and bone [1], [2], along with virtual monochromatic images (VMIs) at energy of interest, are reconstructed from the DECT data. However, there exists interest in obtaining multiple ($\geq$ 3) basis materials directly from DECT data, because the scanned subject might contain additional materials, such as adipose tissue, contrast materials, and metal implants/inserts. Applications that can potentially benefit from reconstructing multiple basis materials from DECT include virtual enhanced images in contrast-enhanced DECT, liver-fat quantification for detecting and diagnosing liver disease, and detection of bone marrow edema [3].

Works have been reported for decomposing DECT images into multiple basis materials by using the basis-material volume conservation constraint [3], [4]. Previous efforts also exist in reconstructing multiple basis images from DECT data by using constraints on basis triplet library and voxel-value bounds, in addition to the volume conservation [5], [6]. These constraints can be interpreted as a data augmentation technique for alleviating the ill-conditionedness in decomposing or reconstructing multiple ($\geq$ 3) basis materials from two sets of DECT data. In this work, we investigate a different approach to reconstructing multiple basis materials directly from DECT. Observing in CT imaging that some basis materials, e.g. bone, metals, and contrast agents of interest, are often confined within spatial supports smaller than the image array, we propose to partition the image array into a spatially complementary "basis region set" (BRS), where there are ≤ 2 basis materials in each basis region, and develop a basis-region data model.

We formulate the reconstruction problem of basis-region images directly from DECT data as a constrained optimization program, and develop a non-convex primal-dual (NCPD) algorithm to reconstruct basis-region images through numerically solving the optimization program. Basis images and VMIs on the entire image array can be obtained subsequently from the basis-region images reconstructed. Numerical studies are carried out with simulated and real DECT data. The NCPD algorithm is first verified for accurately reconstructing multiple basis images only from DECT data and then characterized for its robustness in the presence of inconsistencies, such as noise, and other physical factors inconsistent with the data model based on basis regions. Results are evaluated based on both visual inspection and quantitative analysis.

B. Chen, Z. Zhang, E. Y. Sidky, and D. Xia 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.

II. Materials and methods

A. The standard-data model in DECT

The standard data model in DECT can be written as

$$g_{sj}(\{b_{ki}\}) = -\ln \sum_{m_e=1}^{M_e} q_{sjm_e} \exp\left(-\sum_{i=1}^{I} a_{sji} f_{m_e i}\right), \quad (1)$$

where g_{sj}, q_{sjm_e}, a_{sji}, and $f_{m_e i}$ denote model data, spectrum, X-ray transform matrix element, and VMI, respectively, while s, j, m_e, and i index spectrum, ray, energy bin, and image pixel, ranging from 1 to S, $J^{(s)}$, M_e, and I, respectively. Often, a basis material decomposition model is used to decompose a VMI vector $\mathbf{f}_{m_e}$ into basis images as $f_{m_e i} = \sum_{k=1}^{K} \mu_{km_e} b_{ki}$, where b_{ki} denotes the basis image value of basis material $k = 1, \cdots, K$ at location $i = 1, \cdots, I$, and μ_{km_e} the LAC of basis material k at energy bin m_e. The reconstruction of basis images b_{ki} is tantamount to the determination of $K \times I$ unknowns from measured data with S distinct spectra. In DECT, algorithms exist for $K = 2$ basis images from $S = 2$ (i.e., DECT) data sets. For multiple basis materials with $K > 2$, the total number of unknowns $K \times I$ can be substantially larger than $2 \times I$.

B. The basis-region-data model in DECT

In this work, we develop a new data model based on basis regions to reduce the number of unknowns. We hypothesize

that an image array can be partitioned into spatially complementary basis regions. For each basis region $l = 1, \cdots, L$, it has a spatial mask $\mathcal{R}'_l$ of size $I_l \times I$, which picks out those image pixels inside the basis region l. Thus I_l is the number of image pixels within the basis region l. In addition, for each basis region, it also has a basis material set Φ_l, which includes those indices corresponding to the basis materials used in basis region l. $K_l = \text{card}(\Phi_l) \leq K$ denotes the number of basis materials in basis region l.

Let $\mathbf{b}'_{lk} = \mathcal{R}'_l \mathbf{b}_k$ of size I_l denote a *basis-region image* for material k within basis region l. A VMI $\mathbf{f}_{m_e}$ can then be obtained by combining basis-region images over all basis materials in a basis regions and then over all basis regions. As a result, a basis-region data model in DECT can be obtained by replacing $f_{m_e i}$ with basis-region images as [7]

$$\begin{aligned} & g'_{sj}(\{b'_{lki'}\}) \\ =& -\ln \sum_{m_e=1}^{M_e} q_{sjm_e} \exp\left(-\sum_{l=1}^{L}\sum_{i'=1}^{I_l} a'_{sjli'} \sum_{k\in\Phi_l} \mu_{km_e} b'_{lki'}\right), \end{aligned} \tag{2}$$

where $b'_{lki'}$ is the i'-th pixel in basis-region image $\mathbf{b}'_{lk}$ and $a'_{sjli'}$ is the contribution of the i'-th pixel in basis region l to ray j of spectrum s. There are thus $N = \sum_{l=1}^{L} K_l \times I_l$ total number of unknowns in the basis-region data model. Recall that $K_l \leq K$ and that the spatially complementary basis regions form a partition of the image array, we have $N \leq K \sum_{l=1}^{L} I_l = K \times I$.

In this work, we aim to obtain multiple ($K > 2$) basis images under the condition that $K_l \leq 2$ for all basis regions, so that the total number of unknowns satisfies $N \leq 2 \times N$.

C. The NCPD algorithm for inverting the basis-region data model in DECT

It can be observed that the standard DECT model in Eq. (1) and the basis-region data model in Eq. (2) are similar in form, where the basis-region images are linear transforms of the basis images. It has also been established above that the total number of unknowns in basis-region images is no more than those in a standard DECT under the condition $\mathcal{K}_l \leq 2$. Therefore, we seek to invert the basis-region data model in Eq. (2) for reconstructing basis-region images. Once the basis-region images $\mathbf{b}'^{*}_{lk}$ are reconstructed, basis images $\mathbf{b}^{*}_{k}$ on the full image array can be obtained by applying $\mathcal{R}'^{\top}_l$ to the basis-region images. The image reconstruction is formulated as a total-variation- (TV-) and non-negativity-constrained minimization problem [7], [8] as

$$\begin{aligned} & \min_{\mathbf{b}'} D(\mathbf{g}'(\mathbf{b}'), \mathbf{g}^{[\mathcal{M}]}) \\ & \text{s.t. } ||(|\nabla \mathcal{U}_m \mathbf{b}')|)||_1 \leq \gamma_m, \ m = 1, 2 \\ & \text{and } \mathbf{b}' \geq 0, \end{aligned} \tag{3}$$

where $D(\cdot,\cdot)$ is the data-fidelity term and also the objective function to be minimized, which takes the form of squared ℓ_2-norm in the work, ∇ the discretized matrix of spatial derivative, $\mathcal{U}_m$ a linear transform from basis-region images to VMI of energy m and $\mathbf{b}'$ the congregate basis-region image vector formed by concatenating individual basis-region images $\mathbf{b}'_{lk}$. The TV constraints are imposed on two VMIs at two well-separated energies for yielding accurate VMIs over a broad energy range of interest.

The non-convex primal-dual (NCPD) algorithm is developed using the same strategy as in the previous works [8] for numerically accurately solving the non-convex optimization program in Eq. (3). We first separate the non-linear data model into a linear and a nonlinear terms, then derive an instance of the general primal-dual (PD) algorithm based on the convexified objective function by replacing the non-linear term with a constant term, and finally complement the PD algorithm with a correction step which estimates and updates the non-linear term iteratively. The detailed derivation of the NCPD algorithm and its convergence conditions can be found in Ref. [7].

D. Numerical study designs

We performed evaluation studies using both simulated and real data. The simulated-data study is based on a digital torso phantom consisting of $K = 3$ basis materials, i.e., water, bone, and 20-mg/ml iodine. Within the phantom, the water basis material in included in the entire image array, while the bone and iodine basis materials are within two spatially non-overlapping, complementary basis regions, as shown in Fig. 1. DECT data are generated with a circular fan-beam geometry of 100-cm source-to-center (SOR) and 150-cm source-to-detector (SOD) distances. A linear detector of 70-cm length is in place with 896 detector bins, collecting data at 1440 views over 2π with low- and high-kV spectra at 80 and 140 kV. Images are reconstructed onto an image array of 200×256 square pixels of 0.14 cm in size.

Real dual-energy data of a clinical abdominal phantom are collected with a clinical DECT scanner in axial mode. Data from the central row of the detector are extracted, forming an equivalent fan-beam geometry with 60-cm SOR and 107-cm SOD and a 49° curved detector consisting of 896 bins. Low- and high-kV spectra are set at 80 and 135 kV, and data are collected at 1200 views over 2π. The same three basis materials, water, bone, and 20-mg/ml iodine, are used as the basis, and images are reconstructed onto an image array of 432×656 square pixels of 0.08 cm in size.

For simulated data, a basis-region set that matches exactly the spatial support of the truth basis images is used. Such partition is also referred to as an orthogonal basis-region set (OBRS), where the "orthogonality" comes from the fact $\mathcal{R}_l \mathbf{b}_k = 0$ for $k \notin \Phi_l$. Such a study with OBRS also serves as a verification study for the accurate inversion of the basis-region data model in Eq. (2) by the NCPD algorithm. For real data, a basis-region set is selected by thresholding the FBP image with different thresholds to obtain the basis regions. Reconstructed basis images and VMIs are inspected visually and quantitatively. Basis images are also reconstructed by use of a two-step method, referred to as the data-domain decomposition (DDD) method [2], based upon two basis materials of water and bone. Additional results with quantitative metrics, such as nRMSE and LAC, and with more basis-region sets to investigate the robustness of the algorithm with respect to the selection of basis regions, can be found in Ref. [7].

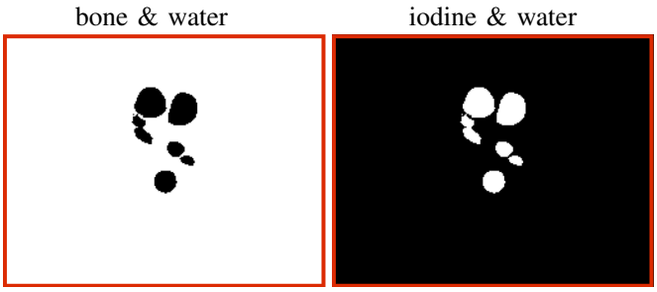


Fig. 1: A basis-region set of two spatially complementary basis regions containing bone & water (left) and iodine & water (right) basis materials.

III. Results

A. Simulated data

The basis-region set used in the simulated-data study is shown in Fig. 1. They are consistent with the spatial support of the truth basis images used to generate the data. As a result, the basis-region set is an OBRS.

We show in Fig. 2 basis images of water, bone, and iodine, obtained from basis-region images and VMIs at 33 keV reconstructed by using the NCPD algorithm with the basis region partion in Fig. 1 and by using DDD method. All three basis images can be accurately reconstructed and recovered by use of the basis-region data model and the NCPD algorithm. The VMI image at 33 keV, just below the iodine K-edge, correctly shows limited to no contrast in the iodine region. The DDD method, with basis materials selected as water and bone, does not contain iodine basis and thus can not yield accurate LAC values, especially in the energy range near the iodine K-edge. As shown, the VMI at 33 keV shows falsely elevated iodine contrast. We also compute quantative metrics nRMSE for VMIS of the energy range 25-120 keV, usnig the truth VMI as the reference. The nRMSE of the VMIs from $K = 3$ basis materials by use of the NCPD algorithm are consistently in the order of 10^{-5} across the energy range, verifying that the NCPD algorithm can numerically accurately solve the optimization program and invert the data model in Eq. (2) with $K = 3$, under the data condition studied. Meanwhile, the VMIs obtained from $K = 2$ basis materials by use of the DDD method are in the order of 10^{-1} across the energy range.

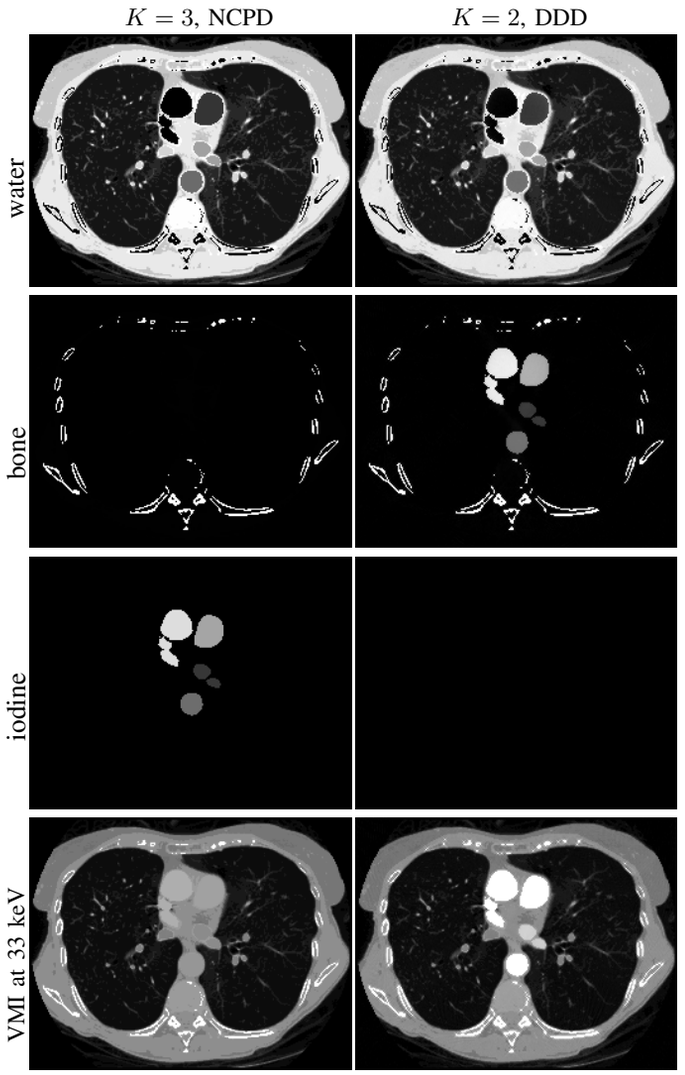


Fig. 2: Basis images of water (row 1), bone (row 2), and iodine (row 3) and VMI at 33 keV (row 4) obtained by using the NCPD algorithm (left) and by using the DDD method (right). Display windows: [0, 1.2] for basis images in rows 1-3 and [-1000, 1000] HU for VMI in row 4.

B. Real data

The real data was collected for a standard DE-472 phantom, which consists of solid water, calcium inserts of varying concentrations, and iodine inserts of varying concentrations. A basis-region set is used, as shown in Fig. 3, where the partition is obtained by applying appropriate threshold values to the image reconstructed from the 80-kV data by using the FBP algorithm. In real-data studies, no knowledge of truth is available, therefore, the basis region partition used here cannot be considered an OBRS.

We show in Fig. 4 basis images of water, bone, and iodine, and VMIs at 33 keV reconstructed by using the NCPD algorithm and by using the DDD method. Images of three basis materials, water, bone, and iodine, can be directly reconstructed from the DECT data by use of the basis-region data model and the NCPD algorithm. The VMI at 33 keV shows appropriate contrast levels for the calcium and iodine inserts, and no visible artifacts. As a result, the iodine concentration map can be generated and overlaid on top of the VMI image (not shown here). On the other hand, the DDD method only reconstructs two basis image of water and bone. The VMI image at 33 keV shows elevated contrast in the iodine inserts, which are likely false given that 33 keV is below the K edge.

IV. Discussion and Conclusion

In the work, we develop a basis-region data model and the NCPD algorithm for accurately reconstructing multiple ($K > 2$) basis images and VMIs directly from DECT data. Results of simulated- and real-data studies verify the accuracy and demonstrate the robustness of the NCPD algorithm. While

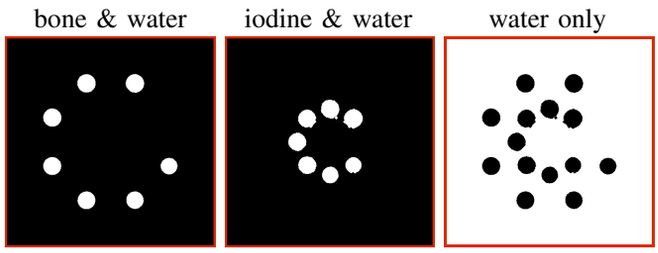


Fig. 3: A basis-region set of three spatially complementary basis regions containing water & bone (left), water & 20-mg/ml iodine (middle), and water-only (right) basis materials.

results included in the summary involve reconstruction of 3 basis materials from DECT data, additional studies involving reconstructing $>$ 3 basis materials from DECT data, more phantoms, and different basis-region sets can be found in Ref. [7].

V. Acknowledgment

This work was supported by NIH Grant Nos. R01EB026282 and 5R21CA263660-02. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

References

[1] T. G. Flohr, C. H. McCollough, H. Bruder, M. Petersilka, K. Gruber, C. Süβ, M. Grasruck, K. Stierstorfer, B. Krauss, R. Raupach *et al.*, "First performance evaluation of a dual-source CT (dsct) system," *Eur. Radiol.*, vol. 16, no. 2, pp. 256–268, 2006.

[2] Y. Zou and M. D. Silver, "Analysis of fast kV-switching in dual energy CT using a pre-reconstruction decomposition technique," in *Proc. SPIE Med. Imag.: Phys. Med. Imag.*, vol. 6913, 2008, p. 691313.

[3] P. R. Mendonça, P. Lamb, and D. V. Sahani, "A flexible method for multi-material decomposition of dual-energy CT images," *IEEE Trans. Med. Imag.*, vol. 33, no. 1, pp. 99–116, 2013.

[4] M. M. Goodsitt, D. I. Rosenthal, W. R. Reinus, and J. Coumas, "Two postprocessing CT techniques for determining the composition of trabecular bone," *Investig. Radiol.*, vol. 22, no. 3, pp. 209–215, 1987.

[5] Y. Long and J. A. Fessler, "Multi-material decomposition using statistical image reconstruction for spectral CT," *IEEE Trans. Med. Imag.*, vol. 33, no. 8, pp. 1614–1626, 2014.

[6] S. Z. Liu, M. Tivnan, G. M. Osgood, J. H. Siewerdsen, J. W. Stayman, and W. Zbijewski, "Model-based three-material decomposition in dual-energy CT using the volume conservation constraint," *Phys. Med. Biol.*, vol. 67, no. 14, p. 145006, 2022.

[7] B. Chen, Z. Zhang, D. Xia, E. Y. Sidky, and X. Pan, "Accurate reconstruction of multiple basis images directly from dual-energy data in ct," *IEEE Trans. Biomed. Eng.*, pp. 1–12, 2024.

[8] ——, "Non-convex primal-dual algorithm for image reconstruction in spectral CT," *Comput. Med. Imaging Graph.*, vol. 87, p. 101821, 2021.

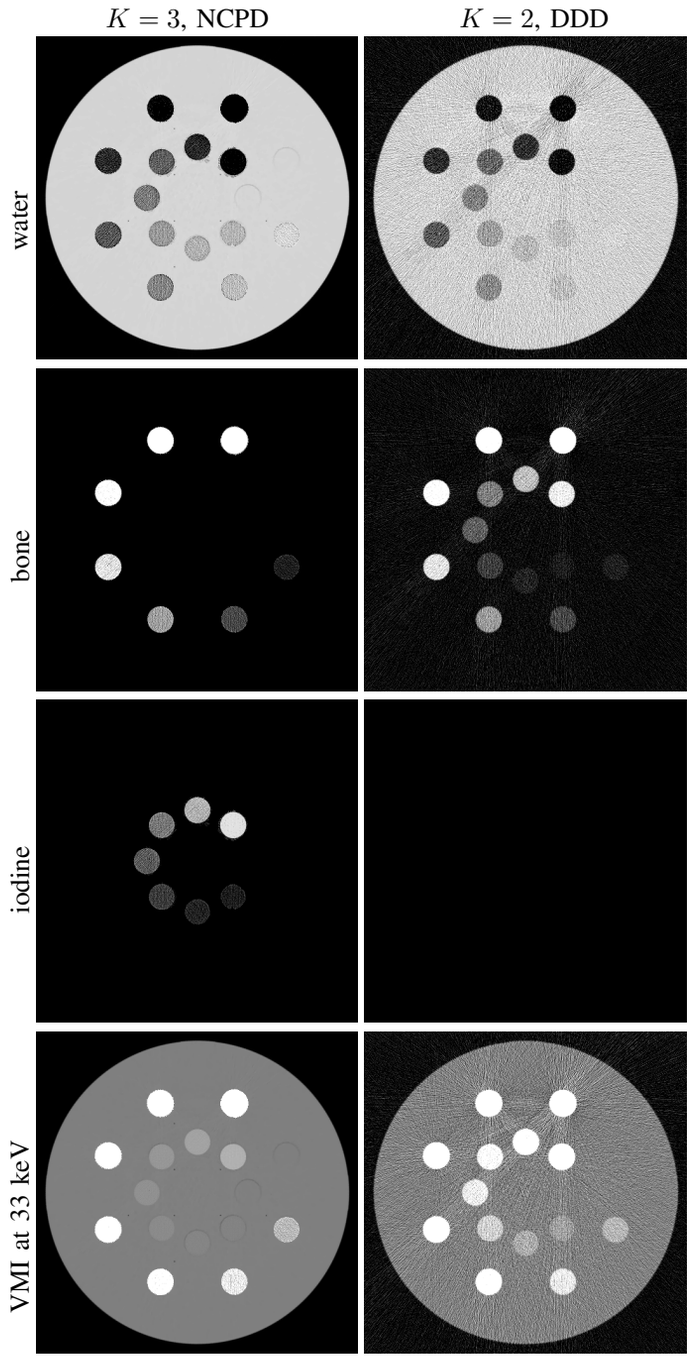


Fig. 4: Basis images of water (row 1), bone (row 2), and iodine (row 3), and VMIs at 33 keV (row 4) obtained by using the NCPD algorithm (left) and by using the DDD method (right). Display windows: [0, 1.2] for basis images and [-1000, 1000] HU for VMIs.

Extreme Few-View Tomography without Training Data

Gengsheng L. Zeng

***Abstract*—There are fewer than 10 projection views in extreme few-view tomography. The state-of-the-art methods to reconstruct images with few-view data are compressed sensing based. Compressed sensing relies on a sparsification transformation and total variation (TV) norm minimization. However, for the extreme few-view tomography, the compressed sensing methods are not powerful enough. This paper seeks additional information as extra constraints so that extreme few-view tomography becomes possible. In transmission tomography, we roughly know the linear attenuation coefficients of the objects to be imaged. We can use these values as the extra constraints. Computer simulations show that these extra constraints are helpful and improve the reconstruction quality.**

***Index Terms*—Inverse problem, optimization, total-variation minimization, few-view tomography, iterative algorithm, image reconstruction**

I. Introduction

EXTREME few-view tomography is referred to the situation where the number of the tomography measurement views is less than 10 [1-3]. If we model the data acquisition as a system of linear equations, the system is extremely under-determined for extreme few-view tomography. Constraints are vital in shrinking the solution space [4]. Iterative algorithms are better than analytical algorithms when the imaging system is under-determined [5-9].

The total-variation (TV) norm of the gradient of an image is a good indicator for the piecewise-constant feature of the image. TV minimization is a popular method for few-view tomography [10-15].

Extreme few-view tomography requires more information about the target image in addition to the piecewise-constant constraint. In the era of machine learning, a large amount of information can be learned from the images similar to the image to be reconstructed [16-22].

This paper assumes that similar images are not available. We must seek for other information. In transmission tomography, we roughly know the values of the attenuation coefficients for the materials in the objects being imaged. We use these known values as the constraints in this paper, as described in the next section.

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

There are many approaches to develop an image reconstruction algorithm. One approach is to set up an objective function, which typically contains a data fidelity term and one or more Bayesian terms. Each Bayesian term represents a constraint. An algorithm that minimizes this objective function minimizes all the terms simultaneously.

Another approach is the projections onto convex sets' (POCS) approach. In this approach, the main algorithm consists of two or more sub-algorithms. These sub-algorithms work separately and sequentially. Each of them has its own goals in mind. For a POCS algorithm, it is not easy to study its convergence. However, it is easy to fine tune each sub-algorithm independently and to adjust the balance between them. The POCS approach is adopted in this paper and is described in Fig. 1.

The POCS algorithm we used in this paper consists of three sub-algorithms. The first sub-algorithm takes care of image reconstruction. Any iterative image reconstruction algorithm can potentially be used to minimize the discrepancy between the forward projection of the reconstructed image and the line-integral measurements. In Fig. 1, the image reconstruction algorithm ① is chosen to be the well-known maximum-likelihood expectation-maximization (MLEM) algorithm [23]:

$$x_{i,j}^{(k+1)} = \frac{x_{i,j}^{(k)}}{\sum_m a_{i,j,m}} \sum_m a_{i,j,m} \frac{p_m}{\sum_{\hat{i},\hat{j}} a_{\hat{i},\hat{j},m} x_{\hat{i},\hat{j}}^{(k)}}, \qquad (1)$$

where p_m is the mth projection, $a_{i,j,m}$ is the projection contribution from the pixel (i,j) to the projection bin m, and k is the iteration index. In fact, the user can choose any justifiable iterative image reconstruction algorithm for algorithm ①. For example, a transmission EM algorithm [24] or a least square minimization algorithm [5].

The second sub-algorithm is a gradient descent algorithm to minimize the TV norm of the reconstructed image. The gradient descent algorithm ② in Fig. 1 is given as

$$x_{i,j}^{(k+1)} = x_{i,j}^{(k)} - \eta \frac{\partial TV(X)}{\partial x_{i,j}}, \qquad (2)$$

where $\eta = 2 \times 10^{-7}$ in our computer simulations, and $\frac{\partial TV(X)}{\partial x_{i,j}}$ is subdifferential of the TV norm of the current reconstructed image X. We use an extremely small step size η to ensure the stability of the algorithm. At the same time, we repeat this step 5000 times to guarantee the TV norm is

effective. The TV norm can be defined as

$$TV(X) = \sum_i \sum_j \sqrt{\left(x_{i+1,j} - x_{i,j}\right)^2 + (x_{i,j+1} - x_{i,j})^2}\,. \quad (3)$$

One can combine algorithm ① and algorithm ② into one Bayesian algorithm [11].

The third sub-algorithm is used to enforce the reconstructed image pixels to take the pre-specified values. The sub-algorithm ③ is the new attempt in this paper. It simply moves its image pixel value to its closest default image values. For torso imaging, the default values can be set up as the linear attenuation coefficients of the air, soft tissues, and bones. In fact, this sub-algorithm is nothing but segmentation. Notice that this step is skipped for most iterations, as enumerated by the variable 'Count.' We only activate this step every 100 counts, as dictated by remainder function 'mod' in mod(Count,100) = 0, which is the reminder of Count/100. We only know the approximate potential values in the image. We must downplay these 'known values' constraint and give the overall POCS algorithm a chance to converge to the true values that may not be the same as our 'set values.' Therefore, it is important not to terminate the POCS with the sub-algorithm ③.

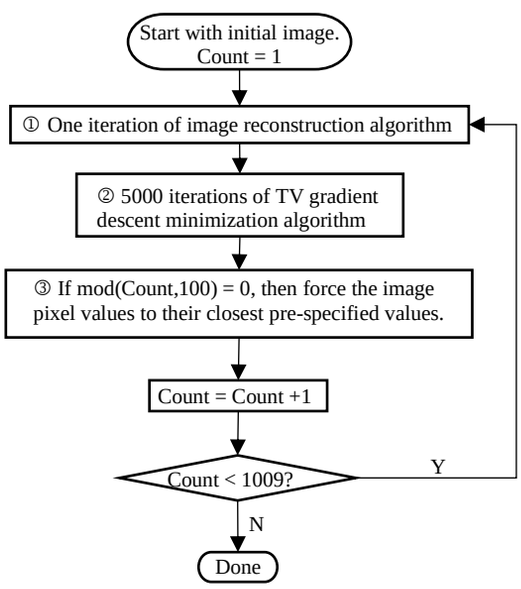


Figure 1. A flowchart of the algorithm used in computer simulations.

The computer simulations in this paper consider a two-dimensional (2D) parallel-beam imaging system, with an image array size of 256 × 256 pixels, the detector size of 256 bins, and 8 views (over 180°).

The projection line integrals were calculated analytically. The POCS algorithm used 1009 iterations used. Notice that 1009 is not a multiple of 100. This gives the pixel values in the reconstructed image a chance to move away from the segmented values set in the third sub-algorithm.

Both noiseless data and noisy data were used in the computer simulations. The noise was Gaussian distributed with a mean value of 0 and a variance of 5.

III. Results

Fig. 2 shows the true phantom used in computer simulations. The large disk has a value of 0.5. There are 8 small discs. Discs 1-5 have a value of 1.5; discs 6-8 have a value of 1.0.

In our implementation of the sub-algorithm ③, the potential pixel values were set at 0.51, 1.01, and 1.51. The corresponding pseudo code to update the pixel $x(i, j)$ is as follows.

When 'Count' is a multiple of 100, execute
if $(0.25 < x(i, j) \leq 0.75)$ then $x(i, j) = 0.51$;
if $(0.75 < x(i, j) \leq 1.25)$ $x(i, j) = 1.01$;
if $(x(i, j) > 1.25)$ $x(i, j) = 1.51$.

We do not force any pixel to a hard zero in the sub-algorithm ③, because the MLEM algorithm cannot update the pixel value zero.

Fig. 3 shows two MLEM reconstructions, one with noiseless data and the other one with noisy data, respectively. Here, the sub-algorithms ② and ③ are disabled in Fig. 1.

Fig. 2 shows two TV reconstructions, one with noiseless data and the other one with noisy data, respectively. Here, the sub-algorithm ③ is disabled in Fig. 1.

Fig. 2 shows two proposed POCS reconstructions, one with noiseless data and the other one with noisy data, respectively. All images are displayed in the gray-scale window of [0, 1.59].

The structure similarity (SSIM), peak signal-to-noise ratio (PSNR), and signal-to-noise ratio (SNR) are compared for the reconstructions. Table 1 compares the results using the noiseless data; Table 2 compares the results using the noisy data. It is shown that the rough knowledge used in sub-algorithm ③ is helpful in obtaining better reconstructions.

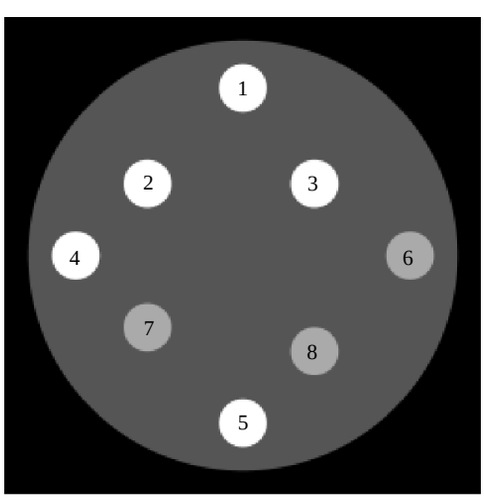
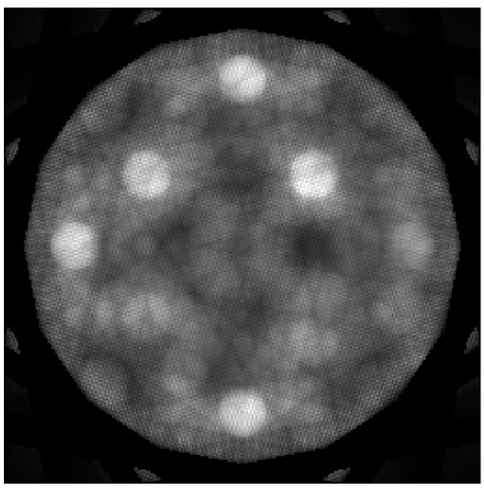


Fig. 2. The true phantom.



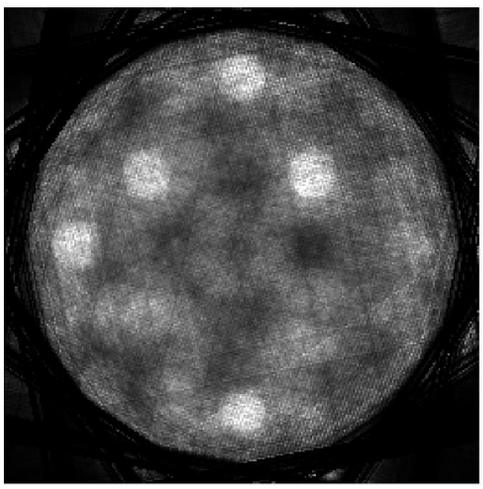


Fig. 3. MLEM reconstructions: (Upper) using noiseless data; (Lower) using noisy data.

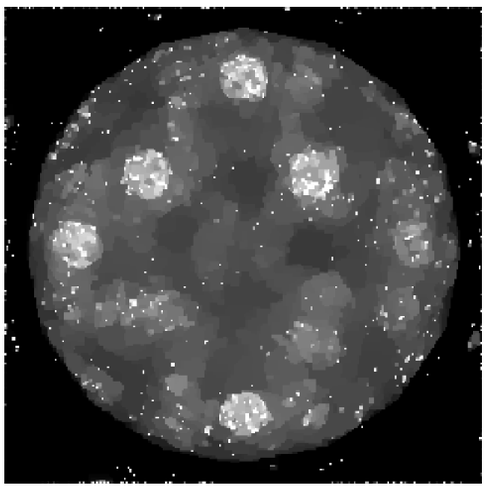
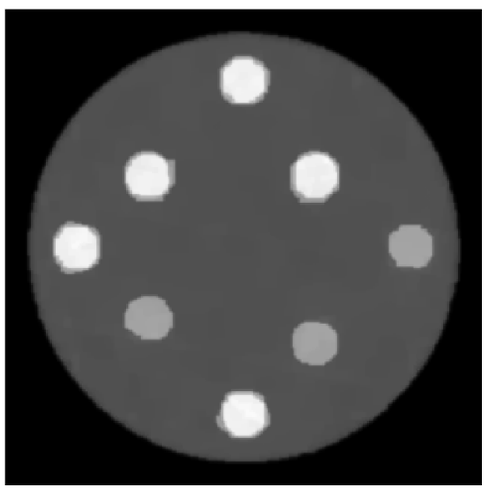
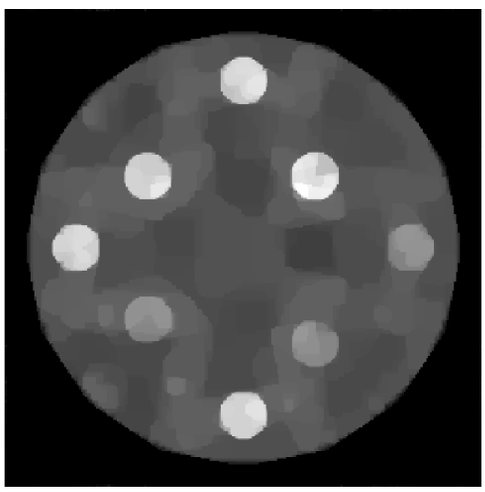


Fig. 4. TV reconstructions: (Upper) using noiseless data; (Lower) using noisy data.





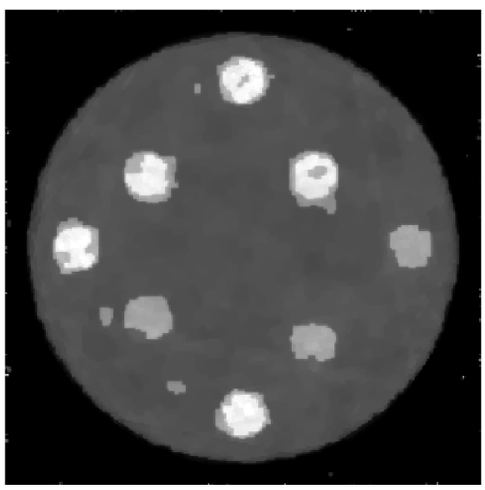


Fig. 5. Proposed POCS reconstructions: (Upper) using noiseless data; (Lower) using noisy data.

Table 1. Comparison studies with noiseless data

Method	SSIM	PSNR	SNR
Ideal case	1.00	∞	∞
MLEM w/o noise	0.4354	17.1834	11.2257
TV w/o noise	0.8692	23.1472	17.1895
Proposed POCS w/o noise	0.9472	26.4425	20.4847

Table 2. Comparison studies with noisy data

Method	SSIM	PSNR	SNR
Ideal case	1.00	∞	∞
MLEM w/ noise	0.2545	15.9300	9.9723
TV w/ noise	0.5058	12.7521	6.7943
Proposed POCS w/ noise	0.8594	22.7771	16.8194

IV. Discussion and Conclusions

One may ask; "What is the objective function of the sub-algorithm ③?" We do not need one. If one insists to have one, we can set something up as

$$F(X) = \sum_i \sum_j \left[x_{i,j}\left(x_{i,j} - 0.51\right)\left(x_{i,j} - 1.01\right)\left(x_{i,j} - 1.51\right)\right]^2. \quad (4)$$

However, we do not suggest a gradient based algorithm to minimize (4).

In our proposed POCS algorithm, we do not simply alternate between the sub-algorithms sequentially. Within each POCS iteration, we execute sub-algorithm ① once, sub-algorithm ② 5000 times, and sub-algorithm ③ 1/100 times.

We run sub-algorithm ② 5000 times because we choose to use a very small step size in a gradient descent algorithm to minimize the TV norm. A larger step size produces worse images. A very small step size makes almost no effect on the reconstructed image. To overcome this difficulty, we use a very small step size and a large iteration number for the sub-algorithm ②.

As the sub-algorithm ③, the number of iterations is 1/100. In other words, this sub-algorithm is skipped 99 times for every 100 global POSC iterations. This is equivalent to re-setting the initial image every 100 global POCS iterations.

One may argue that it is wrong to use the emission MLEM algorithm when the noise is Gaussian. The emission MLEM algorithm was original derived for the Poisson noise. It is more proper to use a transmission EM algorithm or a least square error minimization algorithm for the sub-algorithm ①. Our attempt of using an emission MLEM algorithm is to demonstrate that the TV minimization makes the noise model less important. The Bayesian information dominates the noise model in the maximum likelihood.

When the measurements are incomplete, any prior information and corresponding constraints will help. The piecewise feature of the objects makes the TV norm minimization effective. The rough knowledge of the image values, as demonstrated in this paper, is also effective. We believe that there are other features of the images can be used to supplementing the incomplete data. Machine learning turns out to be an effective way to explore the common features for a group of similar images.

References

[1] K. Champley, *Few-view image reconstruction*. U.S. Patent 10,282,869, 2019.

[2] H. Kim, R. Anirudh, K. A. Mohan, and K. Champley, "Extreme few-view CT reconstruction using deep inference," *Deep Inverse NeurIPS 2019 Workshop*, arXiv, https://arxiv.org/abs/1910.05375, 2019.

[3] K. Champley, J. Smith, J. Kallman, and P. Top, "Automatic threat detection for a dual-energy four-view X-ray carryon luggage scanner," *The 4th Int'l Conference on Image Formation in X-Ray Computed Tomography*, https://www.ct-meeting.org/?p=meeting_2016, pp. 213-216, 2016.

[4] G-H Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): A method to accurately reconstruct dynamic CT images from highly under-sampled projection data sets," *Medical Physics*, vol. 35, pp. 660–663, 2008.

[5] M. C. Pérez, V. E. Vidal, G. J. Verdi, and G. Quintana-Ortí, "Few-view CT image reconstruction via least-squares methods: Assessment and optimization." *Nuclear Science and Engineering*, 2023.

[6] Y. Mao and G. L. Zeng, "Tailored ML-EM algorithm for reconstruction of truncated projection data using few view angles." *Phys. Med. Biol.*, vol. 58, pp. N157-N169, 2013.

[7] G. L. Zeng, "On few-view tomography and staircase artifacts." *IEEE Trans. Nucl. Sci*., vol. 62, no. 3, pp. 851-858, 2015.

[8] G. L. Zeng, "Sparse-view tomography via displacement function interpolation." *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 13, 2019.

[9] G. L. Zeng, "Directly filtering the sparse-view CT images by BM3D." *SL Clinical Medicine: Research*, vol. 5(1), 126, 2022.

[10] K. Li, Z. Sang, X. Zhang, M. Zhang, C. Jiang, Q. Zhang, Y. Ge, D. Liang, Y. Yang, X. Liu, H. Zheng, and Z. Hu, "Few-view CT image reconstruction using improved total variation regularization." *J. Xray Sci. Technol*., vol. 27(4), pp. 739-753, 2019. doi: 10.3233/XST-190506. PMID: 31227684.

[11] V. Y. Panin, G. L. Zeng, and G. T. Gullberg, "Total variation regulated EM algorithm." *IEEE Trans. Nucl. Sci*., vol. 46, no. 6, pp. 2202-2210, 1999.

[12] E. Y. Sidky and X. Pan. "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization." *Physics in Medicine & Biology*, vol. 53, pp. 4777 – 4807, 2008.

[13] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1–4, pp. 259–68, 1992.

[14] Y. Lu, J. Zhao, and G. Wang, "Few-view image reconstruction with dual dictionaries." *Physics in Medicine & Biology*, vol. 57(1), p173, 2011.

[15] Y. Liu, Z. Liang, J. Ma, H. Lu, K. Wang, H. Zhang, and W. Moore, "Total variation-stokes strategy for sparse-view X-ray CT image reconstruction." *IEEE Trans. Med. Imaging*, vol. 33(3), pp. 749-763, 2014.

[16] V. Y. Panin, G. L. Zeng, and G. T. Gullberg, "Reconstructions of truncated projections using an optimal basis expansion derived from the cross correlation of a "knowledge set" of a priori cross sections." *IEEE Trans. Nucl. Sci.*, vol. 45, no. 4, pp. 2119-2125, 1998.

[17] Y. Han and J. C. Ye, "Framing U-net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag*., vol. 37, pp. 1418–1429, 2018.

[18] L. Hoyeon, L. Jongha, K. Hyeongseok, C. Byungchul, and C. Seungryong, "Deep-neural-network-based sinogram synthesis for sparse-view CT image reconstruction." *IEEE Trans. Radiat. Plasma Med. Sci*., vol. 3, pp. 109–119, 2018.

[19] S. Xie, X. Zheng, Y. Chen, L. Xie, J. Liu, Y. Zhang, J. Yan, H. Zhu, and Y. Hu, "Artifact removal using improved GoogLeNet for sparse-view CT reconstruction." Sci. Rep., vol. 8, no. 6700, 2018. doi: 10.1038/s41598-018-25153-w.

[20] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, "A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution." *IEEE Trans. Med. Imaging*, vol. 37, pp. 1407–1417, 2018.

[21] M. Lee, H. Kim, and H. J. Kim, "Sparse-view CT reconstruction based on multi-level wavelet convolution neural network." *Phys. Med*., vol. 80, pp. 352–362, 2020. doi: 10.1016/j.ejmp.2020.11.021.

[22] K. Liang, Y. Xing, H. Yang, and K. Kang, "Improve angular resolution for sparse-view CT with residual convolutional neural network." *Proceedings of the Medical Imaging 2018: Physics of Medical Imaging*; Houston, TX, USA. 12–15 February 2018; p. 105731K.

[23] L. A. Shepp and Y. Vardi, "Maximum likelihood reconstruction for emission tomography." *IEEE Trans Med Imag*., vol. 1, pp. 113–122, 1982.

[24] G. L. Zeng, "Emission expectation-maximization look-alike algorithms for x-ray CT and other applications." *Medical Physics*, vol. 45, pp. 3721-3727, 2018.

Deep Bone Extraction in X-Ray Projection Domain

Jonathan Bollig, Fabian Jäger, and Marc Kachelrieß

***Abstract*—Background: Bones are a defining structure in X-ray images and reconstructed CT volumes. Having accurate knowledge of their contribution to the X-ray attenuation and their position in the volume is of great importance for many applications. Most bone segmentation approaches in the context of CT rely on volumetric information to perform the segmentation.**

Purpose: Removing the need for a three-dimensional CT volume for the bone extraction can greatly accelerate computation times for applications that would normally rely on an initial backprojection. Furthermore, it allows us to deal with purely two-dimensional problems for which volumetric information is not obtainable.

Methods: We developed and trained the deep bone extraction (DBE), a convolutional neural network with UNet architecture to perform the bone extraction in two-dimensional radiographs. DBE outputs the contribution made to the total attenuation by bone and soft tissue separately for each detector pixel.

Results: Different network architectures and approaches to the training process were investigated and the results compared for both the two-dimensional X-ray projections and three-dimensional CT volume.

Conclusion: DBE is able to perform a bone extraction on the two-dimensional radiographs, which can be used to reconstruct CT volumes containing only bone or soft tissue respectively.

***Index Terms*—Cone-Beam CT (CBCT), Radiography, X-Ray Projections, Bone Segmentation, Deep Learning.**

I. Introduction

BONES play a dominant role in diagnostics, reconstruction tasks, and data processing when considering X-ray and CT data across a wide variety of applications. It is therefore often of great interest, to accurately determine their position in an x-ray image or in a reconstructed CT volume. In the scope of this study we distinguish between the two-dimensional projection domain of the X-ray projections and the three-dimensional image, or volume, domain of the reconstructed CT volume. While numerous approaches exist for bone segmentation tasks and related applications in the image domain [1], [2], [3], this may not be adequate for some applications. Imaging techniques like fluoroscopy can only be treated in projection domain, and other tasks may not allow for the extra computation time needed to perform a reconstruction prior to the segmentation. Examples for this include higher-order beam hardening corrections and registration tasks. Being able to perform the segmentation in projection domain can make a significant difference in computation time and feasibility of such tasks in clinical practice [4].

This study investigates a deep learning-based method for performing the bone extraction task, deep bone extraction (DBE), in projection domain. While this is a classic example of a binary segmentation task when considered in image domain, it is instead a two-label prediction problem when considered in projection domain. The pixel value in a two-dimensional projection is given by the log attenuation $q(L)$ detected at that pixel of the detector, given by

$$q(L) = -\ln \int dE\, w(E, L)\, e^{-\int d\lambda\, \mu(E, \boldsymbol{s}(L) + \lambda \Theta(L))} \,. \tag{1}$$

Here, L is a symbolic quantity referencing a beam starting at the focal position $\boldsymbol{s}(L)$ of the X-ray source traveling along the direction of $\Theta(L)$ parameterized by the continuous parameter λ. $\mu(E, \boldsymbol{r})$ is the linear attenuation coefficient of the measured volume for photons at position $\boldsymbol{r}$ and energy E. The normalized spectrum of the X-ray source is given by $w(E, L)$. It includes the directional dependence of the spectrum with L. The quantity in this process that directly relates to the object of interest is $\mu(E, \boldsymbol{r})$. It is the spatial distribution of this quantity that is reconstructed in CT volumes and more generally enables us to draw conclusions about the spatial material composition of the object. We will limit our considerations to the monochromatic case at energy E_0. In that case, we can write $w(E, L) = \delta(E - E_0)$. Then (1) leads to the new quantity

$$p(L) = \int d\lambda\, f(\boldsymbol{s}(L) + \lambda \Theta(L)) \,.$$

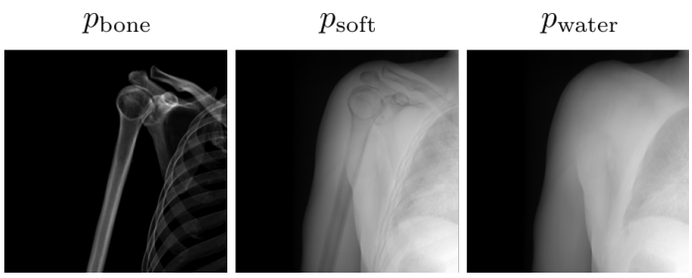


Fig. 1. The line integrals in bone p_{bone} (left), soft tissue p_{soft} (middle) and soft tissue with bones replaced by water at 70 keV p_{water} (right).

The issue at the heart of DBE in projection domain is that both $q(L)$ and $p(L)$ for any given beam L are composed of contributions from different materials encountered along the beam direction. The principle goal of DBE is to decompose these quantities into the contributions made by bone and those made by other materials. These other materials are henceforth collected under the term *soft tissue*. In this framework, we can write

$$p(L) = p_{\text{bone}}(L) + p_{\text{soft}}(L) \,,$$

Jonathan Bollig, Fabian Jäger, and Prof. Dr. Marc Kachelrieß are with the German Cancer Research Center (DKFZ), Heidelberg, Germany, and with Ruprecht-Karls-University, Heidelberg, Germany. Corresponding author: jonathan.bollig@dkfz.de.

where $p_{\text{bone}}(L)$ and $p_{\text{soft}}(L)$ describe the line integral along the beam L accounting only for the bone and soft tissue respectively.

The goal of the DBE is to train a neural network to predict $p_{\text{bone}}(L)$ and $p_{\text{soft}}(L)$ from $p(L)$ for each L in a two-dimensional projection. The network predictions will henceforth be denoted by $\hat{y}_{\text{bone}}$ and $\hat{y}_{\text{soft}}$, respectively.

II. Data and Methods

A. Data Generation

For the training of the neural network we used the following projections:

- Input: X-ray projections of the entire anatomical volume. This is the *full projection*.
- Labels: The corresponding set of two X-ray projections of only the bones or the soft tissue in the anatomical volume respectively. These are the *bone only projection* y_{bone} and *soft tissue only projection* y_{soft}.

As data we used full-body CT scans. Generating the training data from this involves a segmentation step and a forward projection.

1) Segmentation: We used a binary segmentation of the CT volumes with a threshold of 200 HU. To limit false positive classifications due to noise, we only considered bone candidate voxels if they have at least one neighbouring voxel that also fulfills the thresholding criterion. This segmentation yields a bone volume and a soft tissue volume along with the original full volume.

2) Forward Projection: We simulated a CBCT scan of the full volume, the bone volume, and the soft tissue volume by calculating projections from different angles. This effectively corresponds to calculating the line integrals $p_{\text{bone}}(L)$ and $p_{\text{soft}}(L)$ where L describes the beams from the source to the individual detector pixels. This approach effectively substitutes the bones with air in the soft tissue volume. Another possibility is to substitute them with water, specifically to set the linear attenuation coefficient in those regions to $\mu_{\text{water}}(70\ \text{keV}) = 0.0192/\text{mm}$. Forward projecting the resulting soft tissue volume leads to less sharp edges in the volume and may make it easier for the network to learn.

B. Training Data

We used nine full-body CT scans for the data generation in a 7/1/1 split for training, validation, and testing. For the forward projection step we picked isocenters for the simulation so as to avoid artefacts, e.g. such caused by implants. Furthermore, we ensured that the resulting projections could unambiguously be categorized in one of the anatomical categories thorax, abdomen, and legs. Overall, 3600 projections were used for training, 648 for validation, and 684 for testing with 36 projections for each isocenter, i.e. with an angular difference of 10° between neighbouring projections. The entirety of these data was then simulated again after replacing the bone with water at 70 keV instead of air in the soft tissue volume, yielding the line integrals p_{water}. Figure 1 shows an example for the three line integrals considered. The training itself was done on patches with 256×256 pixels.

C. The Network

In this study we used a CNN with UNet architecture [5]. It consists of an encoding and a decoding branch with skip connections between the two. Both branches consist of six blocks of two convolutional layers, each followed by a rectified linear unit, linked by pooling or upsampling layers respectively. We used the Adam optimizer [6] and initialized the parameters with the uniform Xavier initialization [7]. We implemented the network as a 2D and 2.5D UNet. In case of the 2D UNet we simply used one projection as an input, for the 2.5D UNet we used additional projections from the same isocenter as further input channels. Because the 2.5D architecture requires multiple projections, it is not applicable to single X-ray scans. The number of network parameters is approximately the same for the 2D and 2.5D architecture. The idea of the 2.5D UNet is to make use of the three-dimensional spatial information contained in multiple projections while keeping the computational effort comparable to the 2D UNet by using 2D convolutional layers [8]. The 2.5D UNet was also trained for different configurations of the input projections relative to the projection of interest.

D. Loss Functions

As a first approach, we compare predictions and labels using the MSE as given by

$$L_0 = \frac{1}{\#y}(\|\hat{y}_{\text{bone}} - y_{\text{bone}}\|_2^2 + \|\hat{y}_{\text{soft}} - y_{\text{soft}}\|_2^2) \ .$$

Here, $\#y$ denotes the number of pixels in each projection. However, we can exploit the fact that the sum of the labels is equal to our input. One way to use this is by only predicting one of the two labels in contrast to the multitask-training of predicting both projections. Alternatively, we can implement this by enforcing equality of the sum of predictions and the input in the loss function. The corresponding loss term is

$$L_1 = \frac{1}{\#y}\|(\hat{y}_{\text{bone}} + \hat{y}_{\text{soft}}) - (y_{\text{bone}} + y_{\text{soft}})\|_2^2 \ .$$

We combine the two loss functions in a linear combination with $L = L_0 + \lambda L_1$, initially setting $\lambda = 0$.

E. Training Aspects

The division of training, validation, and test set into subsets corresponding to anatomical regions is an essential part of our training and testing process. We always train and validate the network on data from one or all of the three specified regions and evaluate the trained network on the three subsets individually.

While our network is trained in projection domain, it is interesting for a number of reasons to evaluate the results in image domain as well. For this, we used additional projection sets that extend the used test sets to 720 projections per isocenter and reconstruct the network predictions for these sets. This allowed us to evaluate the quality of our DBE in image domain by comparing the results with the original bone only and soft tissue only volumes. In this study, we compared the two volumes using the MSE and the binary segmentation maps calculated from these volumes using the dice coefficient.

III. Results

While a main point of interest of this study is the final performance of the DBE, another essential aspect is the comparison of different decisions that can be made in the network design and training process. While the total loss and the loss on the individual projections are all of interest, we will focus only on the loss on the bone only projection in the following section. The results reported below are based on predictions for 720 projections for one isocenter for the three regions each.

A. Network Architecture and Training Process

1) Network Dimensionality: When trained on data from only one anatomical region, neither of the network dimensions unambiguously showed a better performance than the others. However, using data from all three anatomical regions the 2.5D architecture performed better than the others across test sets from all regions. It further outperformed all networks trained specifically on data from only one region independent of their dimensionality. For the 2.5D UNet we further compared different configurations of the input projections. The best performance among those variations was reached for two additional projections with an angular distance of $\pm 10°$ relative to the projection of interest.

2) Replacing Bone with Water: The effect of replacing bone with standard water instead of air had an ambiguous effect on the network's performance. When trained on data from only one anatomical regions, none of the two options was categorically superior to the other. However, when trained on data from all regions, replacing bone with air showed a better performance in all cases. For the test regions thorax and legs, the 2.5D UNet trained on data from all regions with bone replaced with air remained unmatched in performance. For testing on data from the abdomen, it was surpassed by the 2D network trained only on data from the abdomen while replacing bone with water. This result is surprising in that both network dimensionality and the choice between p_{soft} and p_{water} are in contrast to the otherwise observed trend. However, this may be explained by the anatomical properties of these regions. In the thorax the lungs as a large part of the patient volume are already filled with air. Creating additional air-to-soft tissue transitions by replacing bone with air does not have as significant an effect as it does in the abdomen, where there is much less air contained in the patient volume. In case of the legs, the overall anatomical structure is much more trivial, which allows most networks to learn the bone extraction very well. This also results in good predictions for leg projections even if a network was not trained on data from this region.

3) Multitask Learning: Multitask learning did not have a clear advantage or disadvantage over single task learning across all networks.

4) Loss Functions: Different parameter configurations for the loss function were investigated, but none produced results superior to those of the simple loss function L_0.

5) Results in Projection and Image Domain: In principle, one would expect a monotonous relationship between the error of the projections and the error of the volume reconstructed from those projections. However, an initial overview of our results does not support that. Much more dominant in determining the error of the reconstructions is the anatomical region considered. Whereas the CT volume is largely empty for the legs, a much larger portion of it is filled in case of the thorax or abdomen. In a first approach to correct for this, we can consider only the relationship between the errors calculated within a specific test region. Here, this effect should be much less significant and we did observe a monotonous relationship between the errors across all predictions. Another approach would be to only calculate the loss in semantically relevant regions of the projections and CT volumes, but we have not yet implemented this approach. The first approach however already suggests that improving the quality of the projections translates directly to an improvement in the quality of the reconstructed volumes.

B. Performance of the DBE

Figure 2 shows an example for the network input, labels, predictions, and the differences between labels and predictions. The MSE results used throughout the study are highly relevant

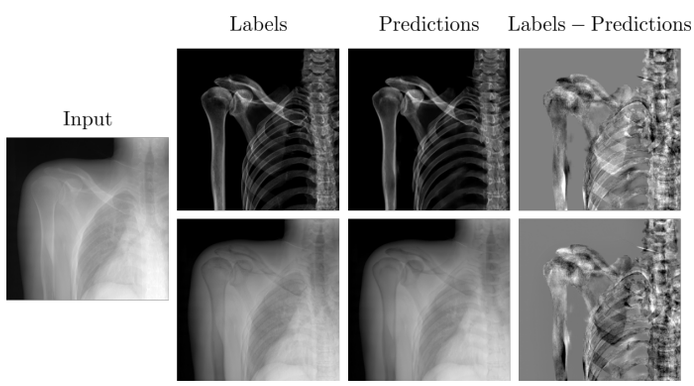


Fig. 2. An example for the input projection (C = 2.3, W = 4.7), the labels, network predictions, and difference between labels and predictions (C = 0.0, W = 0.5), from left to right respectively. The top row shows the bone only projections (C = 0.7, W = 1.4), the bottom row the soft tissue only projections (C = 2.3, W =4.6).

for many applications of the DBE. However, they are of little value for comparison with results outside of this study. Therefore, we include the results of a bone segmentation based on the DBE results using p_{soft} as a soft tissue label by applying a thresholding to the bone label and prediction. These are given in the form of the dice coefficient in table I. As mentioned

TABLE I
Dice Coefficients for a Bone Segmentation in Projection Domain Based on the DBE Results.

Test \ Train	All	Thorax	Abdomen	Legs	Architecture
Thorax	0.946	0.948	0.845	0.840	
Abdomen	0.916	0.886	0.905	0.685	2D
Legs	0.998	0.959	0.993	0.961	
Thorax	**0.950**	0.947	0.873	0.786	
Abdomen	**0.926**	0.878	0.890	0.684	2.5D
Legs	**0.998**	0.980	0.991	0.962	

above, the best performance for each test region is reached by the 2.5D UNet trained on data from all regions (marked bold

in table I). For the networks trained only on data from one region, the best performance for each test region is reached by the network trained only on data from that region. This is an expected result, as the network will focus on learning the features specific to the training region.

Finally, we are interested in the quality of the reconstructed volumes based on the predicted projections. We report these results in table II using the dice coefficient after converting both the ground truth and the reconstruction of our results to a binary map via thresholding. Here, the best results for each test region are not all reached by the 2.5D UNet trained on data from all regions, but the values for this network are very close to the best performance we observed (marked bold in table II). As these results and visual investigation of the results show, the reconstructed volumes show a reasonable similarity to the ground truth. Even given the high quality of the results in projection domain, this is not inherently a necessary consequence. A priori, the errors in projection domain are not correlated in any such way as to support a meaningful reconstruction. Potentially, the errors of the projections may just be small enough so as to limit the reconstruction errors to the observed range. Another hypothesis yet to be investigated is that the error distribution across the projections supports the reconstruction. The largest errors by far were observed for regions where the bone label shows bone tissue. This may be conducive to an accurate reconstruction.

TABLE II
DICE COEFFICIENTS FOR A BONE SEGMENTATION IN IMAGE DOMAIN BASED ON THE DBE RESULTS.

Test \ Train	All	Thorax	Abdomen	Legs	Architecture
Thorax	0.617	**0.625**	0.350	0.386	
Abdomen	0.586	0.563	**0.629**	0.391	2D
Legs	0.974	0.937	0.923	0.975	
Thorax	0.616	0.500	0.305	0.348	
Abdomen	0.620	0.463	0.503	0.386	2.5D
Legs	**0.977**	0.783	0.876	0.909	

IV. DISCUSSION

While we were able to draw some important conclusions about the network architecture and training process from our data, some ambiguities remain in our results. Especially for the networks trained only on data from one anatomical region, a clear advantage of one design decision over the others could not be seen. To further confirm this, testing on larger amounts of data is an important next step. Furthermore, the relationship between the results in projection domain and the quality of the reconstructed results is an interesting question to be investigated. This may include analyzing the structure of the residual error of the network's predictions, the resulting correlation between errors in different projections, and how these translate to errors in the reconstructed volumes both mathematically and empirically.

V. CONCLUSIONS

DBE allows for a bone extraction without need for volumetric information. It was also shown that the results of DBE in projection domain allow for the reconstruction of sensible bone and soft tissue only volumes. This means that the DBE in projection domain not only provides a great foundation for computation-efficient higher-order beam hardening corrections and other two-dimensional applications, but also a novel approach to binary CT volume segmentation. The effect of a number of different choices in the training process were evaluated. This demonstrated some important properties of the training data that influence the performance of the networks such as the air, bone, and soft tissue distribution in the scans.

ACKNOWLEDGEMENT

The scans are from cadaver measurements and have been acquired in scope of a forensic study in close collaboration with the Institute of Forensic and Traffic Medicine (Prof. Dr. Sarah Heinze), Heidelberg University, Heidelberg, Germany, after being approved by the local ethics review board (S 388/2014). The projection simulation software RayConStruct-PS and the image reconstruction software RayConStruct-IR were provided by RayConStruct® GmbH, Nürnberg, Germany.

REFERENCES

[1] K. A, W. J, H. J, and M.-H. KH, "Automatic bone segmentation in whole-body ct images," *Int J Comput Assist Radiol Surg.*, vol. 14, no. 1, pp. 21–29, 2019.

[2] M. van Eijnatten, R. van Dijk, J. Dobbe, G. Streekstra, J. Koivisto, and J. Wolff, "Ct image segmentation methods for bone used in medical additive manufacturing," *Medical Engineering & Physics*, vol. 51, pp. 6–16, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1350453317302631

[3] J. Zhang, C.-H. Yan, C.-K. Chui, and S.-H. Ong, "Fast segmentation of bone in ct images using 3d adaptive thresholding," *Computers in Biology and Medicine*, vol. 40, no. 2, pp. 231–236, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482509002157

[4] P. Després and X. Jia, "A review of gpu-based medical image reconstruction," *Physica Medica*, vol. 42, pp. 76–92, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1120179717302417

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *CoRR*, vol. abs/1505.04597, 2015. [Online]. Available: http://arxiv.org/abs/1505.04597

[6] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2017.

[7] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *Journal of Machine Learning Research - Proceedings Track*, vol. 9, pp. 249–256, 01 2010.

[8] C. Angermann, M. Haltmeier, R. Steiger, S. P. J. au2, and E. Gizewski, "Projection-based 2.5d u-net architecture for fast volumetric segmentation," 2019.

The Impact of Arterial Input Function Selection on Dynamic Liver Perfusion Imaging with Cone-Beam CT

Hana Haseljić, Robert Frysch, Vojtěch Kulvait, Daniel Punzet, Thomas Werncke, Oliver Speck, and Georg Rose

***Abstract*—For the treatment of the liver cancer, the perfusion maps have to be accurate and therefore the arterial input function, needed for their calculation, correctly located. Due to reconstruction artifacts and truncation in cone-beam CT, selecting the same arterial input function as in prior CT volume might not be feasible. Consequently the comparison of the perfusion maps between the cone-beam CT and CT is not easy, and sometimes impossible. This work was concentrated to determine whether the inability to detect the adequate arterial input function in both, cone-beam CT and CT, would affect the accuracy of perfusion maps and to which extent. In reconstructed post embolization CT liver perfusion scan, multiple candidates from different volumes' slices for arterial input function were selected and the one with the shortest first moment and with the highest peak value was selected as the referent one for comparison. The same procedure was performed in simulated cone-beam CT scan. The Pearson correlation coefficients indicated that the CT perfusion maps differed strongly depending on the arterial input function selected and that the arterial input function with peak closer in time to peak of the referent one would result in higher correlation coefficient for blood flow and blood volume. The further in time the peaks of two functions are, the lower the correlation for the mean transit time was ($r < 0.8$). Similar behaviour could be observed for cone-beam CT perfusion maps but with much lower correlation coefficient for mean transit time (< 0.6). The findings suggest that the hypoperfused area - tissue after the embolization procedure, remained relatively consistent across all perfusion maps. However, the overall accuracy of perfusion maps and the values for perfusion parameters were very influenced by the selected arterial input function.**

***Index Terms*—arterial input function, cone-beam computed tomography, liver perfusion imaging**

I. Introduction

THE compute tomography (CT) liver perfusion imaging is a promising modality for tumour diagnosis, treatment planning, posttherapeutic monitoring of possible cancer recurrences and treatment response and to assist the minimally invasive procedures [1], [2]. The minimally invasive treatment would furthermore benefit from the on-site success evaluation using cone-beam CT (CBCT) without transferring the patient [3] out of the operational room. To evaluate the accuracy of perfusion maps, they get compared to perfusion maps calculated by acquired pre-treatment or, in case of such experiments, post-treatment CT perfusion scans [4], [5], [6]. The small field of view, truncation and reconstruction artifacts, low contrast resolution [7] affect the possibility to correctly detect the best suitable arterial input function (AIF), which is required to calculate perfusion maps, and consequently the quality and accuracy of perfusion maps and the comparability to CT perfusion maps. To authors' knowledge, the AIF selection was discussed for brain and lungs CT perfusion imaging [8], [9]. The processing methods of selected AIF in liver CT perfusion were discussed in [10].

As the first step to determine to which extent the position of AIF in liver affects the CT perfusion maps, perfusion maps were calculated using different AIF candidates and compared to a referent perfusion maps. Then, the changes in CBCT perfusion maps related to selection of AIF were inspected in comparison to CT perfusion maps. The following sections are about the data used for this study, the procedure of AIF localization and the algorithm for perfusion maps calculation. Finally, the results are presented and discussed.

This work was by the German Federal Ministry of Education and Research within the Research Campus STIMULATE, grant no. 13GW0473A and 13GW0473B.

Hana Haseljić is with Institute for Medical Engineering and Research Campus STIMULATE, Otto-von-Guericke University, Magdeburg, Germany (e-mail: hana.haseljic@ovgu.de)

Robert Frysch, Daniel Punzet and Georg Rose are with Institute for Medical Engineering and Research Campus STIMULATE, Otto-von-Guericke University, Magdeburg, Germany

Oliver Speck is with Institute for Medical Engineering and Research Campus STIMULATE, Biomedical Magnetic Resonance, Otto-von-Guericke University, Magdeburg, Germany

Thomas Werncke is with Institute of Diagnostic and Interventional Radiology, Hannover Medical School, Hannover, Germany.

Vojtěch Kulvait is with Institute of Materials Physics, Helmholtz-Zentrum hereon, Geesthacht, Germany.

II. Methods

A. Animal Experiments

The hepatic artery of one pig[1] was embolized with tantalum-based embolization material and coils. Then, the CT perfusion scan of duration of $41.2\,\mathrm{s}$ was acquired with SOMATOM Force CT. The contrast material dose was $10.4\,\mathrm{ml}$ and the volume $14.9\,\mathrm{ml}$. This CT scan was used then to simulate the CBCT perfusion scan.

B. Simulation of CBCT scan

The CBCT perfusion scan should consist of ten sweeps, in each acquiring 248 views, and nine pause intervals between the consecutive sweeps [11], [4]. The time attenuation curves were

[1]The research was carried out in compliance with European Directive 2010/63/EU and German animal welfare legislation (TierSchG). All experiments were approved by the local animal ethic committee (Lower Saxony State Office for Consumer Protection and Food Safety, LAVES 18/2809).

interpolated and resampled so they satisfy these conditions and then reprojected as in [12]. The simulated scan was reconstructed by the means of straightforward method which assumes that all views within one sweep have been acquired in single time point as in [13] using [12]. For the CT scan of duration of $41.2\,\mathrm{s}$ one sweep is $3.1\,\mathrm{s}$ long.

C. Localization of Arterial Input Function

The region of interest for AIF should be placed in hepatic artery and portal vein, as both are blood inflows to liver [14]. Considering the local entry point of embolization and the differing visibility of vessels through the organ, multiple AIFs could have been selected, see Figure 1. In total three AIFs were selected, labeled AIF 1, AIF 2 and AIF 3. The AIF with the shortest first moment, highest peak value and early arrival time was selected as referent [15]. The other AIFs are selected based on the slices in which it was possible to locate the AIF in CBCT volumes as well.

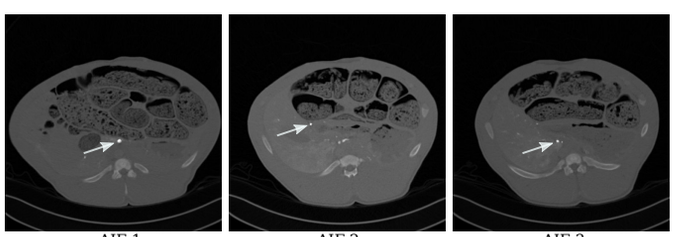


Fig. 1: Locations of selected arterial input functions.

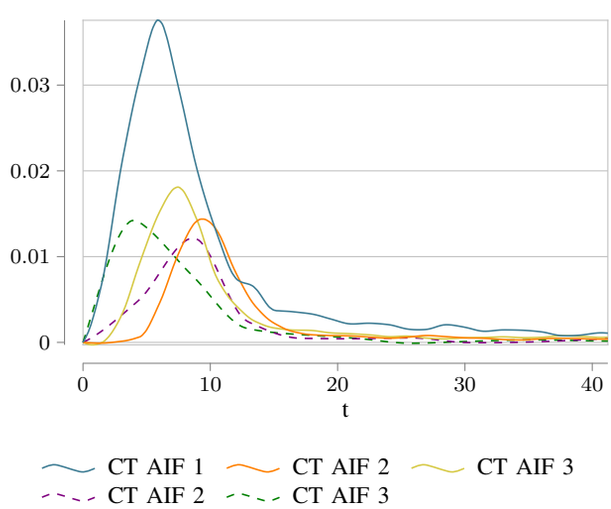


Fig. 2: Time attenuation curves of arterial input functions selected in CT and CBCT reconstructions. The referent CT AIF is in blue. The AIF 2 and AIF 3 are selected in the same regions in both, CT and CBCT reconstructions.

D. Calculation of Perfusion Parameters

The deconvolution approach was used to compute the perfusion maps [16]. The time attenuation curve of every voxel $tac(t)$ is modeled as a convolution of arterial input function $aif(t)$ with residual function $f_r(\mathrm{i})$. Hence, for the blood flow (BF), blood volume(BV) and mean transit time (MTT), see Eq. 1, the selection of AIF is so important.

$$BF = \max f_r(t), \quad BV = \sum_{i=1}^{n} f_r(\mathrm{i}), \quad MTT = \frac{BV}{BF} \tag{1}$$

The time to peak is the time from the acquisition start to the time point at which the maximum value of $tac(t)$ is reached. This means that for any comparison the correlation was 1 and was not to be the subject of further discussion. The given equations can be used for calculation of perfusion parameters considering the duration of the perfusion scan and the vessel being embolized.

For comparison of perfusion maps it is also advisable to observe the form of vessels which were segmented out from the perfusion maps and embolized regions. With the ASIST colormap, see [17], the dark blue in blood flow represents the hypoperfused region of the liver and red the highly perfused regions and, in this, case vessels. The perfusion maps using all selected candidates for AIFs are calculated and compared to perfusion maps calculated using the referent AIF by calculating the Pearson correlation coefficients. Also, the CBCT perfusion maps were compared to CT perfusion maps.

III. Results

In Table I the Pearson correlation coefficients between the referent CT perfusion maps and the CT perfusion maps generated using all other AIFs are given. In Table II the correlation coefficients between AIF 2 and 3 to corresponding and to the others AIFs are given. The AIF labels 1, 2, and 3 correspond to labels given in legends in Figure 2.

TABLE I: Pearson correlation for CT perfusion maps compared to CT perfusion maps calculated using the referent AIF. The highest values are highlighted in bold.

	AIF 2	AIF 3
Blood flow	0.9002	**0.9844**
Blood volume	0.9850	**0.9982**
Mean transit time	0.8104	**0.9578**

TABLE II: Pearson correlation coefficient for CBCT perfusion maps compared to CT perfusion maps calculated using the referent AIF 1 and corresponding AIF. The highest values are highlighted in bold.

		AIF 1	AIF 2	AIF 3
Blood flow	AIF 2	0.6892	**0.8866**	0.7134
	AIF 3	0.7919	**0.8612**	0.8336
Blood volume	AIF 2	0.8294	**0.8600**	0.8443
	AIF 3	0.8479	**0.8642**	0.8584
Mean transit time	AIF 2	0.5749	0.5120	**0.5921**
	AIF 3	**0.4656**	0.2279	0.4517

The mean and standard deviation values for perfusion parameters values for every AIF selected in CT and CBCT perfusion maps are given in Tables III and IV.

The CT and CBCT perfusion maps for different AIFs that could have been selected can be inspected in Figure 3.

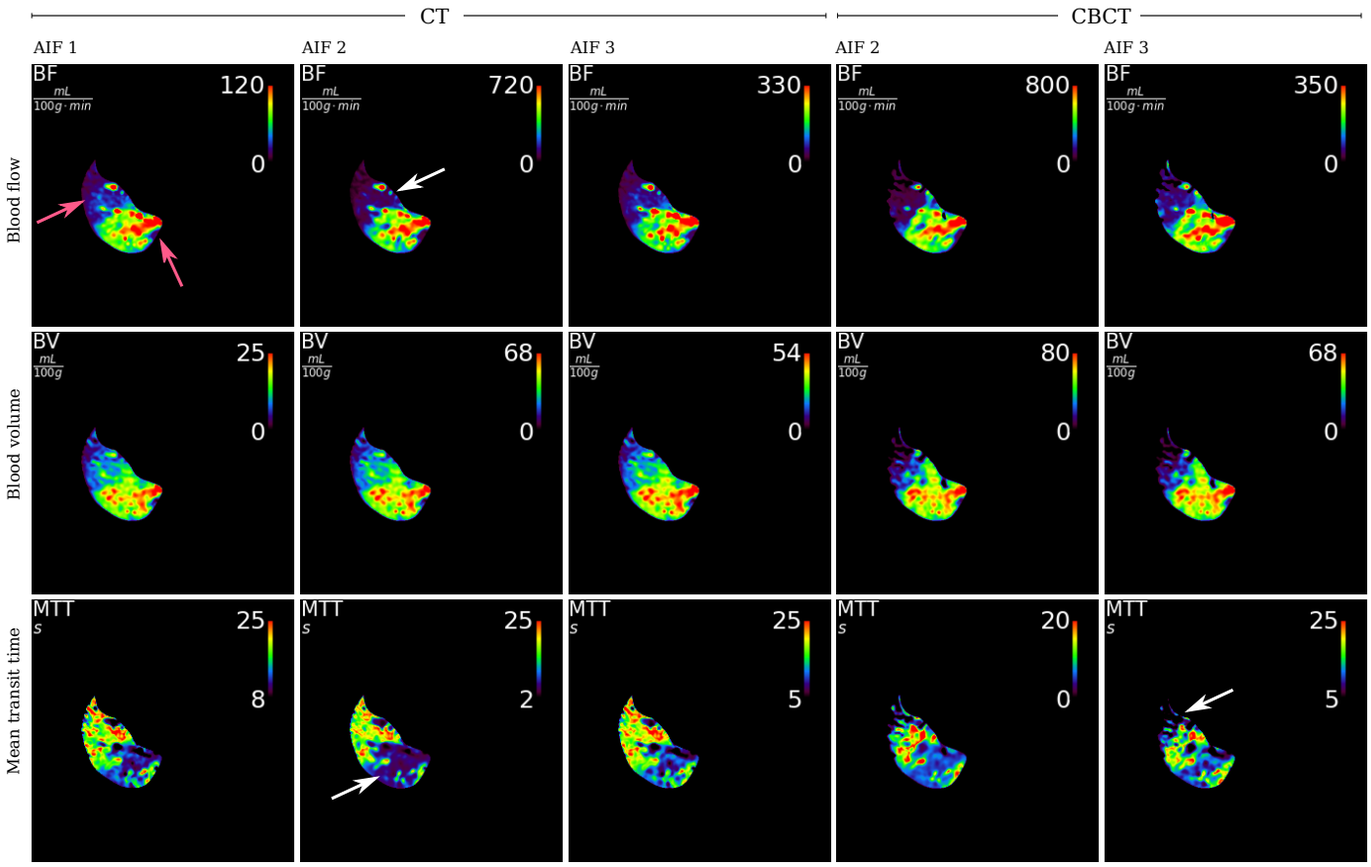


Fig. 3: CT and CBFCT perfusion maps calculated using selected arterial input functions. CT perfusion maps for AIF 1, 2 and 3 are shown in first three columns. In forth and fifth column the CBCT perfusion maps for AIF 2 and 3 are shown. The pink arrows are pointing to the hypoperfused regions. The white arrows are pointing to regions in which the perfusion was possibly inaccurately calculated.

TABLE III: Mean and standard deviation of the values for perfusion parameters in CT perfusion maps.

	Blood flow	Blood volume	Mean transit time
1	56.783 ± 52.209	11.934 ± 5.522	15.557 ± 5.019
2	274.085 ± 235.357	33.217 ± 14.837	10.988 ± 5.900
3	151.976 ± 156.880	27.215 ± 12.453	14.665 ± 5.546

TABLE IV: Mean and standard deviation of the values for perfusion parameters in CBCT perfusion maps.

	Blood flow	Blood volume	Mean transit time
2	262.571 ± 224.625	33.786 ± 20.320	9.625 ± 5.789
3	149.878 ± 135.328	28.406 ± 17.366	12.713 ± 6.398

IV. Discussion

The CT perfusion maps calculated using different AIFs all differ. The perfusion maps calculated using AIF 3, $r > 0.95$ for all perfusion parameters, are more similar to perfusion maps calculated using AIF 1 then those with AIF 2, $r < 0.82$ for MTT. This can be also observed in Figure 3, in the first three columns. For the AIF 2 shown in the second column, the white arrow is pointing to regions of underestimated perfusion. The pink arrow is pointing to hypoperfused region. In all CT perfusion maps the embolized *i.e.* hypoperfused regions are well distinguishable from the healthy tissue. It can be also noticed that the hypoperfused regions in perfusion maps calculated using AIF 2 is much more pronounced than in other perfusion maps.

In Figure 2, it is visible how AIF 1 and AIF 3 have closer peaks in time. This means that AIF 2, even though it is selected in the slice more close to the slice where AIF 1 was selected, was not located in the same vessel, see Figure 1. Unlike the AIF 1, which has a very high peak due to possibly higher contrast agent dose injected, the other two AIFs reach their peaks later in time and also much lower ones. This will obviously affect the values estimated for perfusion parameters resulting in inaccurate estimation.

In CBCT scan the AIF 1 couldn't be selected. Therefore, the CBCT perfusion maps calculated using AIF 2 and AIF 3 were compared to CT perfusion maps calculated using the referent AIF 1 and others as well. As expected, for the BF and BV, the Pearson correlation coefficients were higher when the perfusion maps of corresponding AIFs were compared then with the AIF 1, meaning CBCT and CT perfusion maps both calculated using AIF 2 or AIF 3. Surprisingly, the CBCT BF and BV with AIF 3 were more similar to CT BF and BV with AIF 2 than to AIF 3. For the mean transit time the correlation coefficient was higher when the CBCT perfusion maps were compared to CT perfusion maps calculated using

AIF 1 and, similarly to BF and BV previously discussed, for the comparison between CBCT MTT with AIF 2 and CT MTT with AIF 3. Also, the correlation is much lower for MTT for CBCT maps, $r < 0.6$, compared to CT perfusion maps, $r > 0.8$. In straightforward method for reconstruction of CBCT volumes, it is assumed that the contrast agent flow is constant during one sweep and therefore all the views acquired at the same time point and. Moreover, the AIF is estimated only from ten points, considering ten sweeps. This is what possibly causes the time in shift peaks for AIF 2 and AIF 3 estimated from CBCT volumes compared to corresponding estimated from CT volumes, see Figure 2, and causing the results in Table II. To deal with temporal undersampling the model based perfusion reconstruction method was proposed and analyzed in the liver in [6], [18] which also could incorporate the prior knowledge.

The values of perfusion parameters in CBCT perfusion maps are similar to those of CT perfusion maps, see Tables III and IV and Figure 3, columns 2-4. Here only the simulated CBCT scan has been used where all the slices in the reconstructed volumes correspond exactly to those in CT volumes. This facilitates the selection of the AIF in the same region and slice. For the real CBCT scan this might not be possible and could be as well affected by the possible different positioning of the subject scanned. This also explains why in previous research, such as [6], the values between perfusion parameters in real CBCT and CT scan varied and differed.

V. Conclusion

For the qualitative and quantitative evaluation of the cone-beam CT perfusion maps the usual approach would be to compare them with the CT perfusion maps. The calculation of perfusion parameters is dependent on the arterial input function selected. In this work, it was shown that the selection of the arterial input function affects not only the overall similarity between CBCT and CT perfusion maps but the values of perfusion parameters as well within the perfusion maps calculated from the same volumes but using different AIFs. Nevertheless, selecting the AIF, which is not necessarily the best one according the guidelines, can still be used to generate perfusion maps in which the valuable perfusion information would be preserved. In case of the embolization treatment, the hypoperfused regions have been clearly differentiated in all perfusion maps. The future work should be concentrated around the severity of the influence of selection of the AIF in the model-based perfusion reconstruction with taking into an account the expected values of perfusion parameters.

References

[1] H. Ogul, M. Kantarci, B. Genc, B. Pirimoglu, N. Cullu, Y. Kizrak, O. Yilmaz, and N. Karabulut, "Perfusion ct imaging of the liver: review of clinical applications," *Diagnostic and Interventional Radiology*, vol. 20, no. 5, p. 379–389, Aug. 2014. [Online]. Available: http://dx.doi.org/10.5152/dir.2014.13396

[2] M. Hatem Shalaby and K. A. Ali Shehata, "Ct perfusion in hepatocellular carcinoma: Is it reliable?" *The Egyptian Journal of Radiology and Nuclear Medicine*, vol. 48, no. 4, p. 791–798, Dec. 2017. [Online]. Available: http://dx.doi.org/10.1016/j.ejrnm.2017.07.013

[3] K. A. Kim, S. Y. Choi, M. U. Kim, S. Y. Baek, S. H. Park, K. Yoo, T. H. Kim, and H. Y. Kim, "The efficacy of cone-beam ct–based liver perfusion mapping to predict initial response of hepatocellular carcinoma to transarterial chemoembolization," *Journal of Vascular and Interventional Radiology*, vol. 30, no. 3, p. 358–369, Mar. 2019. [Online]. Available: http://dx.doi.org/10.1016/j.jvir.2018.10.002

[4] S. Datta, K. Müller, T. Moore, L. Molvin, S. Gehrisch, J. Rosenberg, Y. Saenz, M. Manhart, Y. Deuerling-Zheng, N. Kothary, and R. Fahrig, "Dynamic measurement of arterial liver perfusion with an interventional c-arm system," *Investigative Radiology*, vol. 52, no. 8, p. 456–461, Aug. 2017. [Online]. Available: http://dx.doi.org/10.1097/rli.0000000000000368

[5] R. L. O'Donohoe, R. G. Kavanagh, A. M. Cahalane, D. D. Houlihan, J. W. McCann, and E. R. Ryan, "C-arm cone-beam ct parenchymal blood volume imaging for transarterial chemoembolization of hepatocellular carcinoma: implications for treatment planning and response," *European Radiology Experimental*, vol. 3, no. 1, May 2019. [Online]. Available: http://dx.doi.org/10.1186/s41747-019-0099-0

[6] H. Haseljić, V. Kulvait, R. Frysch, B. Hensen, F. Wacker, G. Rose, and T. Werncke, "The Application of Time Separation Technique to Enhance C-arm CT Dynamic Liver Perfusion Imaging," in *Proceedings of the 16th Virtual International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2021, pp. 264 –267. [Online]. Available: https://arxiv.org/abs/2110.04143

[7] L. Lechuga and G. A. Weidlich, "Cone beam ct vs. fan beam ct: A comparison of image quality and dose delivered between two differing ct imaging modalities," *Cureus*, Sep. 2016. [Online]. Available: http://dx.doi.org/10.7759/cureus.778

[8] R. M. e. a. Ferreira, "Arterial input function placement for accurate ct perfusion map construction in acute stroke," *American Journal of Roentgenology*, vol. 194, no. 5, p. 1330–1336, May 2010. [Online]. Available: http://dx.doi.org/10.2214/AJR.09.2845

[9] L. Jimenez-Juan *et al.*, "Arterial input function placement effect on computed tomography lung perfusion maps," *Quantitative Imaging in Medicine and Surgery*, vol. 6, no. 1, 2016. [Online]. Available: https://qims.amegroups.org/article/view/9264

[10] M. D'Anto, M. Cesarelli, P. Bifulco, M. Romano, V. Cerciello, F. Fiore, and A. Vecchione, "Study of different time attenuation curve processing in liver ct perfusion," in *Proceedings of the 10th IEEE International Conference on Information Technology and Applications in Biomedicine*. IEEE, Nov. 2010. [Online]. Available: http://dx.doi.org/10.1109/ITAB.2010.5687750

[11] A. Fieselmann and M. Manhart, "C-arm ct perfusion imaging in the interventional suite," *Current Medical Imaging Reviews*, vol. 9, no. 2, p. 96–101, May 2013. [Online]. Available: http://dx.doi.org/10.2174/1573405611309020004

[12] V. Kulvait and G. Rose, "Software Implementation of the Krylov Methods Based Reconstruction for the 3D Cone Beam CT Operator," in *16th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, 2021.

[13] V. Kulvait, P. Hoelter, R. Frysch, H. Haseljić, A. Doerfler, and G. Rose, "A novel use of time separation technique to improve flat detector ct perfusion imaging in stroke patients," *Medical Physics*, vol. 49, no. 6, p. 3624–3637, Apr. 2022. [Online]. Available: http://dx.doi.org/10.1002/mp.15640

[14] S. H. Kim, A. Kamaya, and J. K. Willmann, "Ct perfusion of the liver: Principles and applications in oncology," *Radiology*, vol. 272, no. 2, p. 322–344, Aug. 2014. [Online]. Available: http://dx.doi.org/10.1148/radiol.14130091

[15] K. Mouridsen, S. Christensen, L. Gyldensted, and L. Østergaard, "Automatic selection of arterial input function using cluster analysis," *Magnetic Resonance in Medicine*, vol. 55, no. 3, p. 524–531, Feb. 2006. [Online]. Available: http://dx.doi.org/10.1002/mrm.20759

[16] A. Fieselmann, M. Kowarschik, A. Ganguly, J. Hornegger, and R. Fahrig, "Deconvolution-based ct and mr brain perfusion measurement: Theoretical model revisited and practical implementation details," *International Journal of Biomedical Imaging*, vol. 2011, p. 1–20, 2011. [Online]. Available: http://dx.doi.org/10.1155/2011/467563

[17] K. Kudo, "Acute Stroke Imaging Standardization Group Japan recommended standard LUT (a-LUT) for perfusion color maps." http://asist.umin.jp/data-e.shtml, 2005, accessed April 9, 2021.

[18] H. Haseljić, V. Kulvait, R. Frysch, F. Saad, B. Hensen, F. Wacker, I. Brüsch, T. Werncke, and G. Rose, "Time separation technique using prior knowledge for dynamic liver perfusion imaging," in *7th International Conference on Image Formation in X-Ray Computed Tomography*, J. W. Stayman, Ed. SPIE, Oct. 2022. [Online]. Available: http://dx.doi.org/10.1117/12.2646449

Development of a Reusable 3D-Printed Phantom for Contrast Agent Evaluation in Photon-Counting Computed Tomography

Zeyu Yang, Raffael Lutz, and Frank G. Zöllner, *Senior Member, IEEE*

***Abstract*—We design and utilize a 3D-printed phantom for the precise assessment of CT imaging. We conduct a comparative quantitative analysis of Photon-Counting CT and Dual-Energy CT across varied contrast agent concentrations and dimensions in different backgrounds. The result demonstrate the superior image quality and contrast resolution of PC-CT over DE-CT. PC-CT shows notably higher Contrast-to-Noise Ratio and more consistent Hounsfield Unit values across different tube sizes and concentrations. Particularly in complex background scenarios, PC-CT exhibits enhanced capability in differentiating between various contrast agent groups. These findings highlight the potential of PC-CT in improving diagnostic accuracy in radiology. Using this phantom, the primary aim of our study is to improve the differentiation capabilities of computed tomography systems for accurate identification of tissues and lesions.**

***Index Terms*—3D printing, computed tomography, phantom.**

SPECTRAL computed tomography (spectral CT) is a cornerstone of modern diagnostic imaging, offering an unparalleled window into the internal architecture of the human body. Its ability to delineate and quantify physical and chemical properties of tissues has made it indispensable in clinical decision-making [1]. Evolving from dual-energy CT (DE-CT) to the latest photon-counting CT (PC-CT), spectral CT demonstrates a notable potential for enhanced tissue differentiation and artifact correction, thereby improving diagnostic precision [8].

In this realm, phantoms play a crucial role, widely employed for calibration [4], quality control testing [6], and technique validation [7] of CT devices, as well as substituting for tissue in imaging experiments when necessary [5]. However, the majority of studies have relied on commercial phantoms, which, despite their utility, often come with high costs and may not adequately meet specific research requirements. This limitation underscores the need for a flexible, cost-effective phantom that can be adapted to various diagnostic dimensions and internal configurations. Addressing this need, the field of 3D printing has risen to the occasion, revolutionizing the creation of custom phantoms for imaging evaluation [2]. The flexibility of 3D printing is particularly beneficial in developing phantoms that can simulate a variety of tissue characteristics and contrast levels, thereby enabling a more nuanced calibration of imaging equipment [3].

In the context of our study, a 3D-printed phantom is designed for repeated measures of contrast agents, facilitating the precise assessment of CT imaging. This phantom, with its array of tubes in different diameters, is tailored to explore the imaging ability of PC-CT across variable concentrations of iodine, within diverse backgrounds. Our research is dedicated to addressing the existing research gap by offering a reproducible and standardizable methodology for evaluating and refining CT technology. This approach aims to enhance the precision of CT systems in distinguishing tissues and lesions, thereby facilitating improved diagnostic accuracy and patient care.

I. Methods

A. Phantom Fabrication and Concentration Planning

The fabrication of the phantom was conducted utilizing a high-resolution 3D printer (Form 3+, Formlabs, Somerville, MA, USA) capable of producing intricate structures with a precision of up to 0.025 mm. The chosen material, a radiolucent resin (Clear Resin), was selected for its compatibility with CT imaging modalities, ensuring minimal artifact generation. The phantom comprises an inner component featuring three rows of eight cylindrical tubes, each with inner diameters ranging from 1.00 to 8.00 mm, as depicted in Fig. 1. Designed to simulate vessels when filled with contrast agents, the tubes mimic the dimension of human blood vessels, as referenced in [9], [10]. The wall thickness of each tube is 1mm. This inner component is inserted into an outer box-shaped structure. The box can be filled with various solvents to create different background densities. Multiple of these inner and outer components, once assembled, are arranged in a defined configuration, with each unit ensuring leak-proof integrity to prevent the mixing of contrast agents and solvents. This arrangement facilitates consistent orientation and spacing.

Concentration planning was performed to evaluate the photon-counting CT's ability to discriminate between varying iodine concentrations under different backgrounds. A range of iodine-based contrast solutions at concentrations of 2mg/ml, 5mg/ml, and 10mg/ml were formulated in the tubes. Two phantoms, each with a rectangular base, were constructed to facilitate the comparison of identical concentration sets within distinct backgrounds. Consequently, this study employed two background scenarios: air and an iodine-based contrast solution of 15mg/ml.

Z. Yang and F. G. Zöllner are with the Computer Assisted Clinical Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany, e-mail: zeyu.yang@medma.uni-heidelberg.de, frank.zoellner@medma.uni-heidelberg.de).

R. Lutz is with the Automation in Medicine and Biology, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany e-mail: raffael.lutz@medma.uni-heidelberg.de

Z. Yang, R. Lutz, and F. G. Zöllner are with the Mannheim Institute for Intelligent Systems in Medicine, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

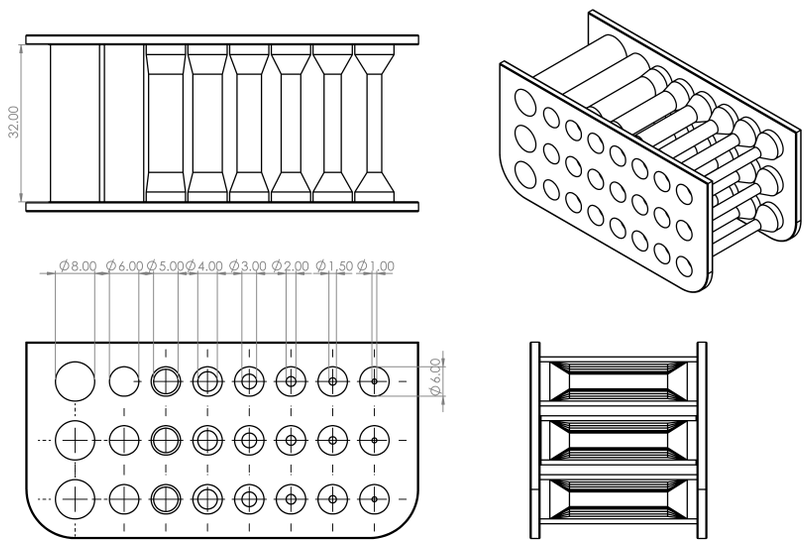


Fig. 1. Three-dimensional schematic representation of the phantom used for evaluation purposes, illustrating the top view, side view, and frontal section. The schematic details various diameters, ensuring precision across multiple measurement scales.

B. Image Acquisition and Data Analysis

Image acquisition was performed on two clinical scanners. In the DE-CT scanner (Somatom Force, Siemens Healthcare, Forchheim, Germany), the settings for the low voltage tube were established at 100 kVp, and the high voltage tube at 150 Sn kVp. In contrast, for the PC-CT (Naeotom Alpha, Siemens Healthcare, Forchheim, Germany), both tubes were set to 120 kVp. All scans were conducted at a pitch of 1. The "Qr40" quantitative reconstruction kernel was utilized with both scanners. Regarding iterative reconstruction strength, QIR 3 was selected for the PC-CT, and ADMIRE 3 for the DE-CT scanner. Reconstructions were executed using a slice thickness of 3 mm, a field of view measuring 244 mm, and a matrix size of 512×512. Since 70keV VMI energy level images are a standard output from the PC-CT, this level was selected for the initial imaging of the phantom.

For each tube, the CNR and mean and standard deviation of the CT numbers (expressed in Hounsfield units (HU)) was determined by placing regions of interest (ROI) and calculated with twenty consecutive images. CNR was defined as the difference in mean CT numbers of iodine solution region and background region, divided by the standard deviation of the background region. The computations were performed with an in-house developed script (MATLAB R2021b, The MathWorks Inc., Natick, MA, USA).

II. Results

The quantitative assessment of the PC-CT and DE-CT is presented in Fig. 2 and summarized in Table. I, II, and III. In this chapter, the results from the analysis are discussed.

Fig. 2 illustrates representative slices from the DE-CT and the PC-CT scanner. Panels (a) and (b) depict images obtained from the PC-CT scanner, while panels (c) and (d) are from the DE-CT scanner. The images were displayed with a window width of -1000–500 HU to ensure adequate visualization of the contrast resolutions. The images produced by the PC-CT exhibit markedly improved quality over those from the DE-CT, with notable enhancements in noise reduction and resolution clarity. Across both background scenarios presented, the PC-CT outperforms the DE-CT in differentiating between the various groups of contrast agent solutions.

The quantitative evaluation is presented in Tables I, II, and III. These tables compare our (dimension, concentration) sets under two distinct background settings: a dark background (dark BG) and a bright background (bright BG), representing air and an iodine-based solution of 15mg/ml, respectively.

Table I reveals that PC-CT demonstrates superior CNR across all test sets. Tables II and III detail the mean $\pm$ standard deviation HU for each test case. The findings indicate that PC-CT generally exhibits a smaller discrepancy in HU values between tube sizes in the dark BG compared to DE-CT. However, in the bright BG, the variation in HU values between small and large tube sizes is similar for both PC-CT and DE-CT. Additionally, a smaller difference in HU values for the same concentration and dimension group under the two backgrounds implies that PC-CT may yield reduced error and enhanced imaging capabilities in diverse background conditions compared to DE-CT.

III. Conclusion

In conclusion, this study designed and utilized a 3D-printed phantom to evaluate the efficacy of iodine-based contrast agent with varying concentrations across different dimensions and backgrounds in CT imaging. The findings indicate a superior performance of the PC-CT in both observational and quantitative assessments compared to DE-CT.

Despite these promising results, areas for improvement have been identified. Primarily, although the influence of tube size on PC-CT is not apparent when exceeding 2mm, its significance should not be overlooked. The problem arises from the influence of tube walls. In particular, with smaller dimensions, it is challenging to exclude the silicon walls from

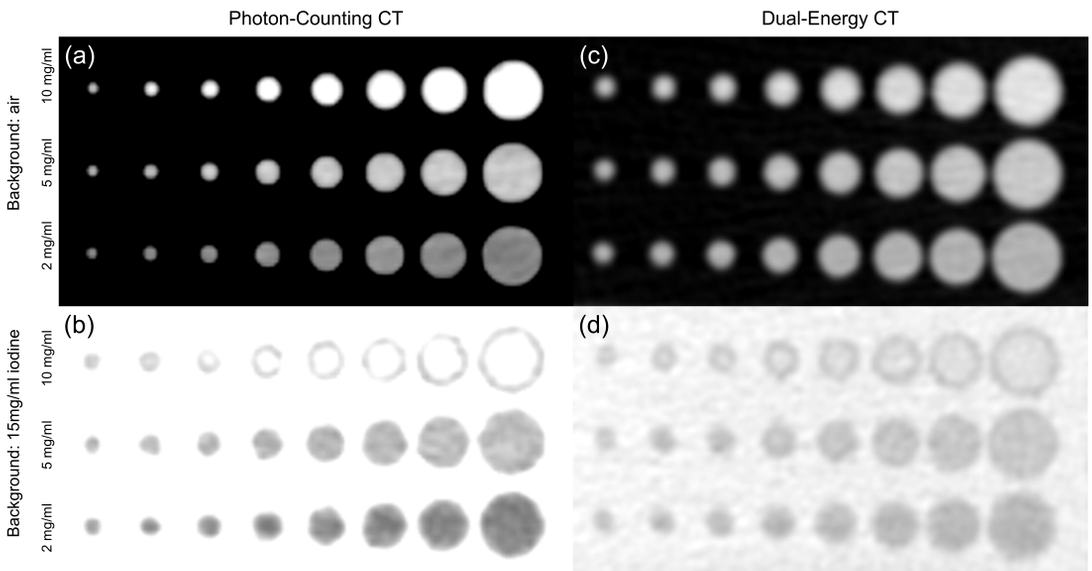


Fig. 2. Comparative slices from DE-CT (Dual-Energy Computed Tomography) and PC-CT (Photon-Counting Computed Tomography) scanners.

TABLE I. CNR Measurements for Photon-Counting CT and Dual-Energy CT

	Photon-Counting CT								Dual-Energy CT							
CNR	1mm	1.5mm	2mm	3mm	4mm	5mm	6mm	8mm	1mm	1.5mm	2mm	3mm	4mm	5mm	6mm	8mm
dark BG																
10mg/ml	155.3	159.8	157.3	143.0	182.5	161.8	154.4	166.6	62.4	68.5	69.9	72.9	73.4	68.7	70,0	72.9
5mg/ml	141.4	146.5	137.6	140.5	150.5	142.3	128.4	116.5	60.0	63.5	61.0	61.7	62.8	55.8	62.7	63.8
2mg/ml	129.3	131.6	131.4	125.6	125.2	126.9	128.6	126.8	53.0	53.1	48.2	52.4	54.1	53.8	55.1	54.9
bright BG																
10mg/ml	16.8	14.0	12.2	9.1	7.9	7.9	7.3	7.9	7.8	7.4	6.1	4.7	3.9	4.6	4.7	4.1
5mg/ml	17.8	19.9	17.9	12.6	15.8	14.0	14.5	15.8	7.7	8.2	8.5	7.9	8.1	7.9	7.4	7.0
2mg/ml	19.4	20.3	21.8	22.0	21.6	22.8	22.2	23.7	8.1	9.7	9.5	10.2	10.3	9.8	9.8	10.6

[1] dark BG - background is air, so the background is darker than signal region
[2] bright BG - background is iodine-based contrast solution of 15mg/ml, so the background is brighter than signal region

TABLE II. Mean±Standard Deviation HU Values for for Photon-Counting CT

mean±std HU	1mm	1.5mm	2mm	3mm	4mm	5mm	6mm	8mm
dark BG								
10mg/ml	154.99±22.11	214.56±32.68	240.63±31.07	276.06±23.70	272.99±46.65	262.81±19.82	280.17±25.25	288.89±37.32
5mg/ml	99.73±42.78	126.63±31.37	145.30±14.51	151.58±23.34	164.63±20.89	164.62±12.01	165.65±11.09	161.18±10.85
2mg/ml	66.63±27.78	66.88±21.88	77.52±12.22	86.08±8.45	83.12±8.70	92.29±9.42	78.07±9.18	70.34±10.48
bright BG								
10mg/ml	179.85±14.01	201.71±17.41	224.93±19.09	257.42±24.62	278.53±25.98	264.35±22.18	272.40±21.11	273.54±17.94
5mg/ml	144.94±14.37	124.01±79.83	146.69±16.94	141.99±15.70	141.97±14.59	147.19±12.10	151.65±11.86	157.57±12.09
2mg/ml	118.80±15.48	107.08±14.31	86.83±17.58	86.76±15.69	89.74±16.05	76.06±13.83	77.80±13.38	73.59±15.50

TABLE III. Mean±Standard Deviation HU Values for Dual-Energy CT

mean±std HU	1mm	1.5mm	2mm	3mm	4mm	5mm	6mm	8mm
dark BG								
10mg/ml	82.36±46.66	168.07±44.67	208.61±47.88	239.78±62.57	262.18±64.74	270.91±26.49	313.73±21.59	347.83±23.33
5mg/ml	45.25±26.38	81.66±58.79	117.02±31.69	140.17±22.64	152.96±21.56	156.92±23.80	157.92±27.41	181.35±23.81
2mg/ml	11.35±17.56	17.86±29.51	29.54±28.94	42.45±20.32	38.80±22.89	55.00±21.54	49.07±24.52	55.67±23.19
bright BG								
10mg/ml	232.75±29.05	251.60±42.45	286.59±39.68	325.75±42.01	337.98±37.31	322.61±37.10	318.12 ± 41.98	331.00 ±35.91
5mg/ml	194.87±25.98	170.88±58.28	185.32±32.71	192.42±34.80	183.28±31.18	187.79±31.64	192.37±27.98	202.42 ±28.41
2mg/ml	187.58±41.06	135.00±30.27	145.84±29.55	120.79±31.93	128.98±30.08	114.69±26.91	110.19±28.41	101.11±26.56

the ROI due to partial volume effect (PVE). This phenomenon, where a single voxel might represent multiple material types simultaneously due to the CT scanner's limited resolution, is a recognized challenge in phantom study [11]. Notably, PC-CT demonstrated superior performance in reducing PVE attributed to its enhanced spatial resolution [12]. Our study employed tubes with 1mm walls to optimize wall strength. Initial prototypes showed that tubes with a 1mm internal diameter tended to crack during the 3D printing washing process. To strengthen these tubes, bases were attached to the ends in the latest version, without increasing wall thickness. Future investigations will focus on determining the minimal wall thickness that maintains structural integrity. Furthermore, the plan includes the implementation of higher spatial resolution scanning, achieved by smaller reconstruction ROI to reduce the PVE.

A comparison across two different backgrounds revealed that, for PC-CT, backgrounds minimally influenced tubes larger than 2mm. However, tubes of (1mm, 5mg/ml), (1mm, 2mg/ml), and (1.5mm, 2mg/ml) were notably affected by background noise. This phenomenon aligns with findings from previous research [14]. Future improvements will focus on mitigating this effect by increasing scanning dose and implementing advanced reconstruction techniques, thereby reducing the background's impact on signal regions [14], [15]. In addition, we aim to enhance our phantoms to more closely mimic liver tissues and tumors. This will involve incorporating a broader range of concentrations and dimensions, as well as integrating tumor and metal artifact inserts, following the methodologies outlined in [17], [18].

Concerning data analysis, previous research has yielded a tool for rapid, automated, and objective analysis of multi-center phantom images for quality control [13]. In forthcoming implementations, we plan to utilize this tool for enhanced automation and accuracy in our data analysis. Moreover, we intend to incorporate calibration marks into our phantoms, facilitating more accurate selection of ROI and enabling automated data analysis.

Overall, our study significantly contributes to the assessment of iodine-based contrast agents' effectiveness at various concentrations and in different dimensions and settings within PC-CT. According to the findings in [16], there is a minimal difference between the HU values of liver and tumor voxels, with a mean difference of 31.94 and a standard deviation of 20 within the Liver Tumour Segmentation Benchmark (LiTS) dataset. Hence, the development of phantoms that accurately mimic these subtle concentration differences, particularly in smaller dimensions, is essential. Utilizing such phantoms, we aim to evaluate and enhance capability to distinguish between tissues and lesions of CT systems, thereby not only advancing post-processing algorithm performance but also elevating diagnostic accuracy at the clinical level.

Acknowledgment

This research project is part of the Research Campus M2OLIE and funded by the German Federal Ministry of Education and Research (BMBF) within the Framework "Forschungscampus: public-private partnership for Innovations" under the funding code 13GW0388A.

References

[1] S. Z. Adam, A. Rabinowich, R. Kessner, and A. Blachar, "Spectral CT of the abdomen: Where are we now?," Insights Imaging, vol. 12, no. 1, p. 138, Dec. 2021, doi: 10.1186/s13244-021-01082-7.

[2] D. F. Bauer et al., "An anthropomorphic pelvis phantom for MR-guided prostate interventions," Magn Reson Med, vol. 87, no. 3, pp. 1605–1612, Mar. 2022, doi: 10.1002/mrm.29043.

[3] D. F. Bauer et al., "Development of an abdominal phantom for the validation of an oligometastatic disease diagnosis workflow," Med Phys, vol. 49, no. 7, pp. 4445–4454, Jul. 2022, doi: 10.1002/mp.15701.

[4] M. Nakao et al., "Development of a CT number calibration audit phantom in photon radiation therapy: A pilot study," Med Phys, vol. 47, no. 4, pp. 1509–1522, Apr. 2020, doi: 10.1002/mp.14077.

[5] S. Leng, L. Yu, T. Vrieze, J. Kuhlmann, B. Chen, and C. H. McCollough, Construction of realistic liver phantoms from patient images using 3D printer and its application in CT image quality assessment. Presented at the SPIE Medical Imaging, C. Hoeschen, D. Kontos, and T. G. Flohr, Eds., Orlando, Florida, United States, Mar. 2015, p. 94124E. doi: 10.1117/12.2082121.

[6] W. A. McDougald and J. G. Mannheim, "Understanding the importance of quality control and quality assurance in preclinical PET/CT imaging," EJNMMI Phys, vol. 9, no. 1, p. 77, Oct. 2022, doi: 10.1186/s40658-022-00503-w .

[7] F. Moloney et al., "A phantom study of the performance of model-based iterative reconstruction in low-dose chest and abdominal CT: When are benefits maximized?," Radiography, vol. 24, no. 4, pp. 345–351, Nov. 2018, doi: 10.1016/j.radi.2018.04.010.

[8] 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, p. mp.16583, Jul. 2023, doi: 10.1002/mp.16583.

[9] S. Scheuer, F. G. Zöllner, E. Tumat, and L. R. Schad, "Untersuchung von Strömungsverhältnissen an künstlichen Modellen mittelgroßer stenotischer Gefäße mit der 3D-Phasenkontrast-Magnetresonanztomographie," Z Med Phys, vol. 20, no. 1, pp. 34–45, Mar. 2010, doi: 10.1016/j.zemedi.2009.10.011.

[10] D. B. Camasão and D. Mantovani, "The mechanical characterization of blood vessels and their substitutes in the continuous quest for physiological-relevant performances. A critical review," Materials Today Bio, vol. 10, p. 100106, Mar. 2021, doi: 10.1016/j.mtbio.2021.100106.

[11] J. . -M. Kuhnigk et al., "Morphological segmentation and partial volume analysis for volumetry of solid pulmonary lesions in thoracic CT scans," IEEE Trans Med Imaging, vol. 25, no. 4, pp. 417-434, April 2006, doi: 10.1109/TMI.2006.871547.

[12] S. Si-Mohamed et al., "Review of an initial experience with an experimental spectral photon-counting computed tomography system," Nucl. Instrum. Methods. Phys. Res. A: Accelerators, Spectrometers, Detectors and Associated Equipment, vol. 873, pp. 27–35, Nov. 2017, doi: 10.1016/j.nima.2017.04.014.

[13] M. Davids et al., "Fully-automated quality assurance in multi-center studies using MRI phantom measurements," Magn Reson Imaging, vol. 32, no. 6, pp. 771–780, Jul. 2014, doi: 10.1016/j.mri.2014.01.017.

[14] J. Conzelmann, U. Genske, A. Emig, M. Scheel, B. Hamm, and P. Jahnke, "Comparison of low-contrast detectability between uniform and anatomically realistic phantoms—influences on CT image quality assessment," Eur Radiol, vol. 32, no. 2, pp. 1267–1275, Feb. 2022, doi: 10.1007/s00330-021-08248-3.

[15] L. M. Popescu and K. J. Myers, "CT image assessment by low contrast signal detectability evaluation with unknown signal location," Med Phys, vol. 40, no. 11, p. 111908, Nov. 2013, doi: 10.1118/1.4824055.

[16] S. Lysdahlgaard, "Comparing Radiomics features of tumour and healthy liver tissue in a limited CT dataset: A machine learning study," Radiography, vol. 28, no. 3, pp. 718–724, Aug. 2022, doi: 10.1016/j.radi.2022.03.015.

[17] N. Subhas et al., "Dose Reduction With Dedicated CT Metal Artifact Reduction Algorithm: CT Phantom Study," Am J Roentgen, vol. 210, no. 3, pp. 593–600, Mar. 2018, doi: 10.2214/AJR.17.18544.

[18] T. Njølstad, K. Jensen, A. Dybwad, Ø. Salvesen, H. K. Andersen, and A. Schulz, "Low-contrast detectability and potential for radiation dose reduction using deep learning image reconstruction—A 20-reader study on a semi-anthropomorphic liver phantom," Eur. J. Radiol. Open, vol. 9, p. 100418, 2022, doi: 10.1016/j.ejro.2022.100418.

Variance-bias analysis of dual-energy CT image reconstruction for objects containing metal

Michael S. Ibrahim, Emil Y. Sidky, Fatma Terzioglu, John Paul Phillips, Ingrid S. Reiser, Guillaume Bal, Xiaochuan Pan

***Abstract*—The goal of this study is to investigate two-step processing in dual energy CT (DECT) when three distinct tissues/materials are present in the scanned subject. In particular, the impact of two material and three material decomposition on metal artifact reduction and imaging with K-edge contrast agents, such as Iodine, on precision and accuracy of the virtual monochromatic image (VMI) is studied. The obtained results are based on a previously developed analytic model for DECT imaging noise properties, which exploits transform linearization and a Gaussian approximation to the accurate noise model. The analytic model is first verified against empirical realizations generated with the accurate DECT data model. Using the analytic model, material decomposition with two and three basis materials is compared within a DECT simulation where the scanned subject is composed of water, bone, and a third material, which is either the metal Titanium or a 10% Iodine insert. Little difference between two and three material decomposition is seen for the case where the third material is metal. When the third material is the Iodine object, however, there is potentially an advantage for use of three material decomposition.**

***Index Terms*—Dual-energy CT image reconstruction, metal artifacts, Iodine-based contrast agent**

I. Introduction

Our recent work on DECT imaging noise properties, developed in Terzioglu *et. al.* [1], presents a model for DECT, based on Alvarez and Macovski's initial introduction of the concept [2], including a model for analytically determining image variance. In this paper, we will test the application of this model on a simulated metal-containing phantom, as well as empirically testing the validity of the variance model through noise realizations.

Metal artifacts continue to pose an issue for CT imaging due to four main factors: beam hardening, photon starvation, scattering, and the partial volume effect [3]. One of the principal causes, beam hardening, is due to the lack of energy resolution in single energy CT. Beam hardening is accounted for in DECT, which acquires images using two distinct energy spectra and uses an updated transmission model. The contribution of this work is to provide a means to optimize parameters of DECT imaging based on the noise properties of the resulting VMI. Thus, indirectly, photon starvation is addressed by selecting DECT parameters that minimize noise variance in the VMI.

M. S. Ibrahim, E.Y Sidky, J. P. Phillips, I. S. Reiser, and X. Pan are with the Department of Radiology MC-2026, The University of Chicago, 5841. S. Maryland Ave., Chicago, IL, 60637.

F. Terzioglu is with the Department of Mathematics, North Carolina State University, 2311 Stinson Dr., Raleigh, NC, 27695

G. Bal is with the Departments of Statistics and Mathematics, The University of Chicago, 5747 S. Ellis Ave., Chicago, IL, 60637

As part of the DECT parameter optimization, we investigate two alternatives for material decomposition. From the two independent acquisitions at the two energies, the pathlengths, or sinograms, of the basis materials for each beam path are solved from the transmission measurements. From the sinograms, the basis material maps are reconstructed by filtered back-projection (FBP). For a two-material decomposition, the DECT transmission-to-sinogram processing can yield unique results. For a three-material decomposition, we can also assume that the object is made of three materials and solve for those pathlengths. This will not give us a unique determination of three material maps, because the system is underdetermined in this case. However, in forming a virtual monochromatic image (VMI), use of three basis materials in DECT may yield accurate results. We investigate this possibility and if there is some potential advantage to using three basis materials in DECT imaging of an object with three materials.

II. Methods

The following methods use a physical model for DECT transmission data that includes a compound Poisson process noise model as presented in Ref. [1].

A. Transmission model

The linear attenuation coefficient at a given point for an object made of N materials is given by

$$M(E,y) = \sum_{j=1}^{N} M_j(E)\rho_j(y)$$

Where the $M_j(E)$ are the linear attenuation coefficients of the various materials at energy E, and the $\rho_j(y)$ are the densities of those materials at a given location y. We can express the linear attenuation coefficients as a vector as follows:

$$M(E) = \left[M_1(E), M_2(E), \ldots, M_N(E)\right]^\top.$$

For a given beam path l, we can also write the x-ray transform of the jth material density map as

$$x_j(\ell) = \int_\ell \rho_j(y)dy.$$

The transmitted intensity for a given path ℓ for a dual energy measurement is given by

$$I_i(x_l) = \int_0^\infty S_i(E)e^{-M(E)\cdot x_l} D(E)dE, \text{ for } i \in 1,2,$$

where $S_i(E)$ are the emitted spectra of the two x-ray sources, and $D(E)$ is the detector response function, which is assumed

proportional to E for energy integrating detectors. We will also be working with the relevant quantity

$$g_i(x) = -\ln I_i(x),$$

and the associated Jacobian

$$J_{ij}(x) = \frac{\partial g_i}{\partial x_j}(x) = \frac{\int_0^\infty D(E)S_i(E)M_j(E)e^{-M(E)\cdot x}dE}{\int_0^\infty D(E)S_i(E)e^{-M(E)\cdot x}dE}.$$

B. Noise model

Noise for the DECT transmission data is simulated through a compound Poisson process. For the analytic noise model, a Gaussian approximation is made and the non-linear transforms of the transmission data are approximated to be linear about the mean. As developed in Ref. [1], the covariance matrix for the two transmitted intensities is

$$\Sigma_g(x) = \begin{bmatrix} \sigma_g(x)_1^2 & 0 \\ 0 & \sigma_g(x)_2^2 \end{bmatrix}$$

where

$$\sigma_g(x)_i^2 = \frac{\int_0^\infty D^2(E)S_i(E)e^{-M(E)\cdot x_l}dE}{\left(\int_0^\infty D(E)S_i(E)e^{-M(E)\cdot x_l}dE\right)^2}$$

The covariance matrix of the material sinograms is then

$$\Sigma_{x_l} = \Sigma_l = J^{-1}(x_l)\Sigma_g(x_l)J^{-\top}(x_l),$$

where $-\top$ denotes the transpose of the inverse matrix. When the Jacobian J is not invertible, J^{-1} denotes the pseudo-inverse. In the image domain the basis image covariance is found by transforming the sinogram covariance with the FBP operator

$$q_{ij} = (1/P)\text{tr}(BC^{(ij)}B^\top), \text{ for } j = 1, 2,$$

where P is the number of pixels over which the variance is being averaged; B is the FBP operator; C is defined as

$$(C^{(ij)})_{ll} = (\Sigma_l)_{ij};$$

and $\text{tr}(\cdot)$ denotes the matrix trace, i.e., the sum of its diagonal entries.

C. Two and three material decomposition

In order to perform two material decomposition in the empirical studies, we solve for the material sinograms x_l from the dual energy measurements g using Newton's method. Applying FBP yields the material maps. For three material decomposition, we apply the same processing with the only difference being how the Newton's algorithm is implemented. This algorithm involves inversion of the Jacobian for the sinogram-to-transmission transform. For three materials, J is clearly not invertible and Newton's algorithm is implemented using the pseudo-inverse of J. This allows us to obtain three material maps which are not accurate on their own, but when combined, weighting them by their respective materials' relative attenuation at a given VMI energy, we obtain a reasonably accurate image.

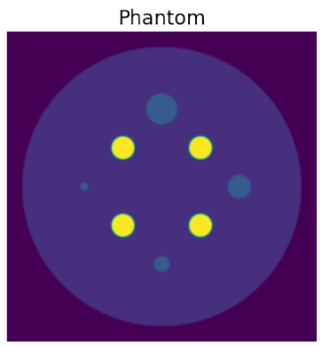


Fig. 1. Test object for the three-material studies. The background disc is modeled as water; the cyan discs are Bone tissue; and the yellow disc is taken to be either the metal Titanium or a 10% Iodine concentration insert.

D. Virtual monochromatic images

Virtual monochromatic images (VMIs) are obtained through a weighted sum of the material maps as follows

$$x_{\text{mono}} = w \cdot x \text{ where } w = \frac{M(E)}{|M(E)|^2}.$$

The variance in the VMI is obtained as follows:

$$\sigma^2_{\text{VMI}} = w^\top \Sigma_l w,$$

where the length of the weight vector w is two or three depending on the number of basis materials considered.

E. Noise Realizations

The analytic formula for the covariance matrix is based on a Gaussian approximation of the Poisson distribution and needed to be tested empirically. A phantom, shown in Fig. 1, consisting of a 9 cm diameter water disc with 4 embedded titanium discs that are 0.6 cm in diameter was used to repeatedly simulate noisy measurements in order to test the noise model. Noisy measurements were simulated as a compound Poisson process, and the noisy measurements were repeated 700 times to get an empirical estimate of the variance and covariance of the material maps and VMI (105 keV). The covariance matrix and VMI variance were estimated in both the sinogram and image domain. This was also repeated with 3 material decomposition. These were then compared to the theoretical values obtained from the analytic formula to test the model.

F. Comparing two and three material decomposition

The accuracy of the images generated through two and three material decomposition was compared over a wide range of VMI energies. Two different three-material phantoms were used: one consisting of water, cortical bone, and titanium, and another consisting of water, cortical bone, and a 10% Iodine disc insert.

At each energy, VMIs were generated using two and three material decomposition. Average relative bias (average difference between the VMI and true image divided by true image) and average analytic standard deviation in the sinogram domain were plotted for each VMI energy. Average relative

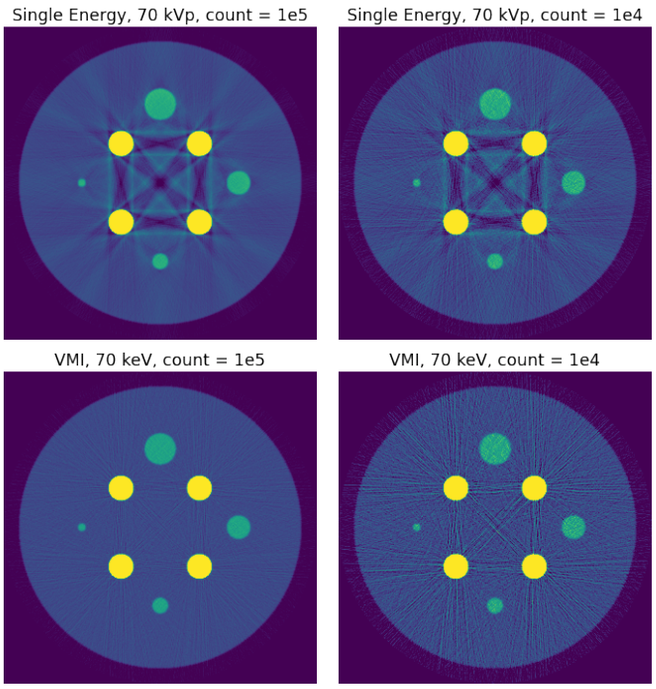


Fig. 2. (Top row) Images reconstructed with standard CT processing from simulated transmission data using a polychromatic model for a 70 kVp spectrum. (Bottom row) DECT VMIs at 70 keV for 70/140 kVp low/high potential settings. (Left column) Simulated fluence incident per pixel is 10^5 photons. (Right column) Simulated fluence incident per pixel is 10^4 photons.

bias and relative standard deviation were also plotted for the bone region (the third material not used for two material decomposition). Converting variance from the sinogram to image domain is very computationally intensive, which is why for analyzing the standard deviation of the whole image we used the sinogram domain. Water and either titanium or Iodine were used as basis materials.

III. Results

A. Effectiveness of DECT in reducing metal artifacts

Using the phantom in Fig. 1, images from simulations of standard single-energy CT and the VMI of DECT are shown in Fig. 2. Images are 512x512, and the scan geometry is circular, fan-beam with 512 views and 512 detector pixels. In the case of sufficient dose of x-rays, where beam hardening is the main cause of metal artifacts, the metal artifacts are all but removed in the dual energy case. For lower dose, streaks due to photon starvation are visible and are amplified relative to the standard CT image, because the process of solving for the pathlengths amplifies much of the noise. This sensitivity to noise due to DECT processing motivates this study on DECT scan parameter optimization.

B. Testing the analytic noise propagation model

The analytic noise model, based on Gaussian and linearization aprroximations, is compared against empirical realizations using the accurate data model for both two and three material decomposition. Only results for the two-material case are

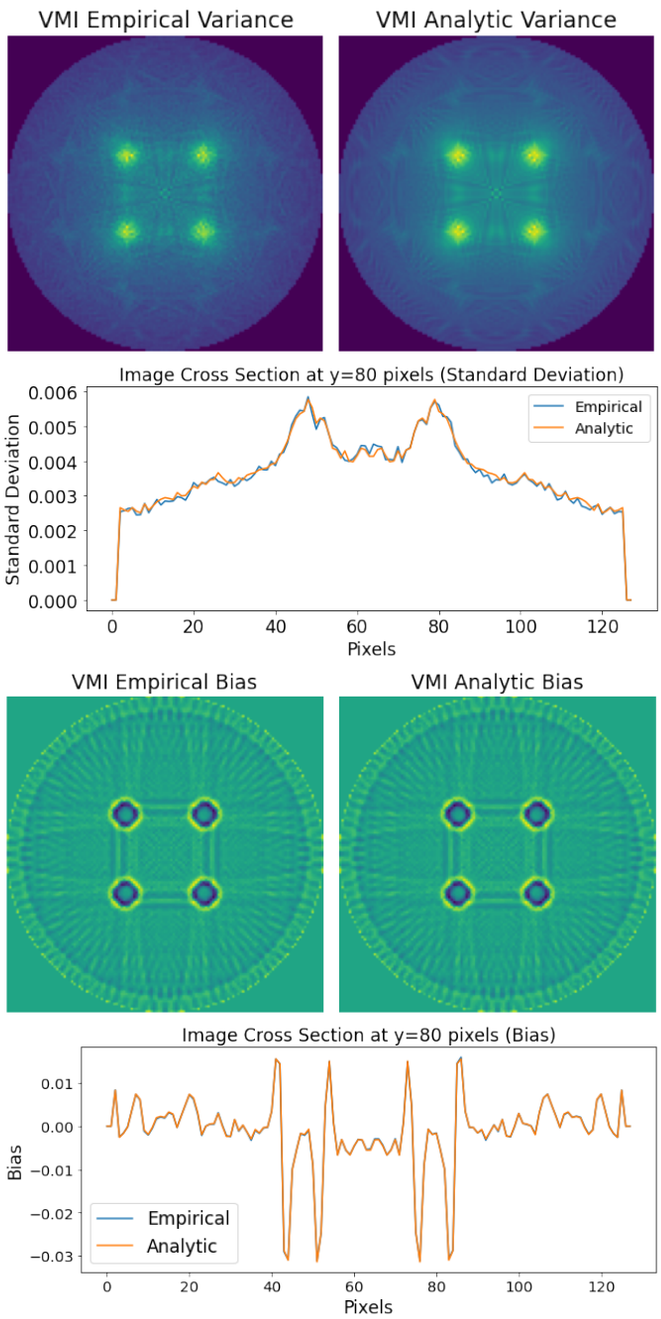


Fig. 3. (First row) Variance images of the DECT VMI derived from noise realizations (left) and the analytic model (right). (Second row) Line profile comparison of the empirical and analytic variance results. (Third row) Bias images of the DECT VMI derived from noise realizations (left) and the analytic model (right). (Bias row) Line profile comparison of the empirical and analytic bias results.

shown in Fig. 3. For these noise studies the dimensions of the system were reduced to a 128x128 image and 128 projections taken on a 128 pixel detector. Noise realizations were done for a two material phantom that consisted of the water disc and the four titanium discs. The model matched up with the empirical testing, indicating the validity of the Gaussian approximation and allowing us to use the analytic formula as a measure of variance rather than having to run noise realizations. The same

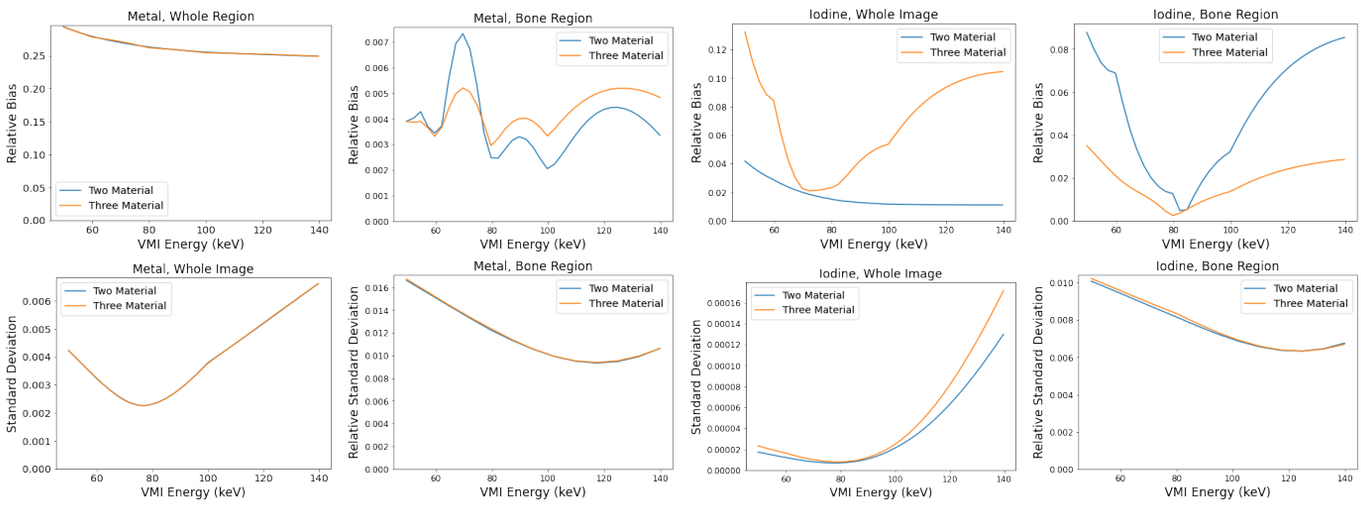


Fig. 4. Statistical properties of VMIs as a function of VMI energy for the three-material study with the third material being Titanium (labeled as "metal") or Iodine. The top row shows bias over the whole image or within only the bone regions. The bottom row shows the corresponding standard deviation.

holds true for the three material case. The same tests were done for all elements of the covariance matrix, but were not pictured.

C. Two vs. three material decomposition and optimal VMI energy

For the final set of results, two and three material decomposition is compared using the methods described in Sec. II-F and the plots are shown in Fig. 4. The DECT source potentials were taken to be 70 and 140 kVp. For the two material basis set either water/Titanium or water/Iodine is used, depending on which third material was used to generate the data.

For all results, it is clear that there is little difference in the standard deviation between using a two- or three-material basis. For the study with Titanium as the third material, bias over the whole object is also similar between the two cases. When focusing on the bone regions there is a clear difference in the use of two- or three-material decomposition with the latter providing a smoother dependence on VMI energy than the former but the bias results do not consistently favor one choice over the other.

In the case where Iodine is present instead of Titanium, there is a clear advantage in using the three-material decomposition for reducing bias in the bone region. This reflects the fact that bone cannot be expressed as simply a linear combination of water and Iodine, which is what the two material decomposition tries to do. The extra degree of freedom offered by three material decomposition likely allowed the representation to be more accurate. For the image as a whole however, we see that the overall bias is larger for three material decomposition, indicating a tradeoff between decreasing overall image bias and decreasing bias of the third non-basis material.

In general, we see that image standard deviation is minimized at around 80 keV, although for bone regions standard deviation tends to decrease with VMI energy. Image bias similarly shows a trend to gradually decrease with VMI energy, with the notable exception of the bone bias in the Iodine case, in which case bias is minimized at around 80 keV. In general, it seems that 80 keV is an optimal VMI energy, but a high energy VMI could be useful in some cases.

IV. Conclusion

The main new result from this work is the comparison of two- and three-material decomposition on DECT VMI noise properties. While three material decomposition is not expected to yield accurate basis images from DECT data, it is seen that bias in the VMI can be highly accurate. Furthermore, there is a clear benefit for the case when the subject is composed of water, bone and a K-edge material such as Iodine-based contrast agent. For this case, two material decomposition struggles in representing the non-basis material accurately. It was also shown that 80 keV is a VMI energy that generally minimizes noise variance, but this could be dependent on the specifics of the phantom that we used.

This work will be extended by use of more realistic phantoms with possibly more materials to obtain a better picture of the effectiveness of three-material decomposition in obtaining the DECT VMI. Also, results on varying the DECT tube settings will be presented at the meeting.

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 R01EB023968 and R21CA263660.

References

[1] F. Terzioglu, E. Y. Sidky, J. P. Phillips, I. Reiser, G. Bal, and X. Pan, "Optimizing dual-energy CT technique for iodine-based contrast-to-noise ratio," *Med. Phys.*, vol. 51, pp. 2871–2881, 2024.

[2] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Phys. Med. Biol.*, vol. 21, pp. 733–744, 1976.

[3] L. Gjesteby, B. De Man, Y. Jin, H. Paganetti, J. Verburg, D. Giantsoudi, and G. Wang, "Metal artifact reduction in CT: where are we after four decades?" *IEEE Access*, pp. 5826–5849, 2016.

Three-mode CT imaging of X-ray transmission, scatter and fluorescence photons

Jiadan Song, Shaozhou Pu, Liang Li

***Abstract*—X-ray transmission computed tomography (XCT) could image the object's structure. X-ray fluorescence computed tomography (XFCT) could perform quantitatively imaging of high-Z element such as Gd and Au. Combining with the reconstruction results of XCT and XFCT, we could realize the specific imaging of tumors, which will provide assistance for tumor diagnosis. The dual-modality imaging system with XCT and XFCT has been widely studied due to its great application potential in the field of tumor diagnosis. However, the performance of XFCT is not good enough for clinical application at present, which limits the further application of this dual-modality system greatly. Therefore, it is crucial for us to improve the performance of XFCT. In the existing system, the detector of XFCT collects not only fluorescence photons but also Compton scattered photons, while Compton scattered photons are usually treated as noise. However, scattered photons contain structural information as well. We could image the electron density through the reconstruction of Compton scatter computed tomography (CSCT), which is a kind of supplement to XCT. Meanwhile, electron density is also helpful for attenuation correction of XFCT. Now, through numerical simulation and phantom experiments, we have realized the three-mode imaging system of XCT, XFCT, and CSCT, which enables the reconstruction of linear attenuation coefficient, high-Z element concentration and electron density. For this system, we also proposed a new algorithm to calculate the linear attenuation coefficient of the object at different energies and different positions, which also assists us to perform quantitative reconstruction of XFCT accurately.**

***Index Terms*—Compton scatter, X-ray fluorescence, transmission, computed tomography, multi-modality imaging system**

I. Introduction

X-ray fluorescence computed tomography (XFCT) is a good tool for quantitative imaging of high-Z elements. When injecting nanoparticles composed of high-Z elements such as Gd and Au into body, high-Z element will accurate in tumor. Then, we could achieve tumor imaging through quantitative reconstruction of high-Z elements [1, 2]. At the same time, X-ray transmission computed tomography (XCT) could provide structural information of body. Combining the reconstruction results of XFCT and XCT, we will get the specific tumor location in *vivo*, which is very important to tumor diagnosis and treatment. Therefore, the dual-modality imaging of XFCT and XCT has been widely studied due to its broad application prospects in the medical field [3-6].

However, this dual-modality imaging system is currently limited by the performance of XFCT and it is important for us to improve the XFCT performance. We find that in the process of fluorescence signal collection, Compton scatter signals are collected simultaneously. For most studies in XFCT, only fluorescence signal is used while Compton scatter photons are treated as noise and need to be removed [1, 2]. But in fact, the scattered signal is useful and contains electron density information, which is related to the linear attenuation coefficient at different energies. In this paper, we design an algorithm to calculate the linear attenuation coefficients at different energies with the help of Compton scatter computed tomography (CSCT). Then, we successfully realize the attenuation correction of XFCT without any prior information. This is the first innovation in our work.

Meanwhile, the reconstruction results of CSCT, electron density, could also provide structure information [7, 8]. Electron density and linear attenuation coefficient are two different physical quantities. Therefore, it is possible that the structures which cannot be distinguished through XCT may be clearly distinguished through CSCT. CSCT could be viewed as a kind of supplement to XCT. As a result, it is meaningful that the dual-modality imaging system of XCT and XFCT is updated to the three-mode imaging system of XCT, XFCT and CSCT. This is the second innovation in this paper.

In this work, a three-mode imaging system of XCT, XFCT and CSCT is proposed and verified by numerical simulation and phantom experiments. Through three-mode reconstruction, we could obtain the linear attenuation coefficient, high-Z element concentration, and electron density. Then we propose a new algorithm to calculate the linear attenuation coefficients at different energies. With the results of this calculation, we could perform the attenuation correction of XFCT accurately.

II. Materials and Method

Fig. 1 is the schematic diagram of the three-mode imaging system. As shown in Fig. 1, transmitted photons are collected through a flat panel detector with high spatial resolution. Fluorescence and scatter signals are collected through a surface array photon counting detector with high energy resolution

This work is supported by the grant from Tsinghua University Initiative Scientific Research Program.

Jiadan Song, Shaozhou Pu and Liang Li are with the department of Engineering Physics, Tsinghua University & Key Laboratory of Particle and Radiation imaging (Tsinghua University), Ministry of Education, Beijing, 100084, China.

Corresponding Author email: lliang@tsinghua.edu.cn (Liang Li)

(~1keV). The material composition of scanned phantom is present in the bottom left of Fig. 1. It is worth mentioning that the High-Z element used in this paper is Gd. Next, combined with Fig.1, we will illustrate the reconstruction of three-mode imaging system as well as the new algorithm to calculate the linear attenuation coefficients at different energies.

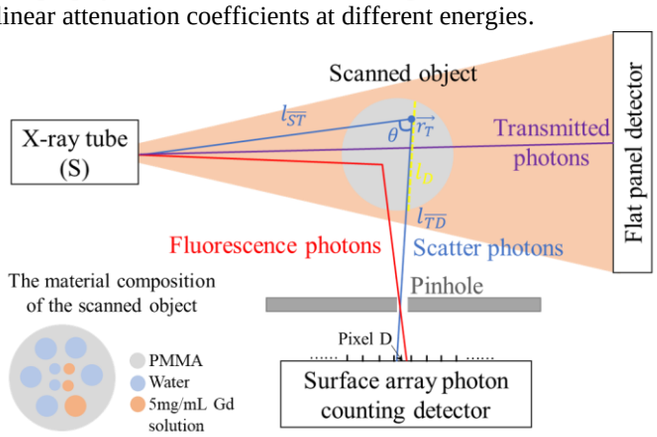


Fig.1 Schematic diagram of three-mode imaging system with XCT, XFCT, and CBCT. The bottom left is the specific configuration of scanned object.

A. *The reconstruction of three-mode imaging system*

The reconstruction of XCT and XFCT has been discussed in a lot of literature [2, 6], so the details of XCT and XFCT will not be presented in this paper. We will focus on CSCT reconstruction.

The Compton scatter signal intensity is as follows:

$$I_{CSCT}(D)=\int_{l_D} A(D,\vec{r}_T)\cdot\rho_e(\vec{r}_T)\,d\vec{r}_T \tag{1}$$

where $I_{CSCT}(D)$ is the Compton scatter signal intensity collected at a detector pixel D, l_D is the integral path marked as the yellow dashed line in Fig.1, $\rho_e(\vec{r}_T)$ is the electron density at position $\vec{r}_T$, $A(D,\vec{r}_T)$ is the system matrix of CSCT. The specific expression of $A(D,\vec{r}_T)$ is as follows:

$$A(D,\vec{r}_T)=\int_{E_{\min}}^{E_{\max}} I_0(E)\frac{d\sigma_{c,e}(E,\theta)}{d\Omega}\Delta\Omega_{TD}e^{-\int_{\overline{ST}}\mu(E,\vec{r})d\vec{r}}e^{-\int_{\overline{TD}}\mu(E_s,\vec{r})d\vec{r}}dE \tag{2}$$

where $E_{\min}$ is the minimum energy of incident photons, $E_{\max}$ is the maximum energy of incident photons, $I_0(E)$ is the intensity of incident photon at energy E, $\frac{d\sigma_{c,e}(E,\theta)}{d\Omega}$ is the differential cross section of Compton scatter, θ is the scatter angle, $\Delta\Omega_{TD}$ is the solid angle between $\vec{r}_T$ and detector pixel D, $\mu(E,\vec{r})$ is the linear attenuation coefficient at position $\vec{r}$ and energy E, E_s is the energy of scatter photons corresponding to the incident energy E, $l_{\overline{ST}}$ is the line connecting the X-ray tube S and $\vec{r}_T$, $l_{\overline{TD}}$ is the line connecting $\vec{r}_T$ and detector pixel D.

Scatter angle θ and solid angle $\Delta\Omega$ could be calculated when the geometric relationship is fixed. E_s could be calculated through Compton scattering principle when E and θ are fixed. $I_0(E)$ could be calculated by SpekCalc software [9], $\frac{d\sigma_{c,e}(E,\theta)}{d\Omega}$ could be calculated through KN formula. For (1) and (2), only $\rho_e(\vec{r}_T)$, $\mu(E,\vec{r})$, and $\mu(E_s,\vec{r})$ are unknown.

When the size of object is small, we could use the approximation of $\mu(E,\vec{r})\approx 0$. After approximation, the value of $A(D,\vec{r}_T)$ is known. Then, we could calculate the $\rho_e(\vec{r}_T)$ through ML-EM algorithm.

B. *Calculating the linear attenuation coefficients at different energies*

Then, we will use the reconstruction results of XCT, XFCT and CSCT to calculate the linear attenuation coefficients at different energies and realize the attenuation correction of XFCT.

The formula of linear attenuation coefficient at position $\vec{r}_T$ and energy E is as follows:

$$\mu(E,\vec{r}_T)=\frac{x(\vec{r}_T)}{E^3}+\rho_{Gd}(\vec{r}_T)\mu_{Pe,Gd,m}(E)+\rho_e(\vec{r}_T)\sigma_{c,e}(E) \tag{3}$$

where $x(\vec{r}_T)$ is an unknown parameter only related to the position $\vec{r}_T$, $\mu_{Pe,Gd,m}(E)$ is the photoelectric-effect mass attenuation coefficient of Gd at energy E, $\sigma_{c,e}(E)$ is the Compton scatter cross section for one electron at energy E, $\rho_{Gd}(\vec{r}_T)$ is the concentration of Gd at position $\vec{r}_T$. The meaning of $\rho_e(\vec{r}_T)$ is the same with (1). $\rho_{Gd}(\vec{r}_T)$ and $\rho_e(\vec{r}_T)$ could be calculated through XFCT and CSCT respectively. $\mu_{Pe,Gd,m}(E)$ could be found in NIST. $\sigma_{c,e}(E)$ could be calculated through KN formula. Therefore, only $x(\vec{r}_T)$ is unknown and we need to get this value.

For XCT, the integral formula is as follows:

$$\frac{I}{I'}=\int_E w(E)\cdot e^{-\int_L\mu(E,\vec{r})d\vec{r}}dE \tag{4}$$

where I is the signal intensity of the flat panel detector with placing the object, I' is the signal intensity of the flat panel detector without placing the object, L is the line of X-ray passing through the object, and $w(E)$ is the spectral weighting function, which is an intrinsic property of the flat panel detector. In (4), I, I' and $w(E)$ are known and only $\mu(E,\vec{r})$ is unknown. Taking (3) into (4), we could calculate the value of $x(\vec{r}_T)$. Then $\mu(E,\vec{r})$ could be solved.

It is worth noting that the value of both $\rho_{Gd}(\vec{r}_T)$ and $\rho_e(\vec{r}_T)$ have errors, because we did not perform attenuation correction. Then, the value of $\mu(E,\vec{r})$ calculated also has error. An iterative approach is proposed to solve this problem. We calculate $\mu(E,\vec{r})$ first. Then we update the values of $\rho_{Gd}(\vec{r}_T)$ and $\rho_e(\vec{r}_T)$ through the calculated value $\mu(E,\vec{r})$. Repeat these two steps until the value of $\mu(E,\vec{r})$ converges.

After calculating the $\mu(E,\vec{r})$ exactly, we could perform attenuation correction for XFCT. At the same time, we could

also eliminate the CSCT reconstruction error caused by the approximation of $\mu(E,\vec{r})\approx 0$.

III. Results

A. Parameters setup

In the phantom experiment, for XFCT and CSCT scan, the X-ray tube voltage is 150kVp and the tube current is 3mA. The projections are acquired with the step of 8° and 45 projections are collected. The exposure time for each projection is 60 seconds. For the CBCT scan, the X-ray tube voltage is 50kVp and the tube current is 1.4mA. The projections are acquired with the step of 0.5° and 720 projections are collected.

In the numerical simulation, the parameters of X-ray tube, phantom, geometric and detector are all same with experiment except for X-ray tube current. We set the number of incident photons to be a little large to simulate the scenario of low noise.

B. Simulation results

Fig. 2 shows the reconstruction results of XCT and CSCT. Fig.2 (a) is the reconstruction result of XCT. Fig.2 (b) is the reconstruction value of electron density. Fig.2 (c) is the true value of electron density. Fig. 2 (d) is the true value and calculated value of electron density corresponding to the red profile line. Comparing Fig. 2 (a) and Fig. 2 (b), we find we could distinguish water and PMMA more easily through CSCT than XCT. Therefore, the introduction of CSCT indeed provide structural information and make up for XCT. In addition, according to Fig. 2 (b-d), it can be seen that the quantitative reconstruction result of electron density is accurate. The relative error of electron density reconstruction result is 3.61%.

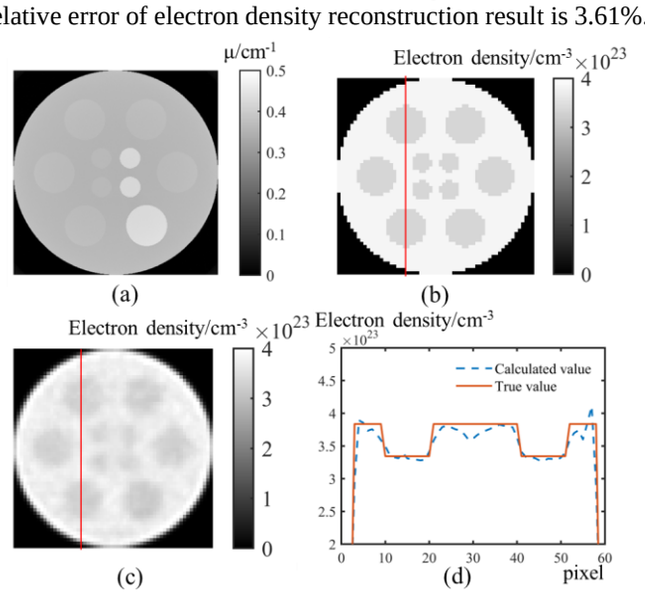


Fig.2 Simulation reconstruction results. (a) The reconstruction results of XCT. (b) The true value of electron density. (c) The quantitative reconstruction results of CSCT. (d) The true value and reconstructed value of electron density corresponding to the red profile line.

Next, we will display the results of $\mu(E,\vec{r})$. In theory, we could calculate the linear attenuation coefficients under arbitrary energy according (3). Due to space limitation, we only show the calculated result of linear attenuation coefficient at 40keV representatively. Fig. 3 (a-b) are the true value and calculated value of linear attenuation coefficient at 55keV respectively. Fig. 3 (c) is the true value and calculated value of linear attenuation coefficient at 55keV corresponding to the red profile line. Fig. 3 illustrates the feasibility of our proposed algorithm to calculate $\mu(E,\vec{r})$. The relative error between true value and calculated value of linear attenuation coefficient under 55keV is 2.73%.

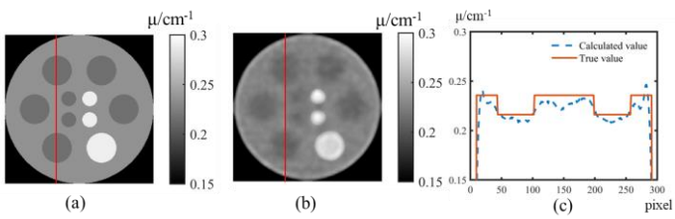


Fig.3 (a) The true value of linear attenuation coefficient at 55keV. (b) The calculated value of linear attenuation coefficient at 55keV. (c) The true value and calculated value of linear attenuation coefficient at 55keV corresponding to the red profile line.

After solving the $\mu(E,\vec{r})$, the attenuation correction of XFCT could be realized. Fig. 4 (a) shows the true value of Gd concentration. Fig. 4 (b) shows the reconstruction result of XFCT without attenuation correction. Fig. 4 (c) shows the reconstruction result of XFCT with attenuation correction. Fig. 4 (d) is the true value and calculated value of Gd concentration corresponding to the red profile line. Based on Fig. 4, we can see that the effect of attenuation correction is very significant. Through introducing CSCT, we realize the attenuation correction of XFCT and improve the accuracy of XFCT reconstruction without other prior information.

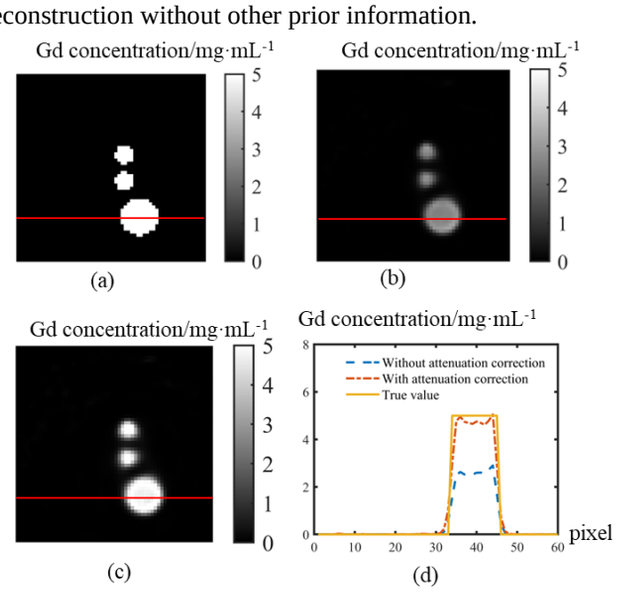


Fig.4 (a) The true value of Gd concentration. (b) The reconstruction result of Gd concentration without attenuation correction. (c) The reconstruction result of Gd concentration with attenuation correction. (d) The true value and calculated value of Gd concentration corresponding to the red profile line.

C. Experiment results

Fig. 5 is the scenario of the phantom experiment, which is corresponding to Fig. 1. Fig. 6 shows the reconstruction results in the phantom experiment. Fig. 6 (a) shows the reconstruction result of CSCT. Fig. 6 (b) shows the reconstruction result of XFCT. Fig. 6 (c) shows the reconstruction result of XCT. Fig. 6

(d) is the linear attenuation coefficient of scanned phantom at 55keV. Since the reconstruction result of CSCT in the phantom experiment is noisy, we use non-local means algorithm to denoise the reconstruction result of CSCT. According to the results in Fig. 6, we achieve the three-mode imaging and the calculation of linear attenuation coefficients at different energies on the real system.

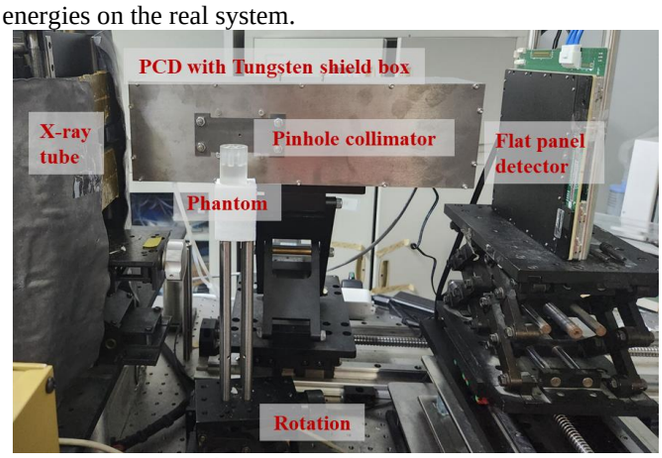


Fig.5 The scenario of phantom experiment.

Comparing Fig. 6 (a) and Fig. 6 (c), it is easier to distinguish water and PMMA by CSCT than by XCT, which is consistent with the deduction obtained by simulation. This phenomenon also indicates the significance of introducing CSCT. For the XFCT results after attenuation correction, we calculate the Gd solution concentration by averaging the values in all red circles of Fig. 6 (b). The calculated Gd concentration is 4.7mg/mL, which is very close to the true value 5mg/mL .

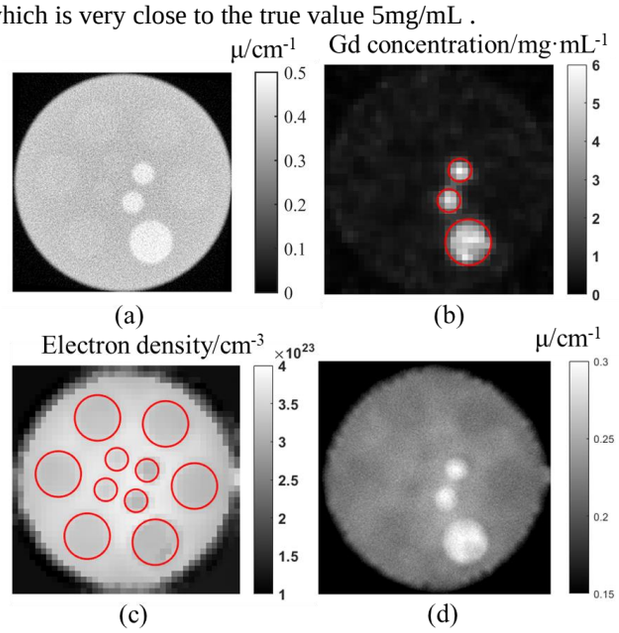


Fig.6 Phantom experiment reconstruction result. (a) The reconstruction results of XCT. (b) The reconstruction results of XFCT. (c) The reconstruction results of CSCT. (d) The calculated value of linear attenuation coefficient at 55keV.

IV. Conclusion

In this paper, we design a new three-mode imaging system of CBCT, XFCT and CSCT. The system feasibility is verified through both simulation and experiment. Compared with the traditional dual-modality system with XCT and XFCT, the new system could achieve the reconstruction of electron density and the attenuation correction of XFCT without any other prior information. In the future, we want to further improve the performance of the three-mode imaging system and apply this system in *vivo* for pre-clinical studies.

References

[1] N. Manohar, F. J. Reynoso, P. Diagaradjane *et al.*, "Quantitative imaging of gold nanoparticle distribution in a tumor-bearing mouse using benchtop x-ray fluorescence computed tomography," *Scientific reports*, vol. 6, no. 1, pp. 22079, 2016.

[2] S. Zhang, L. Li, J. Chen *et al.*, "Quantitative imaging of Gd nanoparticles in mice using benchtop cone-beam X-ray fluorescence computed tomography system," *International journal of molecular sciences*, vol. 20, no. 9, pp. 2315, 2019.

[3] L. Deng, S. Yasar, M. F. Ahmed *et al.*, "Investigation of transmission computed tomography (CT) image quality and x-ray dose achievable from an experimental dual-mode benchtop x-ray fluorescence CT and transmission CT system," *Journal of X-ray science and technology*, vol. 27, no. 3, pp. 431-442, 2019.

[4] S. Vedantham, S. Shrestha, A. Karellas *et al.*, "A framework for optimizing micro-CT in dual-modality micro-CT/XFCT small-animal imaging system." pp. 158-164.

[5] T. Kim, W. S. Lee, M. Jeon *et al.*, "Dual imaging modality of fluorescence and transmission X-rays for gold nanoparticle-injected living mice," *Medical Physics*, vol. 50, no. 1, pp. 529-539, 2023/01/01, 2023.

[6] L. Li, S. Zhang, W. Zhang *et al.*, "Full-field in vivo imaging of nanoparticles using benchtop cone-beam XFCT system with pixelated photon counting detector," *Physics in Medicine & Biology*, vol. 68, no. 3, pp. 035020, 2023.

[7] G. Ruben, I. Pinar, J. M. C. Brown *et al.*, "Full field X-ray Scatter Tomography," *IEEE Transactions on Medical Imaging*, vol. 41, no. 8, pp. 2170-2179, 2022.

[8] A. Brunetti, R. Cesareo, B. Golosio *et al.*, "Cork quality estimation by using Compton tomography," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, vol. 196, no. 1-2, pp. 161-168, 2002.

[9] G. Poludniowski, A. Omar, R. Bujila *et al.*, "Technical Note: SpekPy v2.0—a software toolkit for modeling x-ray tube spectra," *Medical Physics*, vol. 48, no. 7, pp. 3630-3637, 2021/07/01, 2021.

Properties of an optimized estimator for attenuation at very low count rates

Scott S. Hsieh and Paurakh Rajbhandary

***Abstract*—The conversion of raw detector signal into attenuation pathlength is ordinarily performed by inverting the Beer-Lambert law, which prescribes a logarithmic transform. At low signal, however, the curvature of the logarithm induces a statistical bias. When the detector signal is zero or negative, decisions must be made on how to replace this data (e.g., by clipping), which introduces a second form of bias. Improvements to the Beer-Lambert law have been proposed to reduce logarithm-induced bias, but the optimal solution that corrects for both logarithm-induced bias and clipping-induced bias remains unknown. We pose this question as a convex optimization problem and directly solve for the first 100 elements of the optimal unbiased estimator. We numerically define "unbiased" as satisfying a minimum bias of 0.001 and constrain the estimator to be numerically unbiased across a continuum of 1000 different operating points. We prove that a universally unbiased estimator does not exist: for sufficiently small average detector signals (or sufficiently attenuating objects), unbiased estimation is mathematically impossible. Therefore, it is necessary to specify a minimum on the mean detected signal. For a photon counting detector model, we show that when the mean detected signal is guaranteed to be at least 2 counts, the optimal unbiased estimator is very similar to the correction suggested by Li et al [*IEEE Trans Med Img* 42:6, 2023]. When the minimum mean detector signal is further reduced below 1 count, the optimal estimator oscillates, increasing variance. Together, these results suggest that for a photon counting detector: (1) when arbitrarily thick objects are allowed, an unbiased estimator for attenuation does not exist; (2) when a maximum thickness is imposed, numerically unbiased estimators are not unique; (3) as the maximum thickness increases and clipping becomes common, all inverses become non-monotonic, creating a variance penalty that propagates to reconstructed images.**

***Index Terms*—convex optimization, unbiased estimation, Beer-Lambert Law**

I. Introduction

THE estimation of attenuation pathlength is the first step of CT reconstruction. The most common practice is to invert the Beer-Lambert law. Assuming monoenergetic radiation for convenience and assuming a photon counting detector, the Beer-Lambert Law states that:

$$I = I_0 e^{-\int \mu dx} \tag{1}$$

Let us define the attenuation pathlength t as

$$t := \int \mu dx, \tag{2}$$

Solving directly for t, we find that

$$t = -\log(I) + \log(I_0)\,. \tag{3}$$

These equations work well when the area of detector pixels was large, when less attention was paid to radiation dose, and when the minimum dose was primarily constrained by electronic noise limits on the detector. Modern CT scanners have used 1 mm^2 pixels, lower doses, and larger patient sizes. In this context, bias from employing Equation (3) has been observed (1-8). This bias is especially large in spectral CT tasks, because a difference between two datasets is necessary to produce material cancelled images such as iodine maps or virtual noncontrast images. These difference images amplify small biases and reduce quantitative accuracy. Finally, bias is expected to grow increasingly large with newer, photon counting CT scanners because the characteristic pixel area may be reduced by as much as an order of magnitude (7).

It should be understood that low signal bias consists of two components, which have been described before as log-induced bias (8) and clipping-induced bias (1):

Log-induced bias arises from the curvature of the logarithm function itself, and can also be understood as a failure of first-order statistics in error propagation. That is, in ordinary error propagation, we can model the effects of noise by performing a first-order Taylor expansion about the estimator function (in our case, the logarithm). When this linear expansion is accurate, then the bias is zero because the error contributions of below-mean inputs cancel with error contributions of above-mean inputs. However, the logarithm function has high curvature at low signal. A linear expansion is no longer valid. This results in a bias.

Clipping-induced bias arises when the detected signal is nonpositive. For photon counting detectors, this specifically occurs when the detector reads out zero counts. For energy-integrating detectors, this also includes negative signals (i.e., when the detected signal is less than the average of dark current) or very small positive signals, which could lead to very large and nonphysical values according to Equation (3). It is generally necessary to apply "clipping" to such inputs. The

S. S. Hsieh (e-mail: Hsieh.Scott@mayo.edu) is with Mayo Clinic, Rochester, MN 55905. P. Rajbhandary is with Stanford University.

.

exact form of clipping is implementation-dependent. For the purpose of this work, we will assume that the system selects a minimum value I_{clip} such that whenever $I < I_{clip}$, we replace I with I_{clip}.

A number of corrections have been proposed to deal with low signal bias (4, 6, 8). Among recent work, we highlight the correction suggested by Li et al. for log-induced bias in photon counting detectors (3, 7). Li analytically derived an improved formula:

$$t = -\log(I) + \log(I_0) - \frac{1}{2I} + \frac{1}{12I^2} \tag{4}$$

Here, I is the integer number of detected counts, which follows a Poisson distribution in the low signal regime assuming pileup can be neglected. This polynomial correction addresses log-induced bias, but clipping-induced bias is still present. In the original work, Li used $I_{clip} = 0.3$ empirically. Later, Bushe et al. extended Li's correction by testing various I_{clip} values and refitting the polynomial correction to identify a choice that empirically reduces bias (2).

Li's correction is attractive because of its simplicity and because it analytically eliminates log-induced bias, but the question remains: what is the best approach of mitigating clipping-induced bias? While the strategy of Bushe is empirically effective, better solutions may be possible. In this work, we seek to answer these questions directly using optimization.

II. Methods

A. *Convex optimization*

In direct optimization, we seek to construct the estimator through finding each of the constituent elements. We consider only the case of photon counting first, although our methods could be extended to energy integrating detectors also. In this case, the attenuation pathlength estimator is understood as a function that maps from the space of natural numbers (representing photon counts) to the space of real attenuation pathlengths. While the space of natural numbers is infinite, we need only concern ourselves with the flux levels expected in CT. Furthermore, at very high flux levels (say, above 100 counts), we already know that the existing logarithm estimator, perhaps with Li's correction, will be very accurate.

Stated mathematically, we seek to select y, a vector of length 100:

$$y_k, \qquad k = 0, 1, 2 \ldots 99 \tag{5}$$

… such that attenuation pathlength has both minimum error and is unbiased. That is, we seek for a set of attenuation pathlengths t_j that the squared error is minimized (equivalent to minimizing the root-mean-square error):

$$\text{minimize} \sum_j E\left(\left(y - t_j\right)^2 \right) \tag{6}$$

subject to the constraints that the estimator is also unbiased:

$$E\left(y \middle| t_j\right) = t_j$$

The expectation value can be further expanded as follows:

$$\sum_k P_{kj}\, y_k = t_j$$
$$P_{kj} = \frac{\lambda^k e^{-\lambda_j}}{k!}$$
$$\lambda_j = I_0 e^{-t_j}$$

This expresses the expectation in terms of a probability matrix P_{kj} that reflects a Poisson distribution.

In practice, we allow for a small amount of numerical tolerance, so we select a working definition of "numerically unbiased" to have a tolerance of 0.001:

$$\left|E\left(y \middle| t_j\right) - t_j\right| \leq 0.001 \tag{7}$$

The cost function of Equation (6) together with the constraints of Equation (7) define the optimization problem. If we used exact equality instead of allowing a numerical tolerance, one possible solution strategy is to solve with Lagrange multipliers. With the finite width of the tolerance, solutions can be derived from interior point methods, which in practice converged within about 20 iterations. We implemented this using the CVX

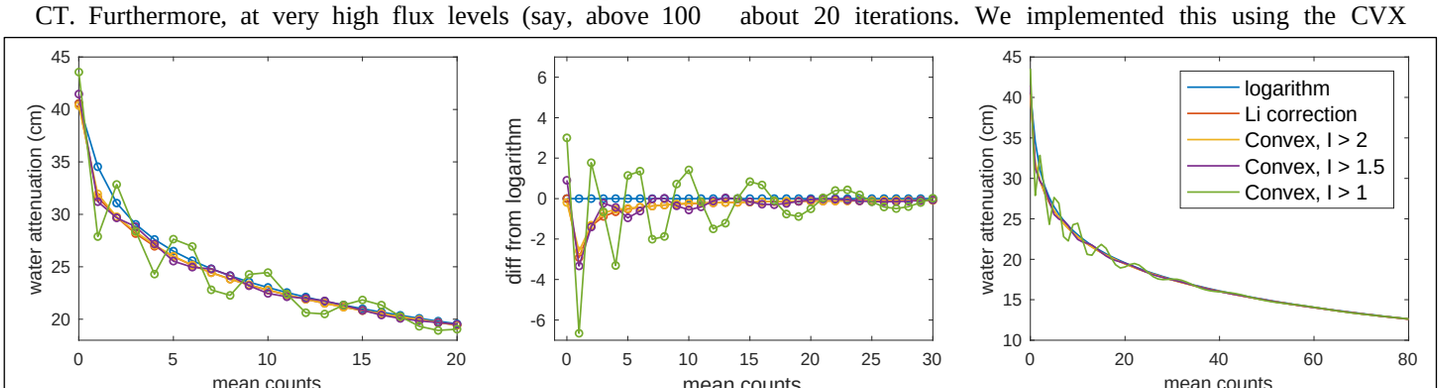


Figure 1. Estimators produced by convex optimization, compared to the simple logarithm estimator and the logarithm estimator with Li's correction. The three convex plots have different maximum attenuations for attenuation, corresponding to average mean counts of 2, 1.5, and 1 counts. All plots assume $I_0 = 1000$ counts. The subplots show differences over different domains.

package and the Mosek solver (9).

A limitation of our numerical formulation is that it becomes necessary to select certain values of t_j. Theoretically, we desire that t_j be the continuous set of pathlengths between 0 and ∞, but as we will see later, large pathlengths for t_j cause problems. We selected 1000 values of t_j that span from some starting to ending attenuation, where the starting attenuation corresponds to 100 mean counts on the detector and the ending attenuation corresponds to 1.0, 1.5, or 2.0 mean counts on the detector.

Because of Poisson noise, more than 100 counts could be received on the detector even when the mean counts was at most 100. One approach is to simply use y_k from the logarithm estimator in these cases. We chose to optimize for y_k spanning from 0 to 140, with the understanding that the behavior for large values of y might be corrupted by the low values of the probability density function. Another possible approach would have been to add more constraints that sample still thinner objects.

B. Images

After constructing the optimal attenuation pathlength using convex optimization, we were interested in understanding the manifestation in images. Therefore, we conducted numerical simulations.

We simulated a parallel beam CT scanner with monoenergetic radiation. The phantom being scanned had a width of 20, 25, or 30 cm and I_0 was 1000 counts before any object. This corresponds to a low-dose CT protocol (120 kV, 125 mA, 0.1 ms readout time) with a 0.2 mm pixel pitch, but we used a pitch of 1 mm to improve computational speed (but with expected equivalent bias characteristics). We used forward projection to create a noiseless sinogram, and sampled noisy realizations using a pure Poisson noise model. We repeated this 1000 times and averaged the result to better isolate the bias.

The noisy raw data was then converted to attenuation path length measurements using one of four estimators: (1) the standard logarithm estimator with $I_{clip} = 0.3$, (2) the logarithm estimator with Li's correction, also using $I_{clip} = 0.3$, (3) the optimized estimator, with the thickest value of t_j corresponding to 2 mean counts; (4) the convex optimized estimator, with the thickest value of t_j corresponding to 1 mean count.

III. Results

A. Non-existence proof

We first briefly prove by contradiction that a universal unbiased estimator does not exist. Suppose such an estimator did exist, such that for any attenuation pathlength T, the expected value of the estimator after Poisson noise would also be T. Let us select an extremely large value of T, such that the

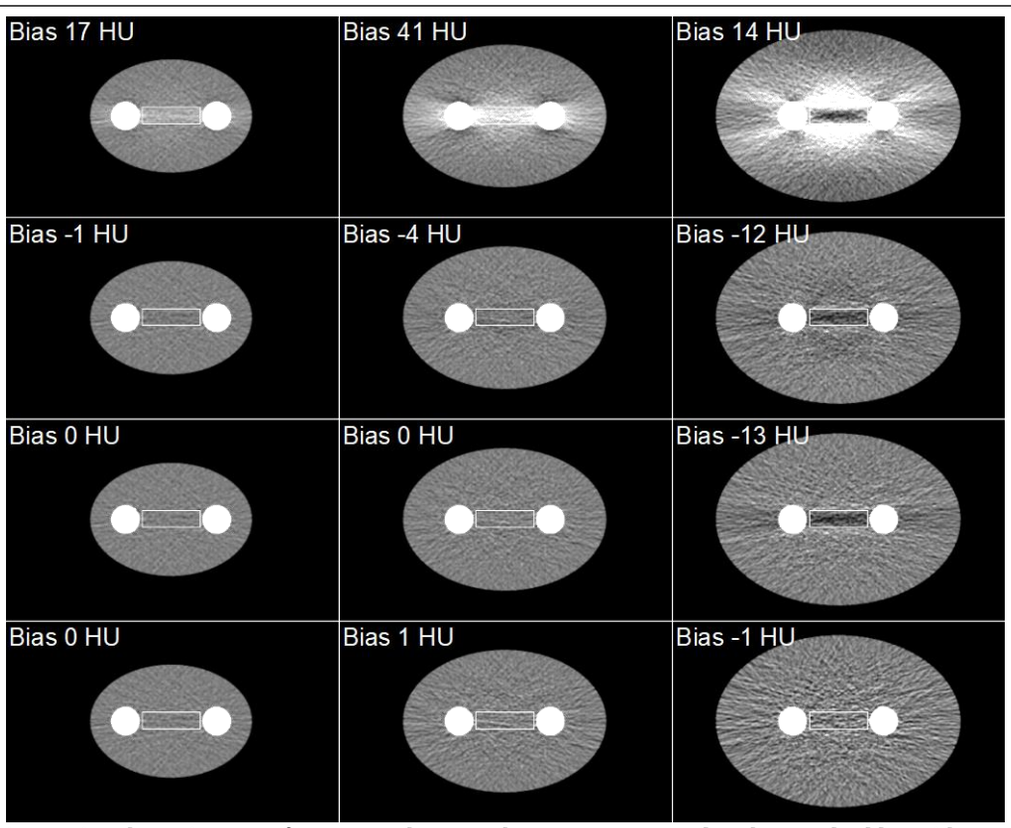


Figure 2. Simulations using the estimators: from top to bottom, the rows correspond to the standard logarithm estimator, the logarithm estimator with Li's correction, the convex estimator with minimum mean intensity of 2 counts, and the convex estimator with minimum mean intensity of 1 count. From left to right, the object thickness is 20 cm (minimum signal, 5.6 counts), 25 cm (minimum signal, 2.0 counts), and 30 cm (minimum signal, 0.8 counts). An average of 1000 noise realizations is used to reduce variance. Bias is calculated from a rectangular region at the center of the phantom.

detector is nearly guaranteed to sample 0 counts. To ensure that the estimator is unbiased, we must have that $y_0 = T$. On the other hand, let us consider another attenuation $2T$. We must have now that $y_0 = 2T$. Because y_0 cannot be both T and $2T$, we conclude that an unbiased estimator cannot exist.

B. Optimization functions

Because of the non-existence of a universal unbiased estimator, it is necessary to place an upper limit on the maximum possible attenuation. Figure 1 shows three different optimal constructed estimators, with maximum attenuation corresponding to 2, 1.5, or 1 mean count. The estimator terminating at 2 mean counts (orange line) is very similar to the logarithm estimator with Li's correction. Note that the similarity at 0 counts is coincidental with the empirical choice of $I_{clip} = 0.3$ that was selected by Li. The estimators terminating at 1.5 and 1 mean counts (purple and green lines, respectively) begin to deviate from Li's correction. In particular, the green estimator is not even monotonic. One expects that more photons should always map to less attenuation, but this is not true with these estimators.

C. Image reconstructions

Figure 2 propagates these estimators into image reconstructions. The greatest attenuation is in the lateral direction, passing through both bones, with minimum average detector signal of 5.6, 2.0, and 0.8 counts for the left, center, and right objects, respectively.

The bias of the standard logarithm estimator for the left, 20 cm left object is a positive streak connecting the two bones. Note that these are scatter-free, monoenergetic simulations so that the only source of bias is from the logarithm estimator. For the thickest object, the bias becomes complicated and shows the interplay of log-induced bias (which is positive) and clipping-induced bias (which is negative), as described previously (8). The listed bias of 14 HU in the region of interest is misleading because the positive and negative portions of the bias cancel.

The bias of Li's correction is strongest for the rightmost, 30 cm object, corresponding to -12 HU. A small amount of bias is present for the center 25 cm object. Because Li's correction eliminates log-induced bias, only clipping induced bias remains.

The bias of the convex optimization estimator assuming I > 2 (third row) is negligible for the left and center objects, where the average detector signal is always at least 2 counts. For the rightmost object, where the detector signal is smaller, residual bias remains that is almost equal to Li's correction.

Finally, the bias for the convex optimization estimator for I > 1 is nearly zero for all objects. However, recall that this estimator was non-monotonic and had some oscillations. This is expected to amplify image noise, and this is apparent from these images: more noise is present in exchange for decreased bias.

IV. Discussion

The bias-variance tradeoff is known to occur in other fields such as machine learning, and it is also present here in attenuation estimation. In this case, our estimator is constrained to be numerically unbiased in a given operating range, and the price for extending that operating range to successively higher attenuations (or smaller detector signals) is that the optimal estimator oscillates and amplifies variance. From these results we can also infer that the set of numerically unbiased estimators is not unique: the oscillating estimators found for I > 1.5 or I > 1 are also valid for the I > 2 case, but they were not selected because they do not minimize mean squared error.

A possible alternative strategy is to pose this problem in a machine learning framework. Our specific problem (mapping a single measurement of detector signal to an estimate of attenuation pathlength) would not be improved with machine learning, but one could directly learn a high-resolution, less noisy image from a noisy or biased image. This could be accomplished by taking high dose images and injecting high amounts of Poisson noise, assuming that access to the raw data exists. Such an approach is likely to reduce noise and bias, but the impact on detectability is uncertain. For example, a large reduction in variance is likely, but it may not correspond to a proportionally large improvement in detection efficiency, as would be true for purely linear reconstruction.

While our tools were demonstrated in simulation only for photon counting detectors, they can be extended to energy-integrating detectors with models of electronic noise. The electronic noise can be convolved with the Poisson noise and inversion can proceed with the same optimization framework.

Our work has been demonstrated in simulation models only. In future work, it is desirable to demonstrate these estimators on experimental data and to more thoroughly characterize these bias-variance tradeoffs.

We acknowledge Norbert Pelc for helpful discussions.

References

1. Brown KM, editor Clipping-induced bias correction for low-dose CT imaging. 15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine; 2019: SPIE.
2. Bushe D, Zhang R, Chen G-H, Li K. Unbiased zero-count correction method in low-dose high-resolution photon counting detector CT. Physics in Medicine & Biology. 2023;68(11):115002.
3. Chen JR, Feng M, Li K, editors. Overcoming the challenges of inaccurate CT numbers in low dose CT. Medical Imaging 2022: Physics of Medical Imaging; 2022: SPIE.
4. Gomez-Cardona D, Hayes JW, Zhang R, Li K, Cruz-Bastida JP, Chen GH. Low-dose cone-beam CT via raw counts domain low-signal correction schemes: Performance assessment and task-based parameter optimization (Part II. Task-based parameter optimization). Medical physics. 2018;45(5):1957-69.
5. Griner D, Lei N, Chen GH, Li K. Correcting statistical CT number biases without access to raw detector counts: Applications to high spatial resolution photon counting CT imaging. Medical Physics. 2023.
6. Hayes JW, Gomez-Cardona D, Zhang R, Li K, Cruz-Bastida JP, Chen GH. Low-dose cone-beam CT via raw counts domain low-signal correction schemes: Performance assessment and task-based parameter optimization (Part I: Assessment of spatial resolution and noise performance). Medical physics. 2018;45(5):1942-56.
7. Li K, Chen JR, Feng M. Construction of a nearly unbiased statistical estimator of sinogram to address CT number bias issues in low-dose photon counting CT. IEEE Transactions on Medical Imaging. 2023.
8. Rajbhandary PL, Pelc NJ, editors. Statistical bias in material decomposition in low photon statistics region. Medical Imaging 2015: Physics of Medical Imaging; 2015: SPIE.
9. Grant MC, Boyd SP, editors. Graph implementations for nonsmooth convex programs. Recent advances in learning and control; 2008: Springer.

Dark-Field CT Reconstruction with a Large Field of View

Peiyuan Guo, Longchao Men, Li Zhang, Zhentian Wang

***Abstract*—X-ray dark-field imaging is sensitive to microstructures and particularly beneficial for lung imaging. Up to now, a human-scale dark-field CT prototype has been developed. Since dark-field signal is determined by the autocorrelation length, the dark-field signal of each voxel in the imaged object varies due to the change of autocorrelation length during the rotation in CT when the distance between each voxel and gratings keeps changing. This phenomenon is significant especially in a large FOV dark-field CT and will cause artefacts on the reconstruction image. An effective calibration method is to average each ray on the sinogram with its conjugate ray, which is valid when the dark-field signal is linear to the distance to gratings. In this study, we studied dependence of PMMA spheres' dark-field signal on the distance to gratings, and applied the calibration method to the sinogram acquired with our large-FOV dark-field CT system, removing the artefact on reconstruction image.**

***Index Terms*—CT, X-ray imaging, Dark-field imaging**

I. Introduction

X-ray dark-field imaging is a novel imaging method based on a grating interferometer. Compared with the conventional X-ray imaging using the attenuation of the beam caused by the imaged object, the dark-field imaging detects the small-angle scattering caused by the object [1], which means it's sensitive to objects containing micro-strucures in the scale of nano- and micrometers, such as alveoli in the lung. There is a wealth of studies focusing on its application in lung imaging, including radiography [2]–[5] and CT [6], [7]. Furthermore, the dark-field lung CT has reached the human scale [8], advancing this novel imaging method into clinical CT application.

Since the discovery of dark-field signal, numerous studies have focuesed on its physical origin [9]–[12]. The dark-field signal has been proved to follow the Beer-Lambert's law similar to conventional attenuation signal [9], laying a solid theoretical foundation for dark-field CT. A comprehensive interpretation of dark-field signal is modeling it as a function of the autocorrelation length [11], [12], which is determined by the grating period, the X-ray energy, and the postion of the object.

Based on the defination of autocorrelation length, one issue can be observed that the object's rotation in CT scanning will cause a change in the autocorrelation length, leading to a variation in its dark-field signal, which means the rotaional invariance doesn't hold in dark-field CT. This issue might not be significant in dark-field CT with a small field of view (FOV), but it cannot be ignored for the human scale and has to be calibrated to avoid artefacts [13].

There are two calibration methods proposed in [14], including iterative and analytical reconstruction. For iterative method, the dependence of dark-field signal on object position is measured in advance, and is considered in the system projection matrix. For analytical method, each ray in the sinogram is averaged with its conjugate ray, and conventional reconstruction methods can be applied. These calibration methods were validated through simulations. However, the iterative methods is time and computation consuming, and the analytical is theoretically unrigorous. In this study, we study lung's dark-field dependence on object position, and propose a reconstruction method for large-FOV dark-field CT based on this dependence. Up to now, we have performed large-FOV dark-field CT with PMMA spheres as the object, and applied the calibration method in [14], validating this method by reconstructing artefact-free dark-field image.

II. Method

A. Dark-field Imaging

A grating-based X-ray imaging system consists of three gratings apart from X-ray source and detector. The core of the system is the beam splitter grating called G1. When X-ray beam passes a grating, a periodic intensity pattern occurs downstream the beam path at a certain distance, caused by either Talbot effect [15] or geometrical propagation [16]. The distortion of this pattern caused by an object contains the information of its composition and structure. An optional source grating called G0 can be placed right close to the source to split the X-ray source into multiple linear sources, reducing the source size and improving its spatial coherence [17]. Since the periodic pattern has a scale of microns, it cannot be directly resolved with current detector whose pixel size is normally much larger. To solve this problem, an analyzer grating called G2 is placed right before the detector. It has the same period as the fringe pattern, and when G2 moves step by step perpendicular to its grating lamella, it blocks different parts of the fringes before the detector. Therefore, the intensity acquired by the detector varies at different step, forming a curve of intensity with the grating position.

When an object is placed in the beam path, it will absorb, refract, and scatter the incoming X-ray, leading to the reduction of amplitude, phase shift, and the reduction of visibility of the phase stepping curve. The dark-field signal is defined as the negative of logarithm of visibility with and without the object. As studied in [11] and [12], it can be modeled as a function of the autocorrelation length of the system:

$$DF(\xi) = \exp[\sigma t(G(\xi) - 1)] \tag{1}$$

where σ represents object's scattering cross section, t represents object's thickness, ξ represents the autocorrelation length, and $G(\xi)$ is the correlation function.

The autocorrelation length can be calculated as follows:

$$\xi = \begin{cases} \dfrac{\lambda d_{S,G_2}}{p_2}, \text{ object between } G_1 \text{ and } G_2 \\ \dfrac{\lambda d_{S,G_0}}{p_0}, \text{ object between } G_0 \text{ and } G_1 \end{cases} \tag{2}$$

where λ represents the wave length of the incoming X-ray, d_{S,G_2} represents the distance between the object and G2, and p_2 represents the period of G2, and so do the subscript 0 for G0.

Based on Eq. 2, it can be observed that dark-field signal is dependent on the object position, indicating the variance of dark-field signal with the rotation in the dark-field CT.

B. Calibration for Dark-Field CT

As proposed in [14], a function between a paper stripe sample's dark-field signal with its distance to G0 was measured. For iterative reconstruction, this function was used to calibrate the system projection matrix according to the position of each voxel in the object. The influence of the varying object position was thus considered pixelwise on the sinogram.

Since iterative reconstruction is time consuming, another method with less computation cost was proposed. It was shown in [14] that the paper stripe sample's dark-field signal was approximately linear to the distance to G0. Therefore, the variance of dark-field signal caused by rotation can be calibrated by averaging each pair of conjugate ray across the sinogram. The pair of conjugate ray is defined as two rays that pass through the same voxels of the object but from different direction due to the rotation symmetry in CT. For each voxel on a ray path, whatever its position is, its dark-field signal becomes euqivalent to the signal when it locates at the midpoint of the intersection line between the FOV and ray path after averaging with its conjugate ray.

For an equal-angle detector as used in the lung CT prototype [8] whose distance between the source focus is the same among all pixels, all the aforementioned midpoints have the same distance from G0 and G2, which means the dark-field signal becomes rotaional invariant through averaging, and makes the dark-field CT meet the requirement for reconstruction. Even for detectors with other types of geometry like flat pannel detector, the position variance of the euqivalent midpoint is much less significant than the voxels at the outer part of the object which is far from the rotation center, reducing the artefact to a negligible level.

However, this conjugate averaging method also has several drawbacks. First, the linear relation between dark-field signal and the object-grating distance in [14] can be attributed to the polychromatic of the X-ray source and a certain relationship between the autocorrelation length and the scale of microstructures in the object. This linear relation may not hold true when the scale of scattering microstructures is close to or even smaller than the autocorrelation length, as measured in [12]. Second, the conjugate ray only exists in fan-beam CT, which limits the usable acquisition geometry. If the dark-field CT reaches the cone-beam or helical CT, this method is no longer valid.

C. Our Current Work

To study the positional dependence of dark-field signal and the reconstruction of dark-field CT, we performed a large-FOV dark-field CT acquisition and applied the conjugate averaging calibration to it.

1) Experimental setup: The experiments were carried out on the geometrical projection system located in Tsinghua University as shown in Fig. 1 (a), whose detail is provided in [18]. In this system, G1 and G2 are stitched with 2 and 4 substrates respectively, providing a large horizontal FOV. The key grating parameters are listed in Table I.

TABLE I
KEY GRATING PARAMETERS OF OUR IMAGING PLATFORM.

Grating	Type	Period (μm)	Duty cycle
G0	Absorption	15.75	0.3
G1	Absorption	7.0	0.5
G2	Absorption	12.6	0.5

A rotation stage is placed between G1 and G2 to perform dark-field CT scan. During this experiment, the distance between X-ray source and isocenter (SID) was set to 980 mm, and the distance between source and detector (SDD) was 1200 mm.

A unique feature of our system is the horizontal layout of the gratings, which means the lamella of the gratings is perpendicular to the rotaion axis. Therefore, the gratings' sensitive vector [19] has a constant direction relative to the object during the rotation, while the vertical layout will cause a varying sensitive vector and thus another source of variant dark-field signal especially for objects with anisotropic microstructures [20].

2) Sample preparation: PMMA microspheres were selected as the imaged sample, whose dark-field signal is close to lung tissue [21]. Microspheres with a diameter of 200 and 400 μm were used, and the former one has a stronger dark-field than the latter at the same autocorrelation length. Each of them was contained in a plastic tube whose diameter was about 1.2 cm, as shown in Fig. 1 (b).

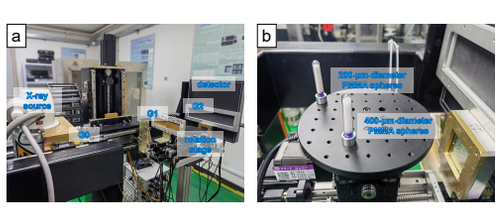


Fig. 1. (a) The dark-field imaging system and (b) two tubes of microspheres placed on the rotaion stage.

3) Acquisition protocol: Two experiments were carried out in this study. One is the radiography studying the positional dependence of microspheres' dark-field signal, and the other is the large-FOV dark-field CT. In both experiments, the X-ray source was operated at 80 kVp and 3.5 mA, the detector's integration time was set to 500 ms for single exposure, and a 9-step phase stepping was done by moving G2 vertically, with a step length of 1.4 μm.

In the radiography experiments, the dark-field signal of the microspheres was measured at different distance to G2, and at each position 10 repeated measurements were done to reduce noise, with the total exposure time of 5 s.

In the CT experiment, the tubes were mounted on the rotation stage as shown in Fig 1 (b). A 720-view rotation was done carrying the whole range of 360°. At each angular position, 3 repeated measurements were made with the total exposure time of 1.5 s. A background measurement without the object was performed after the CT acquisition.

4) Data processing: For the radiography experiments, at each position a region of interest (ROI) was selected. The dark-field and attenuation signal was retrieved through Fourier analysis [17], and averaged across the ROI. To eliminate the influence of thickness, the R value was calculated as the ratio between the dark-field and attenuation signal [22] to study the pure dependence of object distance.

For the CT experiment, the dark-field and attenuation signal was retrieved. The central slice of the beam was selected and was applied filtered back projection (FBP) method. Due to the imperfect stitching of the substrates in G1 and G2, the gratings cannot be aligned simultaneously, and Moiré fringes will appear on the detector [23] and also on the dark-field retreived. During the removal of the objects between the CT scan and the background scan, the fringes shifted due to slight vibration of the gratings, so fringes remained on the signal retrieved, which would cause ring artefacts in reconstruction. Since the Moiré fringes were rather stable during acquisition, the fringe artefacts could be removed on the sinogram with commonly used ring artefact suppression methods.

III. Results and Discussion

A. Positional Dependence of Dark-Field Signal

The R value, or normalized dark-field signal of PMMA spheres with a diameter of 200 and 400 μm at different position was measured, as shown in Fig. 2. It can be observed that for both spheres, the R value could be fitted as a linear function with respect to the distance between the object and G2, which accorded with the result in [14] and [21]. This is because the autocorrelation length of our system ($< 1\mu$m) was much smaller than the diameter of microspheres, and the dark-field signal was measured in a polychromatic spectrum of the X-ray source. This linear dependence made the conjugate averaging method possible in the dark-field CT calibration.

B. Large FOV Dark-Field CT and calibration

The dark-field sinogram was retrieved as mentioned in Sec. II-C4. As shown in the red boxes in Fig. 3 (a) and (b), the dark-field signal of the same odject far from the isocenter varied significantly in rotation due to the changing distance to G2, and averaging each pixel on the sinogram with conjugate ray relieved the signal variance as shown in Fig. 3 (c).

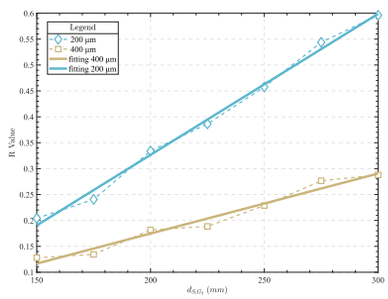


Fig. 2. The R value (normalized dark-field) of microspheres with different diameter at different distance to G2.

If the sinogram was constructed without calibration, severe artefacts would appear on the reconstruction image especially far from the center, as shown in Fig. 3 (d), and this artefact could be removed through the conjugate averaging method in Fig. 3 (e).

IV. Conclusion

In this study, we have studied the dependence of PMMA spheres dark-field signal on its distance to G2, and validated the calibration method in [14] with a large-FOV dark-field CT experiment, removing the artefacts on the reconstruction image. In the future, we will comprehensively study the dependence of dark-field signal of lung alveoli on the autocorrelation length, and propose a reconstruction method for large-FOV dark-field CT based on this dependence.

References

[1] F. Pfeiffer, M. Bech, O. Bunk *et al.*, "Hard-x-ray dark-field imaging using a grating interferometer," *Nature Materials*, vol. 7, no. 2, pp. 134–137, Feb. 2008.

[2] K. Hellbach, A. Baehr, F. De Marco *et al.*, "Depiction of pneumothoraces in a large animal model using x-ray dark-field radiography," *Scientific Reports*, vol. 8, no. 1, p. 2602, Feb. 2018.

[3] K. Willer, A. A. Fingerle, W. Noichl *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, Nov. 2021.

[4] T. Urban, A. P. Sauter, M. Frank *et al.*, "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.

[5] F. T. Gassert, T. Urban, D. Pfeiffer *et al.*, "Dark-field chest radiography of combined pulmonary fibrosis and emphysema," *Radiology: Cardiothoracic Imaging*, vol. 4, no. 4, p. e220085, Aug. 2022.

[6] F. T. Gassert, R. Burkhardt, T. Gora *et al.*, "X-ray dark-field ct for early detection of radiation-induced lung injury in a murine model," *Radiology*, vol. 303, no. 3, pp. 696–698, Jun. 2022.

[7] A. Velroyen, A. Yaroshenko, D. Hahn *et al.*, "Grating-based x-ray dark-field computed tomography of living mice," *EBioMedicine*, vol. 2, no. 10, pp. 1500–1506, Oct. 2015.

[8] M. Viermetz, N. Gustschin, C. Schmid *et al.*, "Dark-field computed tomography reaches the human scale," *Proceedings of the National Academy of Sciences*, vol. 119, no. 8, Feb. 2022.

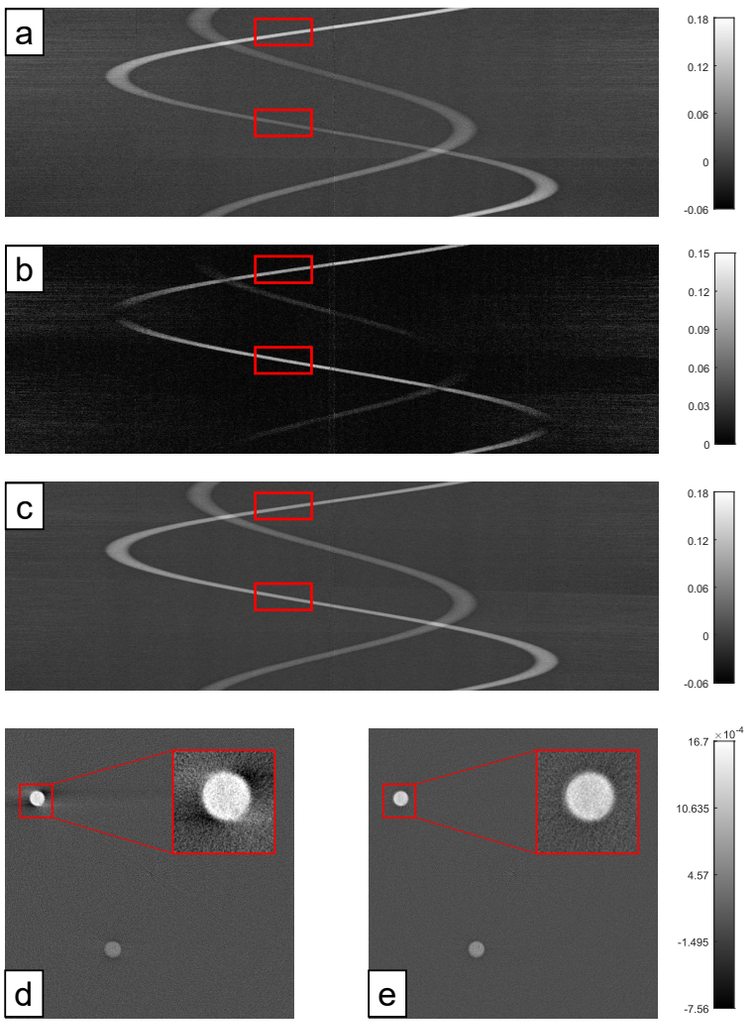


Fig. 3. (a) The dark-field sinogram after fringe removal. (b) The absolute difference between the sinogram and its conjugate one. (c) The sinogram after averaging each pixel and its conjugate ray. (d) The reconstruction image without calibration and (e) after calibration. Red boxes are drawn to indicate the regions of interest.

[9] Z.-T. Wang, K.-J. Kang, Z.-F. Huang *et al.*, "Quantitative grating-based x-ray dark-field computed tomography," *Applied Physics Letters*, vol. 95, no. 9, p. 094105, Aug. 2009.

[10] W. Yashiro, Y. Terui, K. Kawabata *et al.*, "On the origin of visibility contrast in x-ray talbot interferometry," *Optics Express*, vol. 18, no. 16, pp. 16 890–16 901, Aug. 2010.

[11] M. Strobl, "General solution for quantitative dark-field contrast imaging with grating interferometers," *Scientific Reports*, vol. 4, no. 1, p. 7243, Nov. 2014.

[12] F. Prade, A. Yaroshenko, J. Herzen *et al.*, "Short-range order in mesoscale systems probed by x-ray grating interferometry," *EPL (Europhysics Letters)*, vol. 112, no. 6, p. 68002, 2015.

[13] J. Haeusele, C. Schmid, M. Viermetz *et al.*, "Robust sample information retrieval in dark-field computed tomography with a vibrating talbot-lau interferometer," Oct. 2023.

[14] U. van Stevendaal, Z. Wang, T. Köhler *et al.*, "Reconstruction method incorporating the object-position dependence of visibility loss in dark-field imaging," in *Medical Imaging 2013: Physics of Medical Imaging*, vol. 8668. SPIE, Mar. 2013, pp. 260–270.

[15] A. Momose, S. Kawamoto, I. Koyama *et al.*, "Demonstration of x-ray talbot interferometry," *Japanese Journal of Applied Physics*, vol. 42, no. Part 2, No. 7B, pp. L866–L868, Jul. 2003.

[16] Z. F. Huang, K. J. Kang, L. Zhang *et al.*, "Alternative method for differential phase-contrast imaging with weakly coherent hard x rays," *Physical Review A*, vol. 79, no. 1, p. 013815, Jan. 2009.

[17] F. Pfeiffer, T. Weitkamp, O. Bunk *et al.*, "Phase retrieval and differential phase-contrast imaging with low-brilliance x-ray sources," *Nature Physics*, vol. 2, no. 4, pp. 258–261, Apr. 2006.

[18] X. Li, Z. Chen, L. Zhang *et al.*, "A preclinical large-field-of-view x-ray multi-contrast lung imaging prototype," in *Medical Imaging 2020: Physics of Medical Imaging*, vol. 11312. SPIE, Mar. 2020, pp. 312–319.

[19] A. Malecki, G. Potdevin, T. Biernath *et al.*, "Coherent superposition in grating-based directional dark-field imaging," *PLOS ONE*, vol. 8, no. 4, p. e61268, Apr. 2013.

[20] A. Malecki, G. Potdevin, T. Biernath *et al.*, "X-ray tensor tomography," *EPL (Europhysics Letters)*, vol. 105, no. 3, p. 38002, Feb. 2014.

[21] S. Spindler, D. Etter, M. Rawlik *et al.*, "The choice of an autocorrelation length in dark-field lung imaging," *Scientific Reports*, vol. 13, no. 1, pp. 1–11, Feb. 2023.

[22] Z. Wang, N. Hauser, G. Singer *et al.*, "Non-invasive classification of microcalcifications with phase-contrast x-ray mammography," *Nature Communications*, vol. 5, no. 1, p. 3797, May 2014.

[23] K. Deng, J. Li, and W. Xie, "Modeling the moiré fringe visibility of talbot-lau x-ray grating interferometry for single-frame multi-contrast imaging," *Optics Express*, vol. 28, no. 18, pp. 27 107–27 122, Aug. 2020.

curadon: an Open-Source Foundational Library for Highly Efficient Differentiable Projectors in X-ray Computed Tomography

David Frank*, Tobias Lasser*
Department of Computer Science, School of Computation, Information and Technology, Technical University of Munich, Germany
Munich Institute of Biomedical Engineering, Technical University of Munich, Germany

***Abstract*—We introduce *curadon* as an open-source foundational library for highly efficient differentiable projectors, designed to unite model-based and deep learning communities in the X-ray CT field. Positioned as a versatile building block for a large variety of X-ray CT applications, *curadon* offers efficient projection operators, a minimalistic interface, easy-to-use Python bindings, support for diverse geometries and differentiable operations. We demonstrate both the efficiency of *curadon* compared to other standard libraries, and its flexibility to work with real-world data. As such *curadon* is tailored for both classical analytic and model-based iterative approaches, as well as deep learning applications. We invite the X-ray CT community to contribute to *curadon*'s collaborative evolution, ensuring its continuous improvement and adaptation to emerging challenges.**

***Index Terms*—X-ray computed tomography, tomographic reconstruction, efficient projectors, deep learning, software framework, Python**

I. Introduction

X-RAY CT reconstruction has been undergoing a transformative shift towards data-driven methods, spurred on by the success and popularity of deep learning techniques. On the other hand, advancements in imaging hardware, such as multi-energy acquisitions, X-ray phase contrast and dark-field imaging, and non-standard acquisition techniques, require continuing research in how to optimally solve the associated inverse problems.

Numerous open-source frameworks and libraries exist to enable X-ray CT reconstructions, such as the ASTRA-toolbox [1], the Core Imaging Library [2], [3], TIGRE [4], elsa [5], and more. All these libraries provide easy-to-use Python interfaces and friendly installation procedures. However, integration into a deep learning pipeline, for example using PyTorch [6], is less straight-forward and user-friendly due to the complexity of interaction between these libraries that are typically used stand-alone. On the other hand, the most interesting deep learning based reconstruction approaches require differentiable projection operator layers, for example dual domain methods [7] or consistency checks to mitigate the impact of hallucinations [8]. The demand for tailored solutions for this community is illustrated by the popularity of the TorchRadon library [9], which provides fast and easy-to-use 2D projection operations for PyTorch.

The established open-source frameworks were mostly designed for classical reconstruction techniques, both analytical and model-based iterative approaches. A lot of focus was placed on usability in practical scenarios, supporting, for example, 3D cone beam geometries, non-flat detectors, or helical and other acquisitions trajectories. Another focus is computational efficiency, since high-resolution projection operations in 3D, i.e. forward and backward projections, in other words, evaluations of the discrete X-ray transform or its transpose, are very costly and require highly optimized massively parallel computing (e.g. via GPUs).

The deep learning community shares the requirement of efficient projection operations, but has been focusing more on lower resolution 2D data, as computational requirements for training complex deep network architectures are even higher. Consequently, efficient memory management on the GPU and use of low precision floating point arithmetic (FP8, FP16) play a key role here.

Unfortunately, efforts are currently mostly divided between these distinct ecosystems, leading to a duplication of core functionality. Despite these differences, both communities share a common requirement for fast and efficient projection operators. It is within this context that we introduce *curadon*, aiming to address the interoperability challenges by providing highly optimized low level projection operators for both classical analytic and model-based iterative approaches and deep learning approaches.

curadon is positioned as a foundational layer beneath both model-based and deep learning communities, comprising indispensable building blocks enabling a diverse array of applications in X-ray CT reconstruction. The design focuses on efficient memory management on GPUs, while providing highly efficient projection operators with support for 2D and 3D, fan and cone beam geometries, and arbitrary acquisition trajectories. Interoperability is provided via Python and C++ interfaces for easy integration as a differentiable layer in PyTorch as well as easy integration into existing reconstruction frameworks. A unique integration testing approach assures correctness. Thus, *curadon* is intended as a practical tool that streamlines workflows for researches and practitioners in the field.

You can find *curadon* at https://github.com/ner0-m/curadon.

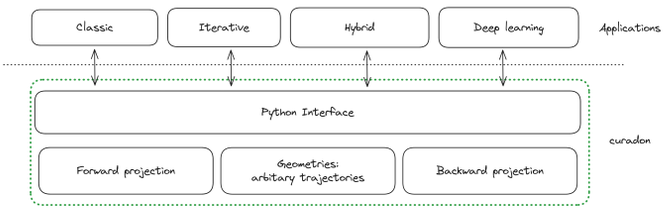


Fig. 1: `curadon` as a foundational layer for a diverse array of applications in X-ray CT reconstruction.

II. *curadon*: A Unifying Framework

At its core, *curadon* is conceived as a foundational layer intended to transcend the limitations of existing frameworks. Highly efficient projection operators supporting 2D and 3D as well as arbitrary acquisition trajectories, are paired with a simple Python and C++ interface, so that *curadon* not only addresses the shared need for fast and effective projection operations, but also lays the groundwork for collaborative and innovative endeavors.

While *curadon* is, at the time of writing of this abstract, at an early stage of development, the framework's vision extends beyond its current capabilities. Building on the experiences with our reconstruction framework elsa [5], we are very positive that we will be able to show-case all the major selling points at the conference. Rapid progress is underway, and while many features are still in the pipeline, we aim to provide a glimpse here into the core functionalities that distinguish *curadon*. As we continue our development efforts, additional features will be introduced, shaping *curadon* into a comprehensive and robust library. In the following, we will present the key features and design principles of *curadon*, showing early results on the ease of use, functionality, and performance.

A. Design principles

In adherence to its core design principles, *curadon* adopts a minimalist yet powerful approach in defining a framework. The primary focus centers on delivering essential functions for forward and backward projections, shaping the library into fundamental building blocks crucial for X-ray CT reconstruction, as shown in Figure 1. The principle of minimizing the library's surface area is synonymous with a commitment to simplicity. This approach ensures that users possess a compact and efficient library, easily integrated into both existing and new tools.

A notable feature arises from the intentional choice to avoid internal memory management within the library, placing the responsibility on the user for allocating device memory. This deliberate design decision not only fosters a more flexible and adaptable environment, but also empowers users. Those already storing data on the GPU can seamlessly integrate *curadon* into their workflows, eliminating inefficiencies. Consequently, the design philosophy of *curadon* revolves around simplicity, efficiency, and, most importantly, user empowerment. This user-centric approach guarantees a streamlined and effective experience for practitioners and researchers in the challenging domain of X-ray CT reconstruction.

Benchmark (mean op / s)	curadon	astra (± vs curadon)	torch-radon (± vs curadon)
forward 2d 256	2216.62	744.92 (-66.4%)	3482.05 (+57.1%)
forward 2d 512	718.70	362.07 (-49.6%)	973.25 (+35.4%)
forward 2d 1024	243.41	135.68 (-44.3%)	141.62 (-41.8%)
forward 2d 2048	57.91	35.72 (-38.3%)	23.80 (-58.9%)
backward 2d 256	2170.62	279.18 (-87.1%)	4025.01 (+85.4%)
backward 2d 512	435.41	81.32 (-81.3%)	1111.43 (+155.3%)
backward 2d 1024	109.90	20.88 (-81.0%)	269.67 (+145.4%)
backward 2d 2048	26.64	5.54 (-79.2%)	64.76 (+143.1%)

Fig. 2: Benchmarks of `curadon` compared to the ASTRA-toolbox and TorchRadon. The mean operations per second over 100 repetitions are shown. Each result is performed for different images sizes n^2. The size of the detector is $\sqrt{2}n^2$, with a circular trajectory using 360 equally spaced angles.

B. Efficient Projection Operators

curadon is engineered to meet the common demand for fast and efficient projection operators, a fundamental need shared by both model-based and deep learning reconstruction approaches. To show-case our commitment to efficiency, we focus here on the forward projection using Joseph's method in Figure 2. This comparison involves TorchRadon [9] and the ASTRA-toolbox [1], demonstrating *curadon*'s speed, especially for larger problems.

The benchmarks are performed on a machine equipped with an AMD Ryzen 9 5900HS (4.7 GHz), 16GB RAM, and an NVIDIA GeForce RTX 3050 Laptop GPU, using version 12.3 of the CUDA toolkit. Each benchmark runs a setup function, to setup volume and sinogram geometry information. The results in Figure 2 are the result of the average performance over 100 repetitions, with 5 warmup runs. Each configuration is run with the image size n^2, a detector resolution of $\sqrt{2}n^2$, and a circular trajectory using 360 equally spaced angles. For comparison, we used ASTRA version 2.1.0 [1] is used, with the commit `4907841` cherry-picked to ensure a successful build on CUDA 12.3. Further, version 2.0.4 of Daniel Ching's TorchRadon fork [10] is used.

The benchmark code is available in our repository https://github.com/ner0-m/curadon.

In both the forward and backward projection operators, certain geometric preprocessing steps are performed on the CPU. These steps calculate essential geometric values, such as the source position, according to the provided trajectory. These values are then uploaded to the device. Notably, the geometric assumptions for the forward and backward projection kernels differ significantly.

For the forward projection, efficiency is achieved by assuming the origin of the object of interest is at the world origin, with the source and detector rotation around the object based on the geometry. This approach facilitates the efficient computation of ray intersections with the object. Conversely, the backprojection kernel assumes the source and detector are along the z-axis, with the object rotated according to the geometry. Each voxel position is projected onto the x-y plane at z = detector-position. Leveraging findings from relevant papers [11], [12], we aim to enhance efficiency and explore further optimizations.

While many implementations use texture memory for fast shared read-only access in GPUs, *curadon* takes a similar but

different approach. Naive implementations relying solely on texture memory may encounter limitations in input data size. *curadon* addresses this by calling the kernel multiple times, mitigating the effects of size limitations. Future releases will progressively ease restrictions and provide projection operators for arbitrarily sized images.

To serve a broad community, *curadon* supports a variety of 2D and 3D X-ray CT setups. Presently, we accommodate arbitrary geometries for fan and cone beam setups, including rotations of the object and tilted detectors.

C. Simple Python Interface

To further enhance usability, *curadon* introduces a simple Python interface, enabling efficient and copyless data transfer from Python to C++. This interface streamlines the implementation of X-ray CT reconstruction tasks, providing a foundation for both analytical and iterative reconstruction frameworks.

Presently, *curadon* tightly integrates with PyTorch, allowing for the creation of simple differentiable functions for end-to-end deep learning tasks. This tight coupling ensures seamless integration with PyTorch-based workflows. However, in future development, we envision reducing this coupling to broaden compatibility with other deep learning frameworks such as TensorFlow [13] or JAX [14]. This step aims to provide users with the flexibility to choose their preferred deep learning environment, while leveraging the efficiency of *curadon*'s projection operations.

Listing 1 exemplifies the ease of using *curadon* in Python, demonstrating forward and backward projection operations. In this example, we leverage PyTorch to filter the sinogram and compute the filtered backprojection. The code is available in the repository's example folder, providing a hands-on illustration of *curadon*'s Python interface in action. Explore the repository https://github.com/ner0-m/curadon for an interactive experience and practical insights into the implementation of X-ray CT tasks using *curadon*.

To show-case reconstructions results, Figure 3 demos *curadon* on the 2DeteCT data set [15], in comparison to results using the ASTRA-toolbox and TorchRadon.

D. Integration testing

Testing projection operations for correctness, especially in the context of performance optimization, poses a unique challenge. While small and synthetic input sizes can be verified manually, they often fall short in testing a large number of projection positions and non-standard setups. This manual verification process is not only tedious but also prone to errors, and correctness in small cases may not necessarily translate to large input sizes.

To address this challenge, *curadon* takes a distinctive approach to testing. We rely on a combination of validation against ground truth and the original version of the forward projection. The primary validation employs perceptual hashes [16] to ensure similarity. Perceptual hashes, unlike cryptographic hashes, produce similar hashes for similar inputs. For the forward projection, we compare the perceptual hash of the Shepp-Logan phantom's analytical forward projection

```python
n = 256
vol_shape = np.array([n, n])

# assume returns a torch tensor
shepp_logan =
 ↪  generate_shepp_logan(vol_shape).cuda()

det_shape = int(n * np.sqrt(2))

DSO = 5000 # distance source to center
DSD = 5500 # distance source to detector

nangles = 360
angles = np.linspace(0, 2 * np.pi,
 ↪  nangles, endpoint=False)

# Define circular fan-beam geometry
geom = curadon.FanGeometry(DSD, DSO,
 ↪  angles, vol_shape, det_shape)

# Compute sinogram
sino = curadon.forward(volume, geom)

# Compute FBP, with Ram-Lak filter, with
 ↪  e.g. PyTroch
filt = get_filter(det_shape, "ram-lak")
fsino = apply_filter(sino, filt)

# Use curadon to backproject
fbp = curadon.backward(fsino, geom)
```

Listing 1: Example Python code to perform a forward and filtered backward projection using *curadon*.

(currently computed by elsa [5]) to our approximated forward projection. While we do not expect identical hashes due to the approximation, we anticipate them to be reasonably close.

In the case of backprojection, we compare the original Shepp-Logan phantom to the Filtered Backprojection (FBP) using the analytical sinogram as input. While not ideal, as we would prefer comparing the backprojection directly, the lack of a method to compute an analytical form of the backprojection necessitates this approach.

To the best of our knowledge, this testing methodology is unique within the context of X-ray CT reconstruction frameworks. We find it to be an elegant and effective solution, ensuring correctness and consistency in *curadon*'s projection operations.

III. Conclusion

We have introduced *curadon* into the X-ray CT reconstruction landscape as a unifying framework. Its foundational layer, featuring highly efficient projection operators and a simple Python interface, establishes a common ground for classical analytical and model-based approaches and deep learning approaches. We have show-cased the ease-of-use and its performance, and how to assure correctness.

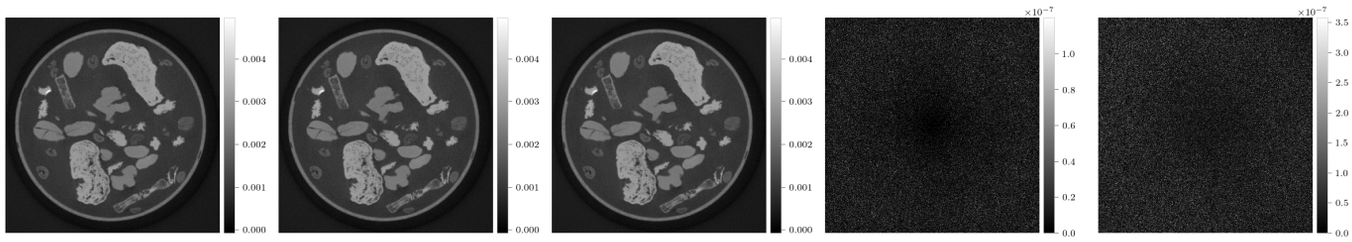


Fig. 3: FBP reconstruction of slice 12 of the 2DeteCT data set [15] using the full angular resolution (3600 angles) binned by a factor of 2 and a reconstruction size of 1024×1024. From left to right, reconstruction using the backprojection of `curadon`, TorchRadon, and the ASTRA-toolbox, followed by the absolute difference images between `curadon` and the other frameworks.

While *curadon* is in an early stage of development, ongoing efforts focus on optimizing the projection operations for even greater efficiency. Plans include providing first-class support for batched projections and mixed precision (including half precision floats), enhancing *curadon*'s utility in deep learning tasks. Expansion of supported X-ray CT geometries, particularly parallel beam geometries and non-flat detector setups, is on the horizon to broaden *curadon*'s applicability. Additionally, we are committed to ensuring ease of use, simplifying installation, and improving interoperability for seamless integration with other frameworks.

We invite researchers, developers, and practitioners to actively contribute to *curadon*'s evolution. Feedback is crucial as we shape *curadon* into a valuable tool for the larger X-ray CT community. Furthermore, we are actively reaching out to frameworks, seeking collaborations to incorporate *curadon* as a fundamental building block for tomographic reconstruction in X-ray CT.

References

[1] W. van Aarle, W. J. Palenstijn, J. Can't, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. Joost Batenburg, and J. Sijbers, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Optics Express*, vol. 24, no. 22, pp. 25 129–25 147, Oct. 2016. [Online]. Available: https://opg.optica.org/oe/abstract.cfm?uri=oe-24-22-25129

[2] J. S. Jørgensen, E. Ametova, G. Burca, G. Fardell, E. Papoutsellis, E. Pasca, K. Thielemans, M. Turner, R. Warr, W. R. B. Lionheart, and P. J. Withers, "Core imaging library - part i: a versatile python framework for tomographic imaging," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2204, p. 20200192, Jul. 2021. [Online]. Available: http://dx.doi.org/10.1098/rsta.2020.0192

[3] E. Papoutsellis, E. Ametova, C. Delplancke, G. Fardell, J. S. Jørgensen, E. Pasca, M. Turner, R. Warr, W. R. B. Lionheart, and P. J. Withers, "Core imaging library - part ii: multichannel reconstruction for dynamic and spectral tomography," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, no. 2204, Jul. 2021. [Online]. Available: http://dx.doi.org/10.1098/rsta.2020.0193

[4] A. Biguri, R. Lindroos, R. Bryll, H. Towsyfyan, H. Deyhle, I. E. k. Harrane, R. Boardman, M. Mavrogordato, M. Dosanjh, S. Hancock, and T. Blumensath, "Arbitrarily large tomography with iterative algorithms on multiple gpus using the tigre toolbox," *Journal of Parallel and Distributed Computing*, vol. 146, p. 52–63, Dec. 2020. [Online]. Available: http://dx.doi.org/10.1016/j.jpdc.2020.07.004

[5] D. Frank, J. Jelten, and T. Lasser, "An update to elsa - an elegant framework for tomographic reconstruction," in *17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2023, pp. 98–101. [Online]. Available: https://arxiv.org/abs/2310.16846

[6] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Z. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "Pytorch: An imperative style, high-performance deep learning library," *CoRR*, vol. abs/1912.01703, 2019. [Online]. Available: http://arxiv.org/abs/1912.01703

[7] T. Chesleren-Boghiu, F. C. Hofmann, M. Schulthei, F. Pfeiffer, D. Pfeiffer, and T. Lasser, "Wnet: A data-driven dual-domain denoising model for sparse-view computed tomography with a trainable reconstruction layer," *IEEE Transactions on Computational Imaging*, vol. 9, p. 120–132, 2023. [Online]. Available: http://dx.doi.org/10.1109/TCI.2023.3240078

[8] M. Patwari, R. Gutjahr, R. Marcus, Y. Thali, A. F. Calvarons, R. Raupach, and A. Maier, "Reducing the risk of hallucinations with interpretable deep learning models for low-dose ct denoising: comparative performance analysis," *Physics in Medicine & Biology*, vol. 68, no. 19, p. 19LT01, Oct. 2023. [Online]. Available: http://dx.doi.org/10.1088/1361-6560/acfc11

[9] M. Ronchetti. (2020) Torchradon: Fast differentiable routines for computed tomography. [Online]. Available: https://arxiv.org/abs/2009.14788

[10] D. Ching and M. Ronchetti, "Torchradon: Fast differentiable routines for computed tomography," https://github.com/carterbox/torch-radon, 2020.

[11] S. Chilingaryan, E. Ametova, A. Kopmann, and A. Mirone, "Balancing load of GPU subsystems to accelerate image reconstruction in parallel beam tomography," in *2018 30th International Symposium on Computer Architecture and High Performance Computing (SBAC-PAD)*. IEEE, Sep. 2018. [Online]. Available: http://dx.doi.org/10.1109/CAHPC.2018.8645862

[12] ——, "Reviewing GPU architectures to build efficient back projection for parallel geometries," *Journal of Real-Time Image Processing*, vol. 17, no. 5, p. 1331–1373, Jun. 2019. [Online]. Available: http://dx.doi.org/10.1007/s11554-019-00883-w

[13] Tensorflow Developers, "Tensorflow," 2023. [Online]. Available: https://zenodo.org/doi/10.5281/zenodo.4724125

[14] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leery, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, "JAX: composable transformations of Python+NumPy programs," 2018. [Online]. Available: http://github.com/google/jax

[15] M. B. Kiss, S. B. Coban, K. J. Batenburg, T. van Leeuwen, and F. Lucka, "2detect - a large 2d expandable, trainable, experimental computed tomography dataset for machine learning," *Scientific Data*, vol. 10, no. 1, Sep. 2023. [Online]. Available: http://dx.doi.org/10.1038/s41597-023-02484-6

[16] R. Biswas and P. Blanco-Medina, "State of the art: Image hashing," *CoRR*, vol. abs/2108.11794, 2021. [Online]. Available: https://arxiv.org/abs/2108.11794

Simultaneous Spectral Calibration and Image Reconstruction in Spectral CT with a Sum of Deltas Spectrum Model

Gregory Ongie, Buxin Chen, Zheng Zhang, Dan Xia, Emil Y. Sidky, & Xiaochuan Pan

***Abstract*—Reconstructing material specific images from spectral CT transmission data is a challenging non-linear inverse problem. Model-based iterative reconstruction (MBIR) algorithms, which rely on a computational forward model of the acquisition physics, have shown promise for solving this inverse problem. An important step for realizing materials decomposition with MBIR algorithms is accurately calibrating the spectral response of the imaging system, which is often difficult to achieve in practice. To mitigate this issue, we propose an "auto-calibrating" MBIR algorithm, where a parametric representation of the spectra is jointly estimated along with the materials decomposition in a unified optimization framework. Key to our approach is a recently proposed sum of deltas surrogate model for the spectral response that achieves accurate forward modelling while dramatically reducing the degrees of freedom. Our numerical experiments on simulated dual-energy breast CT data show that an accurate two-materials decomposition is achievable under spectral miscalibration using the proposed approach.**

***Index Terms*—Spectral CT, Dual-energy CT, Image Reconstruction, Materials Decomposition**

I. Introduction

Spectral CT is a promising technology that enables reconstruction of material specific images by acquiring measurements with several distinct photon energy spectra. The resulting materials decomposition can be used to perform quantitative imaging, reduce beam hardening artifacts, and enhance contrast-to-noise ratio relative to conventional CT.

However, reconstructing material specific images from spectral CT transmission data is a challenging non-linear inverse problem. Traditionally, reconstruction is performed by a two-step procedure that separates the problems of tomographic reconstruction and materials decomposition, but this approach can introduce significant artifacts in the presence of noise and incomplete data acquisition. As an alternative, "one-step" model-based iterative reconstruction (MBIR) algorithms have also been proposed [1], [2], [3], [4], which jointly perform reconstruction and materials decomposition in a unified optimization framework. MBIR algorithms rely on a computational forward model of expected photon counts at each detector pixel. This forward model incorporates an effective spectral response of the imaging system that combines both the source spectrum and detector response. Accurately determining these spectra is challenging in practice, and their miscalibration can severely impact the quality of materials decompositions.

To mitigate this issue, in this paper we propose an MBIR approach that allows for simultaneous estimation of the effective spectra and material decomposition. Our algorithm is based on an alternating minimization strategy that alternates between estimating a material decomposition given a fixed spectrum, and estimating the spectrum assuming the materials decomposition is fixed. We build on recent algorithic approaches designed to address each of these subproblems individually [4], [5].

However, this approach is generally ill-posed without additional constraints on the spectra. To reduce degrees of freedom in the spectrum model, we focus on a sum of delta-functions approximation. Previous work has shown just a few delta functions (e.g., 2–4) are sufficient to accurately approximate the spectrum for the purposes of materials decomposition [6]. We propose using fixed delta locations at each energy and optimizing over the delta weights.

Similar to this work, [7] also investigated an alternating minimization approach for joint spectral calibration and reconstruction. Our approach differs from [7] primarily in the spectral model: [7] attempts to fit a continuous spectral model by regularizing the distance of the estimated spectra from some initial guess of the spectra. In contrast, we do not impose any explicit regularization on the spectral weights besides a normalization constraint.

II. Methods

In this work, we focus on a dual-energy CT acquisition model. For a fixed projection view, we model the expected number of photon counts $\hat{c}_{i,k}$ acquired at detector pixel i and with a source potential of $k \in \{\text{high}, \text{low}\}$ as

$$\hat{c}_{i,k} = N \int_0^\infty s_k(E) \exp\left(-\sum_{b=1}^{B} y_{b,i} \mu_b(E) \right) dE \qquad (1)$$

where $y_{b,i}$ is the line integral of basis material b along the path between source and detector pixel i, $\mu_b(E)$ is the energy-dependent attenuation coefficient of basis material b, $s_k(E)$ is a photon energy spectrum that includes all source and detector effects normalized such that $\int_0^\infty s_k(E) dE = 1$, and N is the expected number of photons per detector pixel and

G. Ongie is with the Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI, USA e-mail: gregory.ongie@marquette.edu

B. Chen, Z. Zhang, D. Xia, E. Sidky, & X. Pan are with the Department of Radiology, University of Chicago, IL, USA

G.O. was supported by NSF CRII award CCF-2153371. B.C., Z.Z., D.X., E.S., & X.P., were supported in part by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health under award numbers R01EB023968 and R21CA263660.

projection measurement without the phantom in the beam path. Here, for simplicity, we assume the spectrum does not vary over detector pixels, i.e., for each tube potential k there is a common spectrum $s_k(E)$ for all detector pixels and all views.

Following [6], we consider an approximation of the spectrum s_k as a weighted sum of delta-functions:

$$s_k(E) = \sum_{d=1}^{D} w_{k,d}\, \delta(E - E_{k,d})$$

where $w_{k,d} \geq 0$, and where the number of δ-functions D is assumed to be small (e.g., $D \approx 2-4$). With this approximation, (1) reduces to a finite weighted sum of exponentials:

$$\hat{c}^{\delta}(\boldsymbol{y}_i, \boldsymbol{w}_k) := N \sum_{d=1}^{D} w_{k,d}\, \exp\left(- \sum_{b=1}^{B} y_{b,i} \mu_b(E_{k,d}) \right) \quad (2)$$

where $\boldsymbol{w}_k := [w_{k,1}, ..., w_{k,D}] \in \mathbb{R}^D$ denotes the spectral weights at energy k, and $\boldsymbol{y}_i = [y_{i,1}, ..., y_{i,B}] \in \mathbb{R}^D$ denotes the collection of line integrals of all basis materials acquired at detector pixel i. We propose using (2) as a parametric model for the spectrum, where the weights $w_{k,d}$ are allowed to vary, while the locations $E_{k,d}$ are assumed to be fixed.

We pose joint reconstruction of the basis material maps $\boldsymbol{X} = [\boldsymbol{x}_1 | \cdots | \boldsymbol{x}_B] \in \mathbb{R}^{M \times B}$ and the collection of spectral weight vectors $\boldsymbol{W} = [\boldsymbol{w}_1 | \cdots | \boldsymbol{w}_K] \in \mathbb{R}^{D \times K}$ as the optimization problem:

$$\min_{\boldsymbol{X}, \boldsymbol{W}} \ell(\boldsymbol{P}\boldsymbol{X}, \boldsymbol{W}) \;\; \text{s.t.} \;\; \boldsymbol{X} \in \mathcal{C}_{\mathcal{X}}, \boldsymbol{W} \in \mathcal{C}_{\mathcal{W}} \quad (3)$$

where $\boldsymbol{P}$ is a discrete ray projection operator and $\ell(\boldsymbol{Y}, \boldsymbol{W})$ is the Poisson negative log-likelihood loss

$$\ell(\boldsymbol{Y}, \boldsymbol{W}) = \sum_{v,i,k} \left[\hat{c}^{\delta}(\boldsymbol{y}_i^v, \boldsymbol{w}_k) - c_{i,k} - c_{i,k} \log\left(\frac{\hat{c}^{\delta}(\boldsymbol{y}_i^v, \boldsymbol{w}_k)}{c_{i,k}} \right) \right],$$

where $\boldsymbol{y}_i^v \in \mathbb{R}^B$ is a row of $\boldsymbol{Y}$ with v being the view index. Here, $\mathcal{C}_{\mathcal{X}}$ and $\mathcal{C}_{\mathcal{W}}$ are convex constraint sets that impose regularization on the basis materials maps, and constrain the spectral weights to be normalized, respectively. For this study, we assume the basis material maps $\boldsymbol{x}_b \in \mathbb{R}^M$ are constrained to be non-negative, have a fixed pixel sum, and have constrained anisotropic total variation. Additionally, we constrain the spectral weights $\boldsymbol{w}_k$ at each energy to be non-negative and sum to one, i.e., each $\boldsymbol{w}_k$ belongs to the probability simplex. Put formally,

$$\mathcal{C}_{\mathcal{X}} = \{\boldsymbol{X} \in \mathbb{R}^{M \times B} : \; \boldsymbol{x}_b \geq \boldsymbol{0}, \|\boldsymbol{x}_b\|_1 = \rho_b, \; \text{TV}(\boldsymbol{x}_b) \leq \gamma_b, \forall b\}$$
$$\mathcal{C}_{\mathcal{W}} = \{\boldsymbol{W} \in \mathbb{R}^{D \times K} : \boldsymbol{w}_k \geq \boldsymbol{0}, \|\boldsymbol{w}_k\|_1 = 1, \forall k\}$$

where ρ_b and γ_b are user-set parameters.

We propose solving the optimization problem (3) via an alternating minimization approach: starting from an initial guess of the spectral weights $\boldsymbol{W}^{(0)} = \boldsymbol{W}^{\text{init}}$, for $n = 0, 1, ...$ we iteratively update

$$\boldsymbol{X}^{(n)} = \underset{\boldsymbol{X} \in \mathcal{C}_{\mathcal{X}}}{\arg\min}\, \ell(\boldsymbol{P}\boldsymbol{X}, \boldsymbol{W}^{(n)}) \quad (4)$$

$$\boldsymbol{W}^{(n+1)} = \underset{\boldsymbol{W} \in \mathcal{C}_{\mathcal{W}}}{\arg\min}\, \ell(\boldsymbol{P}\boldsymbol{X}^{(n)}, \boldsymbol{W}) \quad (5)$$

until convergence.

To solve the $\boldsymbol{X}$-subproblem in (4), we use the recently proposed constrained one-step spectral CT reconstruction (cOSSCIR) algorithm [4]. The cOSSCIR algorithm is based on a variant of the Alternating Directions Method of Multipliers (ADMM) adapted to non-convex objective functions. To accelerate convergence, we warm-start all ADMM auxiliary and dual variables using their values from the previous iteration.

The $\boldsymbol{W}$-subproblem in (5) is convex and separable over energy levels k. In particular, the minimization over $\boldsymbol{W}$ decouples into independent convex optimization problems over each $\boldsymbol{w}_k$ under probability simplex constraints. Following the approach in [5], we apply the exponentiated gradient algorithm, which is specifically designed for optimization problems of this type.

For computational tractability, we run only finitely many iterations of the solver for each subproblem. Since we are not guaranteed to reach optimality after finite iterations, this approach is similar to a block coordinate descent strategy. In future work, we will explore the trade-offs of this strategy with a closer to exact alternating minimization strategy.

III. Results

We illustrate the proposed algorithm on a 2D breast CT simulation. The simulated breast phantom consists of three materials: adipose tissue, fibroglandular tissue, and calcifications. Adipose tissue and calcification are used as the two basis materials. The ground truth basis material maps are shown in the first column of Figure 3.

A dual-energy CT scan is simulated with the X-ray source set to either 50 and 80 kilovolts (kV). For the generation of the tissue sinograms, we implement a standard line-intersection model for the X-ray transform of a discrete image. The images are pixelized on a 512×512 grid with physical area $18\ \text{cm}^2$. The X-ray source follows a circular trajectory with a radius of 50 cm and centered on the middle of the image. The source-to-detector-center distance is 100 cm, and the detector is modeled as a linear array. The high and low kV sinograms are comprised of 512 equally distributed views over a 360 degree scan and the projections are sampled on a 1024-pixel detector.

Figure 1 shows the simulated spectra used in our experiments. We generate a "ground truth" continuous spectra using the SpekPy package [8], modeling a Tungsten anode and 1.6-mm Aluminum filtration. Energy weighting is included in order to model an energy-integrating detector. We also form an approximation of the continuous spectra by a delta-function model (2) with $D = 3$ delta's per source voltage. The delta locations were placed at photon energies representing the lower quartile, median, and upper quartile of the continuous spectra. The delta weighting factors were set by sampling the continuous spectra at the delta locations and normalizing them to sum to one. Finally, we also consider "miscalibration" of the delta-function approximation that assumes uniform weights on the deltas corresponding to each source voltage.

To evaluate the proposed algorithm, we consider three different settings: **no spectral model mismatch, no noise**, where noise-free high/low kV transmission data is simulated

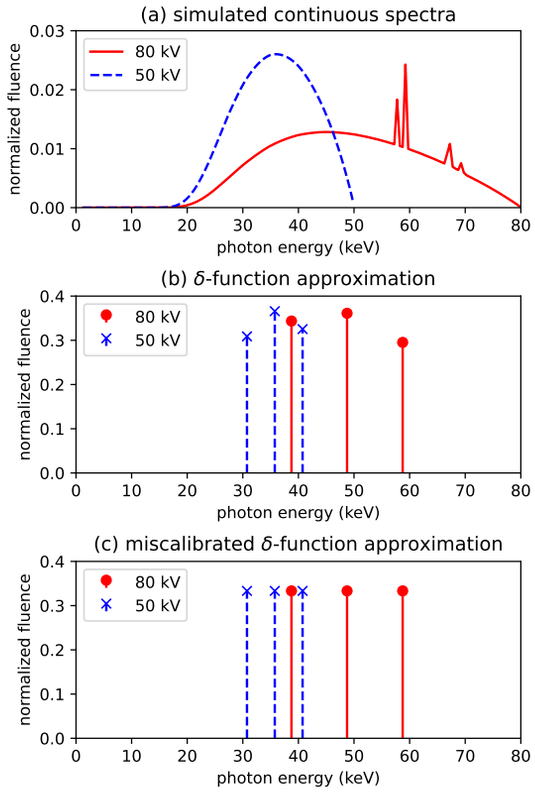


Fig. 1. (a) Continuous spectral response for high and low kilovolt source potentials simulated using SpekPy [8]. The spectra are generated modeling a Tungsten anode and 1.6-mm Aluminum filtration. Energy weighting is included in order to model an energy-integrating detector. (b) δ-function approximation of spectra used in the "no model mismatch" experiment. (c) Miscalibrated δ-function spectra used as initialization in all experiments.

according to the δ-function approximation model shown in Figure 1(b). **spectral model mismatch, no noise**, where noise-free high/low kV transmission data is simulated according to the continuous spectrum model shown in Figure 1(a). **spectral model mismatch, with noise**, where high and low kV transmission data is simulated according to the continuous spectrum model shown in Figure 1(a), with simulated Poisson noise assuming $N = 100,000$ photons per detector pixel.

To illustrate the simulated data in three settings above, in Figure 2 we show a filtered back projection (FBP) reconstruction of the negative log transmission data generated at high and low source potentials. Observe that the FBP images generated using the continuous spectra and delta-function approximation are visibly indistinguishable.

For the regularization constraints, in all settings we set $\rho_k = \mathrm{TV}(\boldsymbol{x}_b^*)$ and $\gamma_b = \|\boldsymbol{x}_b^*\|_1$, where $\boldsymbol{x}_b^*$ is the ground truth basis material map. This ensures the ground truth basis material maps are feasible solutions, and eliminates the need for parameter tuning. For the proposed approach, we run 50 alternating minimization iterations, each with 100 iterations of the cOSSCIR algorithm and 10 iterations of the exponentiated gradient algorithm. As a baseline, we also compare with the standard cOSSCIR algorithm using the miscalibrated spectra (Figure 1(c)) in the forward model, which we call the "no recalibration" approach. In this case, we run 5000 iterations of cOSSCIR. All algorithms were observed to reach convergence with this number of iterations.

Figure 3 shows our main results. We display reconstructions of the adipose basis map, calcification basis map, and a virtual monochromatic image (VMI) computed at 46.25 keV, along with plots of the absolute error, and report the root mean squared error (RMSE). In the first setting, with no spectral model mismatch, we see mild cupping artifacts in the adipose basis map with no recalibration, whereas the proposed approach gives nearly exact recovery (RMSE=0.0003). In the second setting, where there is now spectral model mismatch, the no recalibration approach shows worse cupping artifacts, which the proposed approach is able to mitigate. Finally, in the third setting where we additionally simulate noise in the transmission data, some fine detail is lost, but again we observe substantially lower error in the adipose basis map. Fewer differences are seen in the VMI images between the no recalibration approach and the proposed approach, which suggests that the primary benefit of the proposed approach is improved accuracy in the materials decompositions, though not necessarily in VMI reconstructions.

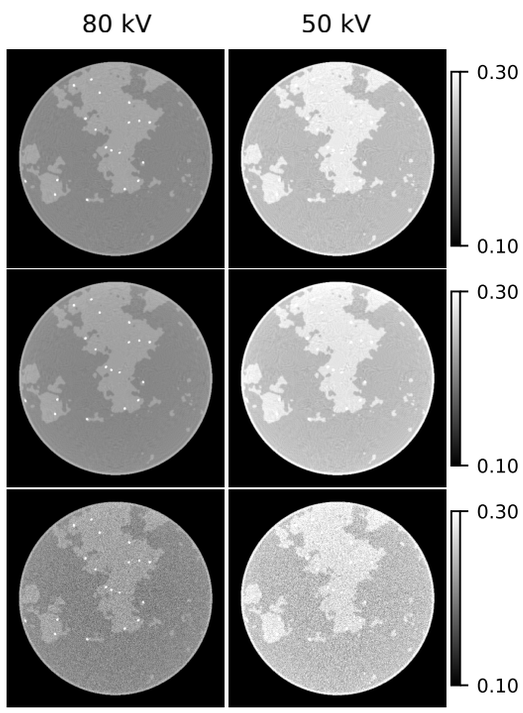


Fig. 2. FBP of negative logarithm processed transmission data according to different spectral models used in our experiments. **Top row:** δ-function approximation to continuous spectra, no noise. **Middle row**: continuous spectra, no noise. **Bottom row**: continuous spectra and simulated Poisson noise.

IV. Conclusion

We propose a new one-step MBIR algorithm to perform spectrum calibration and materials decomposition in spectral CT. In order to constrain the degrees of freedom of the spectra, we use a sum of delta functions model. Our experiments on simulated dual-energy breast CT data illustrate that the proposed approach has the potential to significant reduce artifacts due to spectral miscalibration.

While this study was focused on a dual-energy setup with registered views, in future work we plan to investigate the case

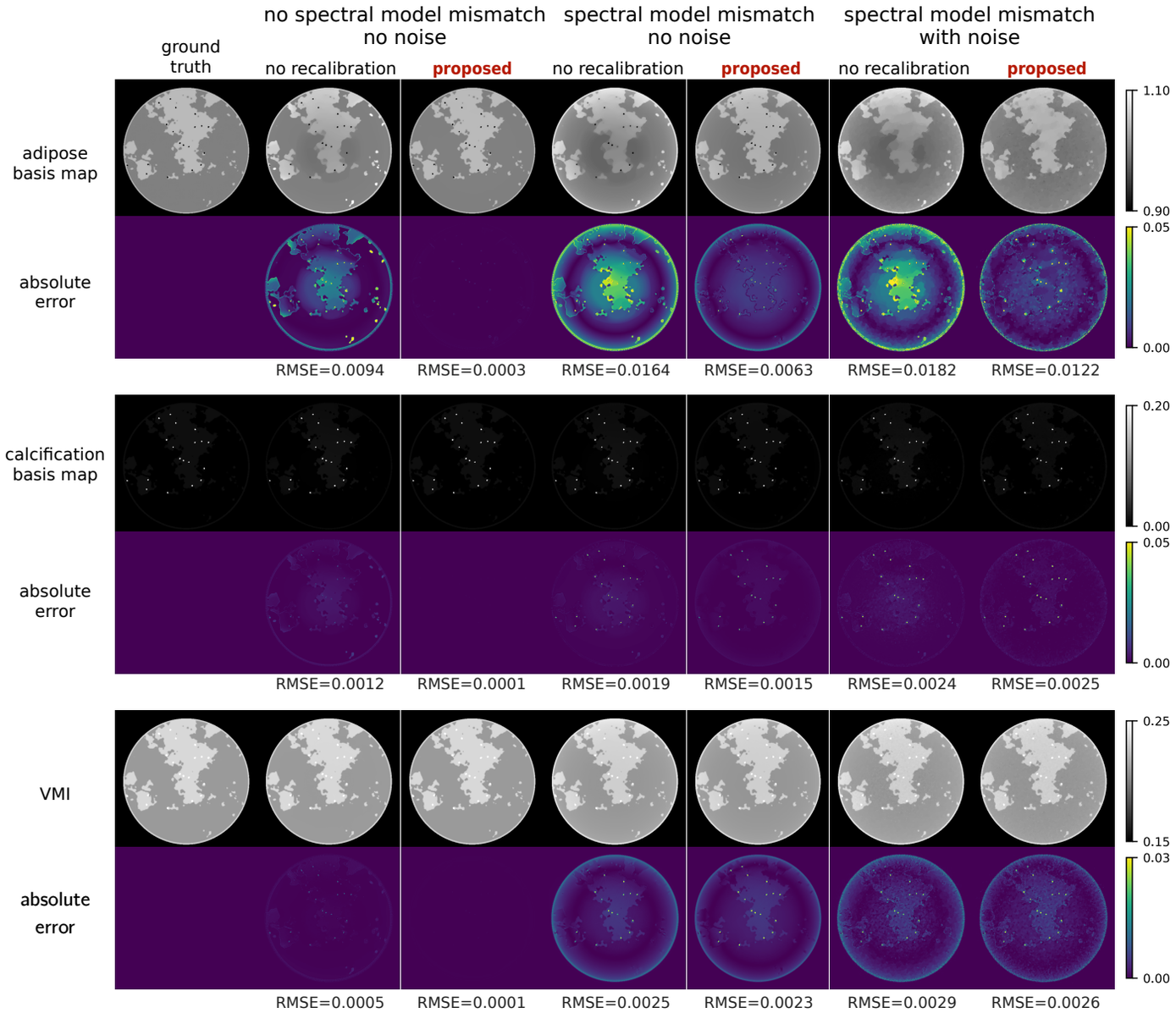


Fig. 3. Reconstruction of adipose and calcification basis material maps and a virtual monochromatic image (VMI) from simulated dual-energy breast CT data. The VMI is computed at 46 keV.

of unregistered views that arise when using fast kV-switching. Additionally, we plan to extend our model to allow for detector pixel specific spectral responses, which is essential for photon counting CT. Finally, we will explore the possibility of a more efficient algorithm that updates the basis material images and spectral weights at each iteration, rather than requiring alternating between two distinct solvers.

References

[1] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE Transactions on Medical Imaging*, vol. 21, no. 2, pp. 89–99, 2002.

[2] 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.

[3] B. Chen, Z. Zhang, E. Y. Sidky, D. Xia, and X. Pan, "Image reconstruction and scan configurations enabled by optimization-based algorithms in multispectral CT," *Physics in Medicine & Biology*, vol. 62, no. 22, p. 8763, 2017.

[4] T. G. Schmidt, R. F. Barber, and E. Y. Sidky, "A spectral CT method to directly estimate basis material maps from experimental photon-counting data," *IEEE Transactions on Medical Imaging*, vol. 36, no. 9, pp. 1808–1819, 2017.

[5] W. Ha, E. Y. Sidky, R. F. Barber, T. G. Schmidt, and X. Pan, "Estimating the spectrum in computed tomography via kullback–leibler divergence constrained optimization," *Medical Physics*, vol. 46, no. 1, pp. 81–92, 2019.

[6] K. Mechlem, S. Ehn, T. Sellerer, E. Braig, D. Münzel, F. Pfeiffer, and P. B. Noël, "Joint statistical iterative material image reconstruction for spectral computed tomography using a semi-empirical forward model," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 68–80, 2017.

[7] W. Ha, E. Y. Sidky, R. F. Barber, T. G. Schmidt, and X. Pan, "Alternating minimization based framework for simultaneous spectral calibration and image reconstruction in spectral CT," in *2018 IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC)*. IEEE, 2018, pp. 1–5.

[8] G. Poludniowski, A. Omar, R. Bujila, and P. Andreo, "SpekPy v2.0–a software toolkit for modeling x-ray tube spectra," *Medical Physics*, vol. 48, no. 7, pp. 3630–3637, 2021.

Impact of patient respiratory motion variability on respiratory-gated PET with CT-based attenuation correction

Paul E. Kinahan[1], Darrin W. Byrd[1], Paul Desari[2], W. Paul Segars[3], Ehsan Abadi[3], Katie Olivas[3], Ehsan Samei[3]

***Abstract*— Due to the variability of intra-patient respiratory motion patterns, CT acquisitions used in respiratory gated PET/CT may not match the PET scan data, resulting in biased attenuation correction that compromises the accuracy of the PET image. In this work, we evaluate the impact of respiratory motion variability through both the CT and PET acquisition chains using simulation studies. The respiratory motion patterns are derived from a library of 700+ patient chest motion patterns collected during clinical PET/CT scans. The chest motion patterns drive the respiratory motion of the XCAT phantom which is in turn used as an input for the DukeSim and ASIM scanner simulations for CT and PET. The reconstructed CT images are used for attenuation correction of the PET data, using the standard clinical method. Initial results indicate that the root mean square error (RMSE) in the estimated attenuation correction values are 61.6 ± 20.8 HU and the corresponding normalized RMSE in the PET images is 7.2% ± 4.2%, even with respiratory gating. There is, however, a considerable variation between simulated patients. These results are being expanded to use a larger selection of the respiratory motion pattern library, variations in patients sizes and habitus, and more detailed CT simulations. The goal is to provide a more accurate population-based understanding of the PET image measurement errors that occur in clinical practice and also as a basis for improved motion compensation methods.**

***Index Terms*— Respiratory gated PET/CT, cine CT, extended cine duration, attenuation correction, motion mismatch.**

I. Introduction

Positron emission tomography/computed tomography (PET/CT) imaging has emerged as a standard component for cancer diagnosis, staging, treatment planning, and assessing response to therapy. However, PET/CT imaging of the lung and abdomen is usually compromised by respiratory motion, where the attenuation correction error caused by mismatch between PET and CT images is one of the major effects. To reduce the mismatch, gated CT imaging has been evaluated to provide phase-matched respiratory-gated CT data for respiratory-gated PET attenuation correction. However, due to a broad variability of intra-scan motion and inter-PET-CT motion in practice [1,2], gated CT acquisitions may be inadequate to match PET scan data and consequently compromise the quality of attenuation correction [3]. Multiple methods have been proposed to reduce errors caused by motion mismatches, including data-driven methodologies, external-gating methodologies, and joint estimation methods [4]. However, prior evaluations of the ranges in error effects in phase-matched respiratory gated PET/CT attenuation correction have been largely anecdotal. In this study we are evaluating population-based estimates using a wide range of patients sizes and motions. These are used with an accurate biodistribution model and detailed physics models of the CT and PET scanners. The goals are to provide a robust understanding of the PET image measurement errors that occur in clinical practice and also reference for evaluating the impact of improved motion compensation methods.

II. Methods

A. Respiratory patterns

We acquired respiratory traces from patients who underwent whole body diagnostic PET/CT scans at the University of Washington Medical Center. The scans were performed on a General Electric (GE) Discovery STE PET/CT scanner (GE Healthcare, Waukesha, WI.). Most protocols The duration of the PET/CT scan was typically 30-45 minutes with seven minutes per bed position. Respiratory motion was monitored for each patient using the Real-Time Position Management (RPM) gating system that acquires a respiratory signal by utilizing a fixed infrared camera mounted to the imaging bed and a block with reflective markers that is secured to the upper abdomen of the patient [5]. Respiratory traces are time dependent recordings of relative anterior-posterior motion at the chest wall as reported by the RPM system and displacement is measured in centimeters. From this collection we have evaluated 10 patient-derived respiratory patterns as described below.

B. Simulation pipeline

We simulated FDG-PET/CT scans with the 4D extended Cardiac-Torso (XCAT) phantom [6,7] to generate 3D distributions of radiotracers and linear attenuation coefficients of the patient body, with spherical objects added to simulate lung tumors. The RPM traces were used to control the motion of the rendering of XCAT images with simulated tumors, with amplitudes of anterior-posterior expansion in XCAT images matching amplitudes of RPM displacements. The XCAT image can be used as inputs for both the DukeSim CT simulator [8] and the ASIM PET simulator [9] (Figure 1).

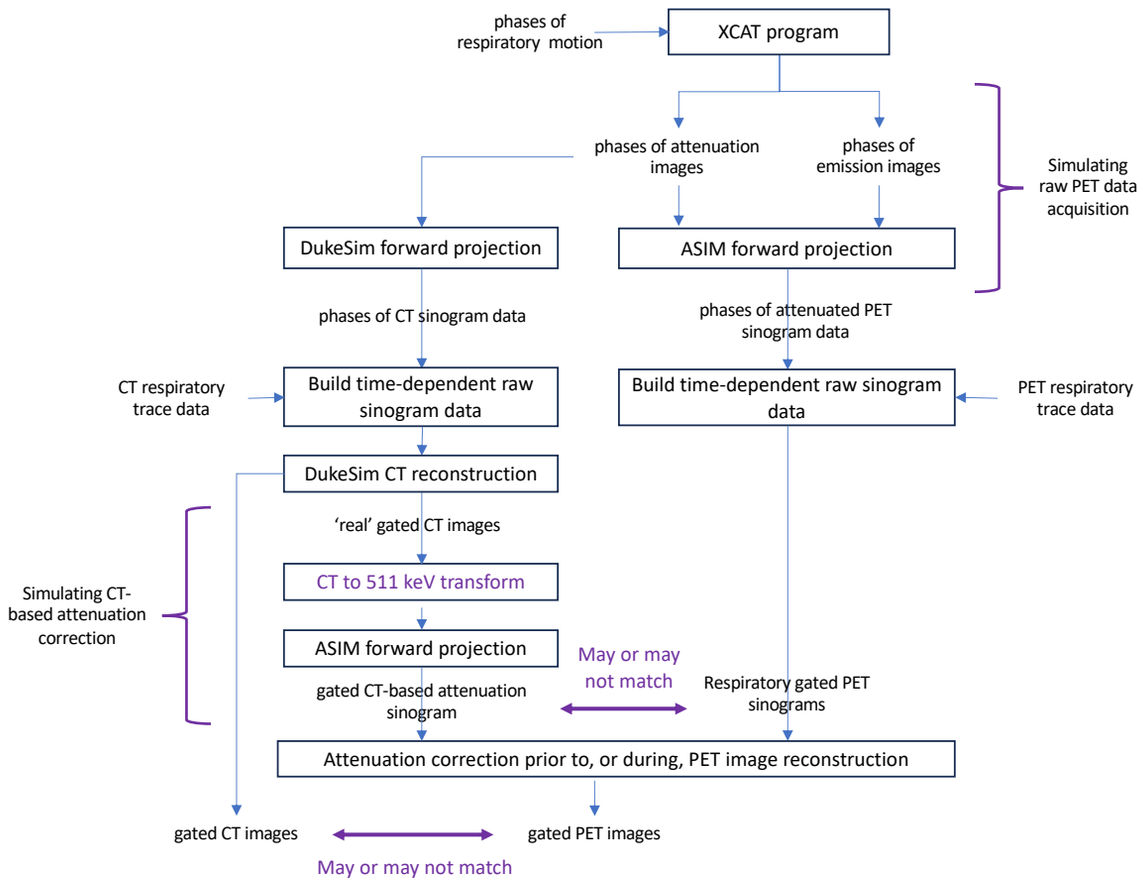


Figure 1. Simulation pipeline. XCAT = 4D extended Cardiac-Torso phantom [6,7]. DukeSim = CT simulator [8]. ASIM = PET simulator [9].

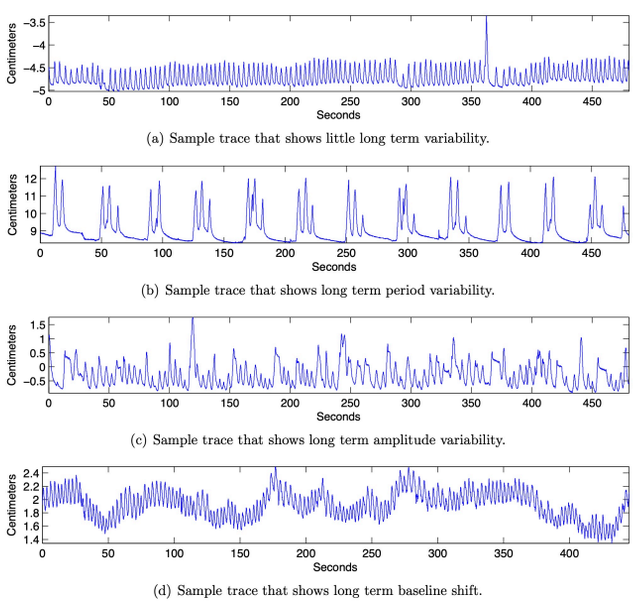
C. *Analysis*

For comparison, we calculated the root-mean-square-error (RMSE) between gated average attenuation anatomy images and associated cine CT images for each phase, which reveals the mismatch between PET and CT acquisition.

We also computed the RMSE between reference reconstruction using average anatomy as CTAC and reconstructions using subsequent cine CT images as CTAC, which indicates the accuracy of using cine CT images for PET attenuation correction. Note that the RMSE was evaluated within the support of the XCAT phantom with all surrounding air pixels excluded.

III. Results

Figure 2. Presents illustrative examples respiratory traces. Figure 3 shows the RMSE in attenuation images and PET uptake images, averaged over all respiratory phases. Initial results indicate that the root mean square error (RMSE) in the estimated attenuation correction values are 61.6 ± 20.8 HU and the corresponding normalized RMSE in the PET images is 7.2% ± 4.2%, even with respiratory gating. There is, however, a considerable variation between simulated patients.



(a) Sample trace that shows little long term variability.

(b) Sample trace that shows long term period variability.

(c) Sample trace that shows long term amplitude variability.

(d) Sample trace that shows long term baseline shift.

Figure 2. Sample respiratory traces that show various short term patterns.

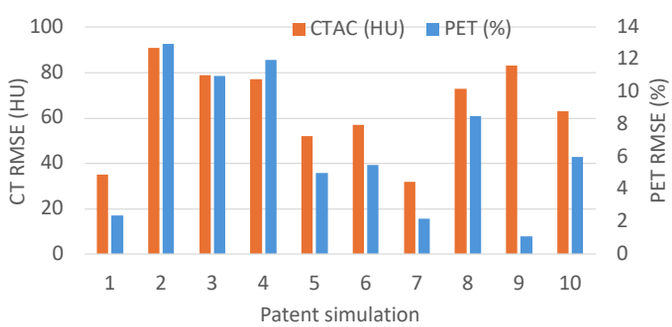


Figure 3. Average RMSE in attenuation images and PET uptake images..

IV. Discussion

The goal of this study is to provide a more accurate population-based understanding of the PET image measurement errors that occur in clinical practice and also as a basis for improved motion compensation methods. Our initial results of phantom simulation with real respiratory traces show that there is considerable inter-and intra-patient variability that is potentially not well-understood on a population basis. The results depend on the patient, with potentially greater needs for improvement for patients with irregular breathing pattern and larger diaphragm movement. These results are being expanded to use a larger selection of the respiratory motion pattern library, variations in patients sizes and habitus, and more detailed CT simulations.

Acknowledgments

This work is supported in part by NIH grants R01CA258298, R01EB001838, and P41EB028744.

References

[1] *Kinahan, P., Wollenweber, S., Alessio, A., Kohlmyer, S., MacDonald, L., Lewellen, and T., Ganin, A., "Impact of respiration variability on respiratory gated whole-body PET/CT imaging," J. Nucl. Med. 48(supplement 2), 196P (2007).*

[2] *Liu, C., Pierce, L.A.II, Alessio, A.M., and Kinahan, P.E., "The impact of respiratory motion on tumor quantification and delineation in static PET/CT imaging, " Phys. Med. Biol. 54(24), 7345–7362 (2009).*

[3] *Chi, P.C., Mawlawi, O., Nehmeh, S.A., Erdi, Y.E., Balter, P.A., Luo, D., Mohan, R., and Pan, T., "Design of respiration averaged CT for attenuation correction of the PET data from PET/CT, " Med. Phys. 34(6), 2039-2047 (2007).*

[4] *Lamare F, Bousse A, Thielemans K, Liu C, Merlin T, Fayad H, Visvikis D. PET respiratory motion correction: quo vadis. Physics in Medicine & Biology. 2022;67.*

[5] *Vedam SS, Kini VR, Keall PJ, Ramakrishnan V, Mostafavi H, Mohan R. Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker. Med Phys. 2003;30:505-513. PMID: 12722802.*

[6] *Segars, W.P. and Tsui, B., "MCAT to XCAT: The Evolution of 4-D Computerized Phantoms for Imaging Research," Proc. IEEE 97(12), 1954- 1968 (2009).*

[7] *Segars WP, Tsui BMW, Jing C, Fang-Fang Y, Fung GSK, Samei E. Application of the 4-D XCAT Phantoms in Biomedical Imaging and Beyond. IEEE Trans Med Imaging. 2018;37:680-692. PMCID: PMC5809240. PMID: 28809677.*

[8] *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;38:1457-1465. PMCID: PMC6598436. PMID: 30561344.*

[9] *Elston B, Comtat C, Harrison RL, Kinahan P. ASIM: an analytic PET simulator. CRC Press, Boca Raton, FL. 2012.*

Tomographic reconstruction for chest tomosynthesis based on Deep Learning

Lorente Mur. A, F Del Cerro. C, Desco. M, Abella M.

***Abstract*—Tomosynthesis presents a viable alternative to planar radiography by offering pseudo-tomographic data with a significantly reduced radiation dose compared to CT scans. However, a notable drawback lies in its inability to provide information about density, posing a significant limitation for chest imaging, in particular to detect pathologies that present an increase in the density of the pulmonary parenchyma.**

Previous works have sought to enhance the quality of tomosythesis images by employing advanced analytical or iterative algorithms, or by incorporating a deep-learning based post-processing step, but the outcomes presented still fall short of achieving the quantitative precision characteristic of computed tomography (CT) scans.

In this work, we explore different reconstructions methodologies consisting of a filtered back-projection step followed by post-processing based on Deep Learning to obtain a tomographic image closer to CT. Preliminary results show the potential of the proposed methodology to obtain true tomographic information from tomosynthesis data, which could replace CT scans in applications where the radiation dose is critical.

***Index Terms*—Chest tomosynthesis, Image reconstruction, Quantitative image, Computed Tomography, Deep Learning, Transfer Learning.**

I. Introduction

CHEST radiography is widely used for the diagnosis of lung diseases. It has a high spatial resolution, but being a projection image, tissues are shown overlapped without depth information and with a masking of low-density structures behind the bones. The only available technique to retrieve a density map is computed tomography (CT), but its use is limited by issues of availability and/or radiation dose delivered to the patient. Tomosynthesis offers an alternative with pseudo-tomographic information with a significantly reduced radiation dose compared to CT scans, but it still cannot convey density information. This is a major limitation for the use of tomosynthesis in chest imaging, due to the wide range of pathologies with increased lung parenchymal density.

Tomosynthesis reconstruction using deep learning (DL) can be divided into two main approaches: 2D-based methods and 3D-based methods. The 2D-based approaches apply a 2D model either to the projections before reconstruction with a traditional algorithm [1], which does not significantly improve the result, or to the slices of the reconstructed volume [2], which leads to discontinuities in the 3D space. In contrast, several works based on 3D approaches either perform 3D convolutions as a post-processing step after reconstruction using traditional methods [3], or map a traditional reconstruction algorithm onto a DL model [4], or embed DL model within an iterative reconstruction algorithm, known as unrolling methods [5, 6]. The main problem of 3D methods are that 3D convolutions have large memory requirements and with the increased number of parameters, the risk of overfitting is higher [7]. These challenges can be diminished by lowering the dimension of the problem through a patch-based approach [8] but, due to the highly anisotropic nature of Tomosynthesis volumes, the design of the patches is non-trivial.

In this work, we evaluate four post-processing strategies based on DL to enhance a preliminar FDK-based reconstruction. Our experiments include 2D single-step, 2.5D, multi-step 2D, and 3D patch-based post-processing approaches.

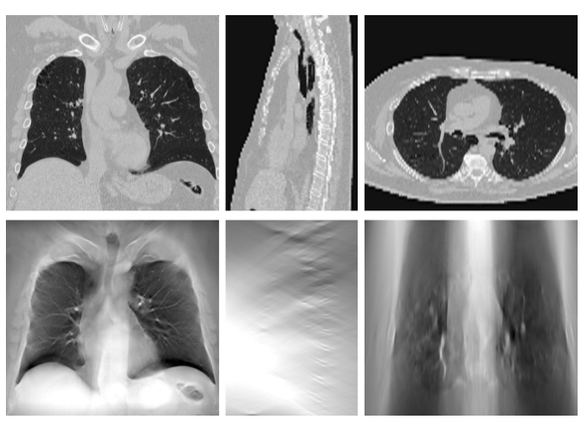


Fig. 1. Coronal, sagittal and axial view of CT (top) and FDK reconstructions from tomothyntesis data (bottom).

II. Materials and methods

A. Database

The database consists of 75 chest CT studies, 58 provided by the Hospital General Universitario Gregorio Marañón and 17 extracted from the Medical Imaging and Data Resource Center (MIDRC) [9].

From these CT volumes, we generated the tomosynthesis projections using FuxSim [10], a simulation software of flexible acquisition geometries based on the density maps. The simulations were based on a standard chest tomosynthesis protocol on a GE VolumeRAD tomosynthesis system: 61 projections with a total source-to-detector distance of 1800 mm and a vertical source displacement of 16mm per projection.

Lorente Mur. A, C. F Del Cerro, M. Abella and M. Desco are with the Dpto. Bioingeniería e Ingeniería Aeroespacial, Universidad Carlos III de Madrid and the Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, España (e-mail: mabella@ing.uc3m.es, cfernandez@hggm.es, alorente@ing.uc3m.es).

M. Abella and M. Desco are with the Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, España (e-mail: manuel.desco@cnic.es).

M. Desco is with the Centro de investigación en red salud mental (CIBERSAM), Madrid, España. (e-mail: desco@hggm.es).

The resulting projections had a matrix size of 1072×1072 pixels with a pixel size of 0.4×0.4 mm.

The FDK reconstructions were performed using FuxSim to a common anisotropic voxel size larger than that of any of the original CT volumes, $1.25\times5.00\times1.25$ mm, with $400 \times 80 \times 200$ voxels. The slice thickness was selected to match the minimum slice thickness of commercial GE VolumeRAD chest tomosynthesis system (GE Healthcare, Chicago, IL)[11]. The original CT studies were also interpolated to a common anisotropic voxel of $1.25\times5.00\times1.25$ mm in order to create the targets for the neural networks.

The database was divided into the following sets:

- Training set: Composed of 59 volumes, giving a total of 4720 coronal, 14076 axial and 15340 sagital slices.
- Validation set: Composed of 8 volumes, giving a total of 640 coronal, 1809 axial and 2080 sagital slices.
- Test set: Composed of 8 volumes, giving a total of 640 coronal, 1809 axial and 2080 sagital slices.

B. Deep-learning based post-processing strategies

As shown in Fig. 1, in the coronal view the main structures of the tomosynthesis image overlap with those of the corresponding CT, though it also has some overlapping structures from other slices. Axial slices show some of the structures of the CT (such as some lung vessels and some ribs), though they are significantly distorted by the smearing resulting from the backprojection. Sagital slices appear to have the least distinguishable visible features amongst the 3 tomographic views.

We have evaluated the following four approaches to post-process the FDK volume:

1) Single-step 2D post-processing: A modified version of the U-Net architecture proposed in [2] was used on either coronal, axial, or sagital slices, resulting in the methods **2DCor** or **2DAx**, **2DSag**, respectively.

*2) **2.5D** post-processing:* In the **2.5D** approach, the resulting volume is derived by averaging the volumes from **2DAx**, **2DCor**, and **2DSag**.

3) Multi-step 2D post-processing: The multi-step approach applies two correction steps to the FDK reconstruction following the methodology presented in [2]. The first correction step is applied on the coronal, axial, or sagital plane and the second on a plane orthogonal to it, resulting in algorithms **Cor2Ax**, **Ax2Cor** and **Sag2Ax**, respectively.

4) 3D patch-based post-processing: We evaluated two state-of-the-art models based on the Unet encoder-decoder structure [12] for reconstructing 3D patches composed of 16 coronal slices (of size 386×192 pixels) from the tomosynthesis volume. The first architecture we considered is the 3D-Unet [13], which is the 3D extension of the Unet architecture, where 2D convolutions are replaced with 3D convolutions. We refer to the 3D patch-based post-processing using 3D-Unet as **3D-P-U**. The second architecture that we evaluated is the 3D UneTR [14], which is a neural network based on the combination of attention-based mechanisms [15] and the encoder-decoder structure of Unet architectures. It formulates the 3D volume-to-volume mapping as a sequence-to-sequence prediction problem by dividing the volume into a sequence of uniform non-overlapping patches. We refer to this method as **3D-P-UTr**.

C. Training details

All neural networks were implemented using Pytorch [16] for 50 epochs to minimise the mean squared error, using Adam [17] as the optimizer, and a weight decay equal to 10^{-3}. The learning rate is initialized to 10^{-3} and divided by 2 when the train loss function has reached a plateau.

III. Results and discussion

We can see in Fig. 2 that 2D approaches lead to discontinuities on the planes where the neural network was not trained. These discontinuities are shown as high frequency bands indicated by the red arrows: **2DAx** shows some vertical bands in the coronal and sagital views, **2DCor** in the axial and sagital views, and **2DSag** in the coronal and axial planes. The **2.5D** strategy, which averages the single-step 2D methods in order to incorporate some information from all planes, inherits the discontinuities from all approaches without any real benefit over the other methods. We can see that **2DAx** and **2DCor** fail to maintain the density mass on the lungs shown by the blue arrow in the first column, significantly reducing their potential clinical use. Surprisingly, while the sagital slices do not appear to have many visible structures in the FDK, **2DSag** is the only approach that seems to correctly recover the higher density mass in the lungs shown by the blue arrow in the 5th column. This suggests that the sagital view may have more features than what our initial evaluation lead to believe.

Fig. 3 shows that the multi-step post-processing visually reduces the discontinuities in the reconstructed volume, but lost details during the first correction step are not recovered during the second correction step: **Ax2Cor** and **Cor2Ax** lack the mass in the lungs (blue arrows in the 2nd and 3rd columns), whereas **Sag2Ax** manages to recover it (see the blue arrow in the 4th column). This shows that the information on the sagital plane is important not to lose details within the lungs. Nevertheless, since **2DSag** has the most discontinuities among the single-step 2D methods, the second correction step retains some of these artefacts.

3D patch-based post-processing post-processing better recover the shape of the front and the back of the patient compared to the other methods. In particular, **3D-P-U** shows the best visualization of the general shape of the structures of the ground truth volume. Regarding the mass in the lungs, **3D-P-U** fails to recover it, similarly to the 2D methods, while **3D-P-UTr** is able to keep part of if.

IV. Conclusion and perspectives

This work presents the evaluation of post-processing strategies using DL to correct an FDK-based tomosynthesis reconstruction. Against all expectation, the 2D post-processing of sagital slices has shown high potential for the reconstruction of details within the lungs, and 3D patch-based post-processing has shown the best reconstruction of the overall shape of the patient. It would be interesting for future work to consider

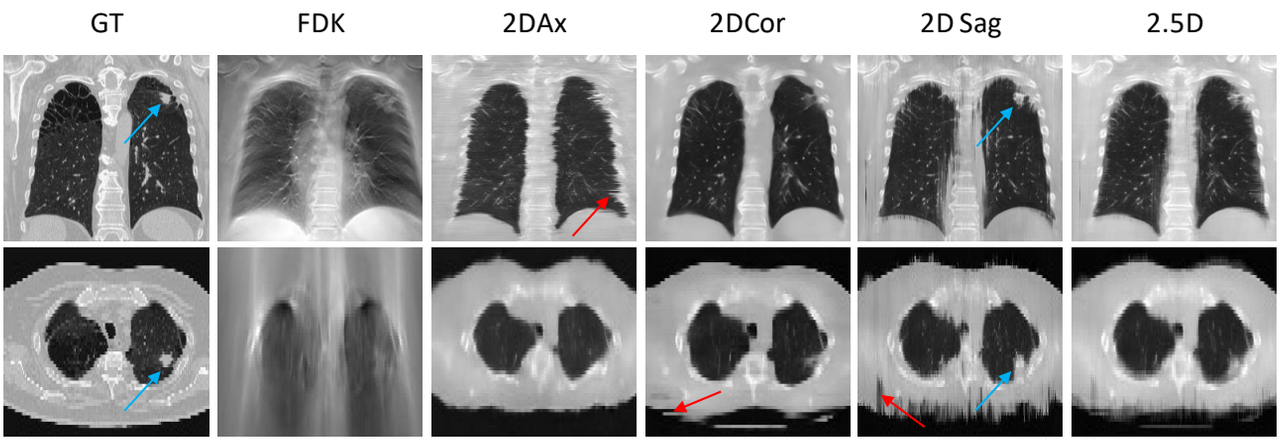


Fig. 2. Coronal and axial slices of the reconstructed volume from tomosynthesis projections with 2D approaches.

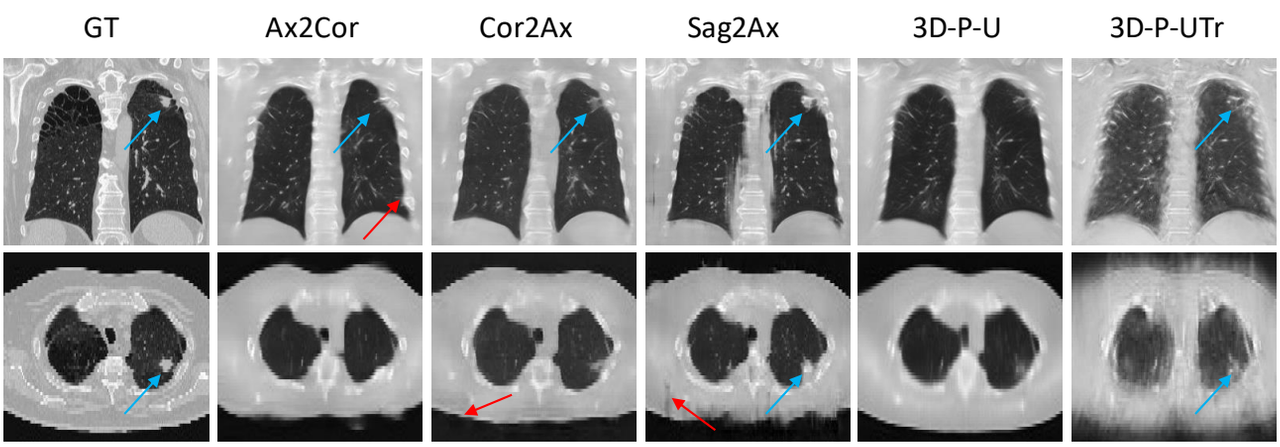


Fig. 3. Coronal and axial slices of the reconstructed volume from tomosynthesis projections with 3D approaches and multi-step approaches.

methods based on 3D patches with sagital slices and axial slices. Our evaluation must further extend to the evaluation of the consistency of the images with respect to the original images. The evaluated methods show that pseudo-tomographic quality can be obtained from tomosynthesis projections. This opens the path to low-dose screening for lung illnesses with tomosynthesis devices.

Acknowledgment

This work was funded by Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación MCIN/AEI/10.13039/501100011033 and by the European Union 'NextGenerationEU'/PRTR through project PDC2021-121656-I00. Also funded by Instituto de Salud Carlos III through the projects PMPTA22/00121 and PMPTA22/00118, co-funded by the European Union NextGenerationEU/MRR. The CNIC is supported by Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación, and the Pro CNIC Foundation.

The imaging and associated clinical data downloaded from MIDRC (The Medical Imaging Data Resource Center) and used for research in this publication was made possible by the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under contracts 75N92020C00008 and 75N92020C00021. 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] Tsutomu Gomi, Hidetake Hara, Yusuke Watanabe, and Shinya Mizukami. Improved digital chest tomosynthesis image quality by use of a projection-based dual-energy virtual monochromatic convolutional neural network with super resolution. *PLoS One*, 15(12):e0244745, 2020.

[2] CF Del Cerro, A Galán, J García Blas, M Desco, and M Abella. New reconstruction methodology for chest tomosynthesis based on deep learning. In *7th International Conference on Image Formation in X-Ray Computed Tomography*, volume 12304, pages 661–668. SPIE, 2022.

[3] Donghoon Lee and Hee-Joung Kim. Restoration of full data from sparse data in low-dose chest digital

tomosynthesis using deep convolutional neural networks. *Journal of digital imaging*, 32(3):489–498, 2019.

[4] Yufeng Wang, Zhan-Li Sun, Zhigang Zeng, and Kin-Man Lam. Trct-gan: Ct reconstruction from biplane x-rays using transformer and generative adversarial networks. *Digital Signal Processing*, 140:104123, 2023.

[5] Nikita Moriakov, Koen Michielsen, Jonas Adler, Ritse Mann, Ioannis Sechopoulos, and Jonas Teuwen. Deep learning framework for digital breast tomosynthesis reconstruction. In *Medical Imaging 2019: Physics of Medical Imaging*, volume 10948, page 1094804. International Society for Optics and Photonics, 2019.

[6] Ting Su, Xiaolei Deng, Jiecheng Yang, Zhenwei Wang, Shibo Fang, Hairong Zheng, Dong Liang, and Yongshuai Ge. Dir-dbtnet: Deep iterative reconstruction network for three-dimensional digital breast tomosynthesis imaging. *Medical Physics*, 2021.

[7] Yichi Zhang, Qingcheng Liao, Le Ding, and Jicong Zhang. Bridging 2d and 3d segmentation networks for computation-efficient volumetric medical image segmentation: An empirical study of 2.5d solutions. *Computerized Medical Imaging and Graphics*, 99:102088, 2022.

[8] Yunsu Choi, Hanjoo Jang, and Jongduk Baek. Chest tomosynthesis deblurring using cnn with deconvolution layer for vertebrae segmentation. *Medical Physics*, 2023.

[9] Medical imaging & data resource center (midrc) data commons. https://data.midrc.org/. Accessed: 2020-01-03.

[10] Monica Abella, Estefania Serrano, Javier Garcia-Blas, Ines García, Claudia De Molina, Jesus Carretero, and Manuel Desco. Fux-sim: Implementation of a fast universal simulation/reconstruction framework for x-ray systems. *PloS one*, 12(7):e0180363, 2017.

[11] New advanced applications in digital x-ray. https://www.apt.drg.de/media/document/1272/APT-2005-Tuerkay-GE-Advanced-Applications.pdf. Accessed: 2020-01-03.

[12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[13] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II 19*, pages 424–432. Springer, 2016.

[14] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584, 2022.

[15] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[16] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Z. Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019.

[17] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

Adaptation of a 4D noise filter for implementation on GPU

Mikhail Mikerov, Sjoerd A.M. Tunissen, Koen Michielsen, Ewoud J. Smit, and Ioannis Sechopoulos

***Abstract*—Functional 4D CT imaging including contrast injection can be beneficial when evaluating pathology. However, acquisition of long image sequences implies the need to use a low dose per phase to limit total radiation dose. This calls for development of advanced processing methods that can handle the resulting high image noise efficiently. Noise filters specific for dynamic imaging can use temporal information to achieve higher noise reduction but have a high computational cost. In this work, we accelerated an existing clinical 4D noise filter by making changes to its algorithm that enabled it to run approximately 13 times faster on GPU compared to a highly optimized CPU implementation without any obvious reduction in its effectiveness.**

I. Introduction

Despite significant progress in CT hardware and software that improved image quality and clinical decision making, standard CT provides only anatomical information of the patient condition. In many clinical situations it would be beneficial to also obtain functional information by studying the enhancement with iodinated contrast agents, especially when evaluating pathology for treatment planning or response.

Usually three- or four-phase functional imaging protocols are implemented in the clinic for patients with stroke, cardiac ischemia, or liver lesions. They can still be performed at relatively low doses but have the risk of mistimed acquisitions, resulting in missing important aspects of contrast dynamics. More densely sampled sequences would eliminate the risk of missing important points on perfusion curves.

Functional imaging with long sequences consisting of 10 to 30 phases depending on the protocol comes at the cost of an increased dose. To reduce this, each phase in the sequence needs to be acquired at considerably lower dose than a standard anatomical CT image. Noise reduction filters can then be applied as a postprocessing step to enable the extraction of useful information from otherwise too-noisy images.

Traditional denoising filters can be used in each individual phase, however they often use information from a small spatial neighborhood, leading to loss of resolution. To solve this, information from the other images in the sequence can also be used to decrease noise in the current phase. Therefore noise filters specific for dynamic imaging that leverage the temporal information such as the 4D similarity filter have been proposed [1].

M. Mikerov, S.A.M. Tunissen, E.J. Smit, 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.

In this work, we aim to improve runtime of the 4D similarity filter by introducing a small modification to the algorithm that enables its implementation on GPU. We performed the first comparison with the exact CPU version and did not notice a performance drop.

II. Filter implementation

A. 4D similarity filter

Prokop and Smit introduced a 4D similarity filter to reduce noise in low dose 4D imaging [1]. The idea of this filter is to find, for each voxel v in the image at a given timepoint t, a set of voxels $\{V_c\}$ that have similar perfusion curves and use those with the least differences to calculate an average voxel value that can then be used to update the image.

The search for similar voxels is performed by comparing the perfusion curves of the voxel of interest v (the one being updated) with a large pool of candidate voxels $v_c \in \{V_c\}$ from the entire image according to how close the temporal mean of each voxel v_c is to that of v. If the perfusion curve of a candidate voxel v_c is similar according to some threshold TH, it is saved in an array A. When the array is full, it is sorted based on the difference of perfusion curves and only the best FS candidate voxels are used to calculate the average value for this voxel. This filter can be implemented both on CPU and GPU, but the latter has not been achieved before. There exist multi-threaded CPU realizations of this filter, which are already used in clinical practice [2]–[4].

One of the drawbacks of the 4D similarity filter is long computational times since every voxel in the image needs to be processed independently. In a typical 320x512x512 volume, with just 1/3 of all voxels being in the mask, the total number of voxels is around 27 million. On a 32-core CPU, each core would need to process more than 870 thousand voxels. The independence of each voxel invites use of the GPU for this task, since it has many more cores, which can process even more voxels in parallel. However, we encountered major problems when implementing the 4D similarity filter on the GPU.

B. Difficulties of implementation on GPU

In general, low computational times on the GPU are achieved when it is possible to execute one thread per output element that can run independently from other threads and consists mainly of arithmetic operations with very limited memory accesses. This is due to the fact that arithmetic operations need up to 20 times fewer cycles to complete than operations on device (global) memory. If a high number of

memory accesses are required to compute a single output element, several techniques can be employed. First, one wants to ensure that memory reads are coalesced, i.e., during the calculation of neighboring output elements, memory reads from the threads in the same block should be performed in accordance to memory layout. Second, if multiple threads in the same block can reuse some of the information, shared memory can be used to specify a user-defined and -maintained cache. Third, if memory reads are random but there is a high chance of having close spatial proximity, textures can be utilized due to their large cache, which is checked first before reading the data from the (slow) device memory.

Unfortunately, none of these optimization techniques are applicable in our problem since each candidate voxel can be located in a completely different part of the image, therefore making coalesced memory reads impossible. Moreover, the probability of hitting the same element is very low, which removes the advantages of using the fast texture cache. Shared memory cannot be used either since it is not known at execution time whether a candidate voxel in one of the threads will be reused in another thread within the same block and, if so, in which one specifically. Thus, it is not possible to save this value in a structured cache for future use.

Another major limiting factor is the need to sort the selected similar voxels in array A according to the difference in perfusion curves. This is difficult due to two reasons. To begin with, keeping track of sometimes thousands of data pairs (difference and voxel value) requires sufficient memory that quickly exceeds the amount of fast on-chip shared memory. Moreover, since shared memory needs to be shared between all threads in a thread block, it greatly reduces the number of voxels that can be processed in parallel. Thus there is no other choice left other than accept writes into slow device memory. The fastest sorting algorithms require on average $\mathcal{O}(n \log n)$ comparisons, which can become a large number for typical settings of the filter. From the computational point of view this method becomes inefficient on the GPU, since memory reads and writes are done in device memory and if not all threads in a thread block (*warp*) have finished their sorting, the core cannot be released to process another block of threads.

A very straightforward approach to reduce the computational burden due to sorting at the end is to sort each element at insertion by using, e.g., min-heap structure of size FS. With this approach, one does still need to keep the differences and values of FS elements. However, FS is almost always smaller than the size of A and it can be possible to store all FS elements in fast memory. Theoretically, on average the insertion takes again $\mathcal{O}(n \log n)$ comparisons, but now the memory throughput is very high since we are working with fast shared memory only. In our case, this number will be even lower since the min-heap structure is sorted before each insertion. Nevertheless, the insertion must be performed as many times as the number of elements we wanted to have in our original array A. Furthermore, element insertion into a heap has multiple if-statements, which increases the likelihood of branching and therefore they could lead to the cores being idle more often.

C. Our solution

Based on these algorithmic bottlenecks hindering an efficient implementation on GPU, we propose a simple change to the algorithm that enables us to work around these issues and attain fast computational times on the GPU at the cost of a small reduction in the numerical accuracy of the solution.

Figure 1 schematically represents our way of storing the same number of elements as in the original A with limited space requirements and a fixed number of operations ($\mathcal{O}(1)$) on insertion and reading in the worst-case scenario.

As in the original version, a candidate voxel v_c is considered for future processing only if the difference between its and voxel v's perfusion curves is below a given threshold TH. By carefully selecting TH, it is possible to save all good candidate values in a sorted manner with limited memory resources. Since the end-result of each kernel execution is an average of FS best candidate voxels v_c, it makes sense to store the average and corresponding count utilizing as little memory space as possible. Assuming that a typical CT image is in the range from -1024 to 3071 HU, storing each value using only positive numbers requires just 12 bits. In case of 32-bit integers it leaves us with 20 bits that can be used to store metadata corresponding to this number. For us, the only metadata we have for the average value is the number of voxels that were used to calculate it. Now we only need to store it in a sorted manner in a small data structure. We propose to create an array H of size N where each n-th bin stores the average value of the candidate voxels v_c with the differences between $n \cdot TH/N$ and $(n+1) \cdot TH/N$. For example, if $N = 100$, $TH = 2000$, the first bin is responsible for storing the average value of candidate voxels with the difference between 0 and 19. The size of the array N is a trade-off between the intended bin width for a certain TH and the number of threads that can run in parallel. For example, for the Nvidia A6000 GPU with 99 KB of shared memory, one can theoretically have 193 parallel threads with $N = 128$ or just 96 threads with $N = 256$ in each block.

D. Limitations

The changes introduced to the algorithm have a negative effect on the numerical accuracy of the 4D similarity filter, which is important to understand and to explore possible solutions. To start, some accuracy is lost when calculating the average due to rounding, since only integer values can be stored. Theoretically, one could use fixed point number arithmetic to mitigate this problem, if one can accept that fewer bits will be used to represent the integer part of the number or the maximum count will be smaller. Following that, the width of the bin could be too broad so that almost all of FS best voxels can be found in one of the first bins. If the number of candidate voxels in one of the bins near the start of the array H greatly exceeds FS at some locations in the image, the noise reduction will not be homogeneous across the whole image. When it is not possible to increase the number of bins anymore due to memory constraints, one could move the array H into device memory, at the cost of increasing memory latency. Alternative solution could be to use the array H with

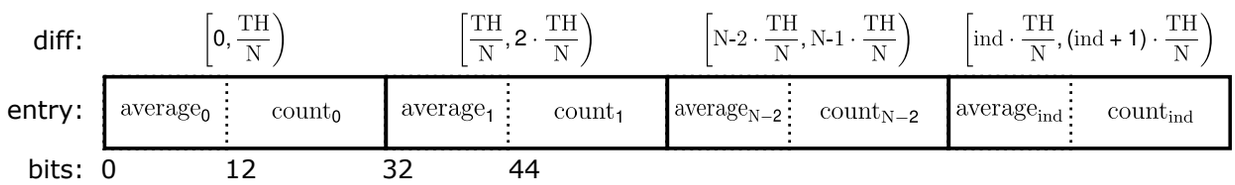


Fig. 1. A sorted array structure is used to store the average value of voxels within a specific range of differences, along with the corresponding voxel count required for this average calculation. The insertion and extraction operations have a worst-case performance of $\mathcal{O}(1)$.

non-uniform bin width. This way, the first bins can be made smaller to enable more granular summation of FS.

III. Materials and methods

We compared the performance of our and commercial CPU implementations of the 4D similarity filter on simulated CT images of an anthropomorphic XCAT perfusion phantom. The acquisitions were simulated in low-dose mode ($\mathrm{CTDI}_{\mathrm{vol}} = 1.3$ mGy per phase) using a previously developed and validated CT simulator [5]. The images were reconstructed using an own implementation of the standard FDK algorithm. To avoid negative values and allow the custom array structure we applied the following linear scaling to arbitrary units: $\mathrm{AU} = (\mathrm{HU} + 3000)/6$. n this preliminary work, we performed very simple analysis that included quantification of noise reduction in terms of signal-to-noise ratio (SNR) in a homogeneous region of the liver and estimation of the bias.

The GPU implementation has fewer and slightly different parameters than the CPU version. Even though we adjusted the parameters according to our understanding of the filter in order to achieve comparable performance, we note that both filters are not expected to lead to identical results. The size of array H was 128 bins to have a multiple of 32 to avoid bank conflicts when writing and reading from H. Mean squared error (MSE) TH was 12800, leading to each bin being 100 MSE units wide. FS was set to 100. For each voxel in the mask, 40 000 candidate voxels were checked.

IV. Results

Figure 2 shows a slice of the original reconstructed phantom and two slices filtered using the CPU and GPU versions of the algorithm. Both results look very similar in terms of representing anatomical structures.

We calculated the SNR in the entire liver as we had perfect segmentation from the XCAT phantom. The GPU version has achieved slightly higher noise reduction than the GPU version, (SNR increased by a factor of 7.3 and 7.7, respectively). Since the noise is expected to decrease mainly as a function of the number of averaged voxels FS provided that they are sampled from the same distribution, and unlike in the CPU version, FS is not fixed, we calculated the actually used FS for each voxel in the output. Figure 3 shows the actually used FS to calculate the average for each voxel in the image. In the whole liver, we found FS to be 126 ± 19 which is higher than the intended $FS = 100$ used in the CPU version.

No visible bias is observed when comparing images filtered on CPU and GPU. To verify this we constructed the Bland-Altman plot shown in Figure 4. There is a small negative bias of -6 HU. There appears to be no clear dependency of the difference between two versions of the filter and voxel grayscale value. Moreover, we observed that both the CPU and GPU versions introduce either positive or negative bias depending on the location in the liver or other organs. There is no clear dependency on anatomy, functional or image properties.

The GPU result can be obtained within 44 seconds on an Nvidia A6000 GPU, which is on average 13 times faster than the multi-threaded commercial CPU implementation that is executed on 32 cores, based on the mean of three runs. The Nvidia profiler has also shown that the L1 and L2 caches are not fully utilized. In order to improve reading from device memory, we implemented memory prefetching – eight candidate voxels are loaded from the device memory into registers and are then processed. This modification improved the utilization of the caches. Better memory utilization increased compute throughput by 385% and enabled, thus, the execution of the filter within 27 seconds, which is roughly 23 times faster than the CPU version.

V. Discussion

We modified the original algorithm of the 4D similarity filter in order to avoid GPU-specific computational constraints. With this we were able to run it on GPU several times faster than with highly-optimized multi-threaded code on the CPU. The main change was replacing the sorting operation by using an ordered hash-like structure that can reside in fast but limited on-chip shared memory and thereby enabling $\mathcal{O}(1)$ writing and reading. However, we needed to sacrifice accuracy to use this data structure. We performed first tests with both versions of the similarity filter to study the effect of this accuracy loss.

Overall, both versions produce very similar outputs with comparable set of parameters. The main difference is a higher level of denoising than anticipated, which is explained by the fact that FS is not a fixed parameter anymore. Therefore it leads to different level of noise reduction in various parts of the image, as demonstrated. To counteract this effect, the bin width in the hash H could be chosen smaller, therefore enabling more granular division of candidate voxels and thus increasing the chance of using the sought number of voxels FS for averaging.

An important advantage of our implementation is significantly reduced complexity of the code. This simplification will enable us to add modifications to the algorithm and explore additional features such as looking for neighbors in a local region or using different metrics to select the voxels. Furthermore, our implementation enables the filtering of single

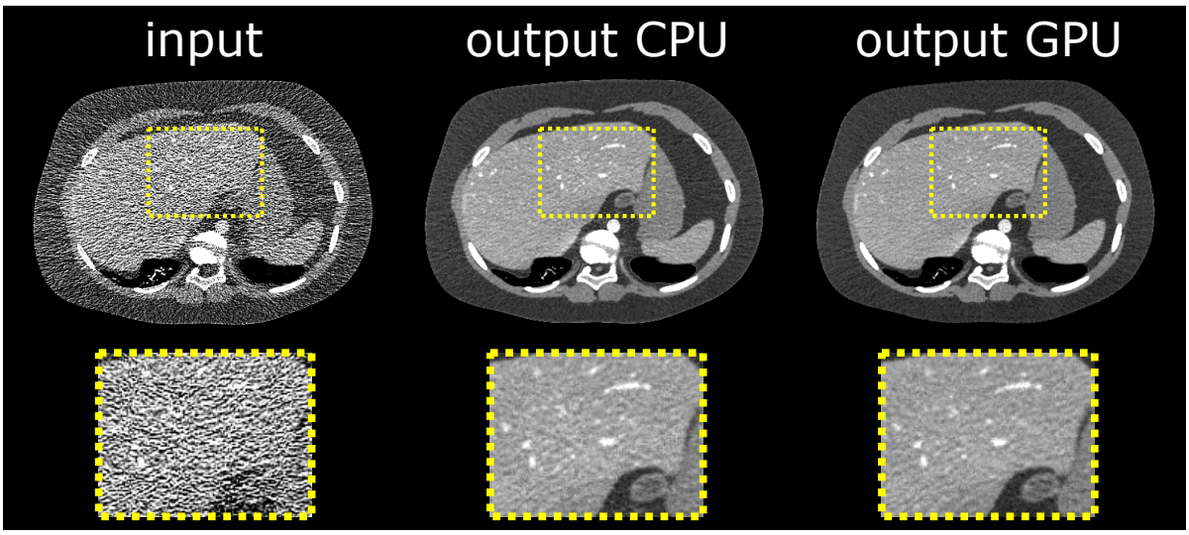


Fig. 2. Comparison of exact CPU and approximate GPU 4D similarity filters. W/L: 200/50 HU.

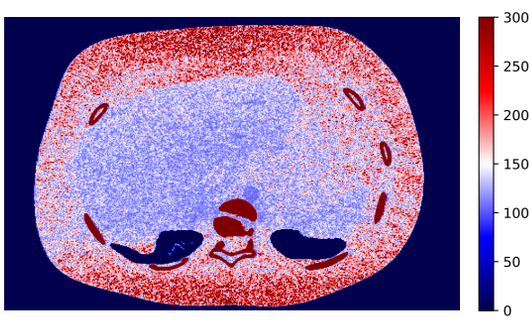


Fig. 3. Actually used FS for each voxel measured as a number of voxels used to calculate the average. The values are clipped at 300.

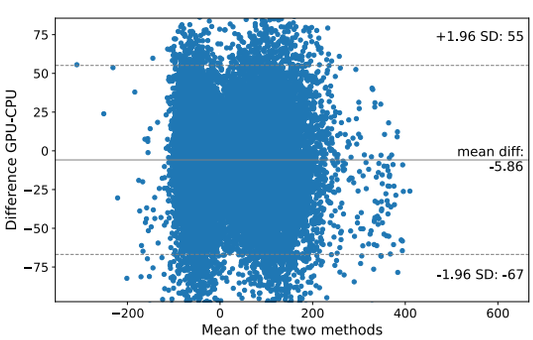


Fig. 4. Bland-Altman plot comparing GPU and CPU versions of the filter. The output of the GPU version is on average lower.

phase in the sequence. An interesting scenario to investigate would be the application of the filter to reduce the noise in every second or third image in the sequence for calculation of approximate perfusion maps in clinical situations in which the time of diagnosis is the main concern, while more images are being processed in the background to update the perfusion maps with more accurate data.

In this work we performed an initial evaluation of how the filter is affected by the algorithm changes in a digital phantom. At the moment it is not clear whether the introduction of slightly non-uniform noise reduction throughout the image and different anatomies we found will have negative effects on the performance of radiologists or further processing of the images, e.g., to calculate perfusion maps. Based on these initial results we plan to continue with further analysis of the performance by, e.g., quantifying bias of the perfusion curves in different anatomies. Such an analysis combined with hyperparameter tuning must be done in order to ensure correct and trustworthy performance.

VI. Conclusion and Outlook

We implemented a GPU version of the 4D similarity filter that is on average approximately 13 times faster than a highly optimized commercial multi-threaded CPU realization by introducing a change into the original algorithm to make it more GPU-friendly at the cost of lowering accuracy. Our modification to the algorithm removed a significant bottleneck for GPU implementation. Our further non algorithm-specific modification enabled further speedup by factor of 1.7 compared to the already modified algorithm.

References

[1] E. J. Smit and W. M. Prokop, “Noise reduction in image data,” WO 2017/194787 A1, 2017.

[2] T. Kouchi, Y. Tanabe, E. J. Smit, T. Kido, A. Kurata, Y. Kouchi, H. Nishiyama, T. Uetani, S. Ikeda, O. Yamaguchi, M. Prokop, and T. Mochizuki, “Clinical application of four-dimensional noise reduction filtering with a similarity algorithm in dynamic myocardial computed tomography perfusion imaging,” *The International Journal of Cardiovascular Imaging*, vol. 36, pp. 1781–1789, 2020.

[3] O. Sliwicka, Z. Swiderska-Chadaj, M. Snoeren, M. Brink, K. Salah, L. Peters-Bax, T. Stille, M. J. van Amerongen, I. Sechopoulos, and J. Habets, “Multireader image quality evaluation of dynamic myocardial computed tomography perfusion imaging with a novel 4d noise reduction filter,” *Acta Radiologica*, vol. 64, no. 3, pp. 999–1006, 2023.

[4] K. Murayama, E. J. Smit, M. Prokop, Y. Ikeda, K. Fukui, I. Nakahara, S. Hanamatsu, K. Katada, Y. Ohno, and H. Toyama, “A bayesian estimation method for cerebral blood flow measurement by area-detector ct perfusion imaging,” *Neuroradiology*, vol. 65, pp. 65–75, 2022.

[5] S. A. M. Tunissen, L. J. Oostveen, N. Moriakov, J. Teuwen, K. Michielsen, E. J. Smit, and I. Sechopoulos, “Development, validation, and simplification of a scanner-specific ct simulator,” *Medical Physics*.

An Iterative 3D Ring Artifacts Correction Algorithm for CBCT

Yanwei Qin, Xin Lu, Hongyang Zhu, Yunsong Zhao*, Xing Zhao

***Abstract*—In recent years, flat panel detector (FPD) based cone beam CT (CBCT) has gained popularity in medical examinations and industrial inspections. But faulty pixels of FPD may cause light or dark stripes in the obtained projections that result in bright or dark concentric rings or half-rings in the reconstructed CT images. These artifacts can significantly affect the accuracy of CT values and even introduce pseudo structures, potentially leading to misjudgment of tissue structures or misdetection of industrial products.**

In this paper, we propose an optimization model that utilizes inter-layer correlations of 3D projections and 3D CT images to correct ring artifacts for CBCT. The stripes artifact's inter-layer correlation in 3D projection manifest as the group sparsity property, which can be expressed by the $L_{2,1}$ norm, so the group sparse constraints for stripes is added to the model. In addition, since 3D reconstructed images usually have the piecewise constant property, we add the 3D TV constraints to the model expected a stable solution. Further, an iterative algorithm is proposed to solve the optimization model in dual domain. Real data experiments are carried out to illustrate the effectiveness of the proposed method in removing ring artifacts.

***Index Terms*—Ring artifact removal, image reconstruction algorithm, group sparse, dual domain, CBCT**

I. Introduction

FLAT panel detector (FPD) based CBCT has many advantages, including rapid scanning speed, high X-ray utilization, isotropic image resolution, etc., which are extensively utilized in both medical area and industry area. But due to technical limitations, FPD usually contains detector pixels with inconsistent response, leading to stripe artifacts in projection data, and forming the so-called ring artifacts in the reconstructed CT images.

To remove ring artifacts, various correction methods have been proposed, including hardware correction methods and software correction methods. Hardware correction methods [1, 2] alter the data acquisition by moving the sample. After this movement, the originally fixed correspondence between the detector pixel response channel and X-rays is changed. In this way, the errors caused by inconsistent responses are distributed to the entire projections, so the impact of detector pixel responses inconsistent on the projected image is reduced.

Yanwei Qin, Xin Lu, Hongyang Zhu, Yunsong Zhao and Xing Zhao are with School of Mathematical Sciences, Capital Normal University, Beijing, 100048, China, and with Beijing Advanced Innovation Center for Imaging Technology, Capital Normal University, Beijing, 100048, China.

E-mail: zhao_yunsong@cnu.edu.cn.

This work is supported in part by the National Natural Science Foundation of China (NSFC) (61871275, 61771324, and 61827809), and is supported in part by the National Key R&D Program of China (2023YFA1011400).

The authors are grateful to Beijing Higher Institution Engineering Research Center of Testing and Imaging as well as Beijing Advanced Innovation Center for Imaging Technology for funding this research work.

Software correction methods can be further categorized into flat field correction methods [3–7], image processing methods [8–12], and neural network methods [13–15].

In X-ray imaging, the flat field correction method comprises a series of methods that are derived from a fundamental formula as follows. If the dark field data I_{dark} is obtained when X-ray is turned off and the flat-field data I_{flat} is the sample-free data when X-ray turned on, the corrected data I_{free} can be caculated by:

$$I_{\text{free}} = \frac{I_{\text{raw}} - I_{\text{dark}}}{I_{\text{flat}} - I_{\text{dark}}}$$

where I_{raw} is the raw data obtained by CT system directly. This method is a physical principle-based method, while it may be not thorough enough. Image processing methods use common image processing methods such as filtering, wavelets, interpolation, and smoothing to eliminate ring artifacts, which effect is relatively stable. Recently, the neural network methods demonstrated advantageous performance in many vision tasks, including ring artifact removal in CT image reconstruction. However, since neural network methods usually require large training data sets, they may have some limitations in practical applications.

Among image processing methods, dual-domain methods refer to a series of methods that simultaneously remove stripe artifacts on sinograms and ring artifacts in reconstructed CT images. In this paper, we propose a novel dual domain method for ring artifacts removal. In the proposed method, we divide the projection data into stripe-free part and stripe part, and then use tensor decomposition to analyze the three-dimensional projection data. For the strip part, we found that the elements within each column of the stripe exhibit a shared group sparse pattern. Additionally, a total variation (TV) constraint is applied to the "clean" reconstructed image, which is widely used in image processing to preserve local spatial piecewise constancy. In summary, a new iterative format is proposed with its corresponding solving algorithm. Real data experiments verify the effectiveness of the proposed method.

II. Materials and methods

A. TLT model

Let $I_{\text{raw}}(x, z)$ be the acquired normalized raw data after dark field correction and offset correction [3], and let $I_{\text{free}}\ (x, z)$ represent the corresponding ideal data. Similar to [4], we have:

$$I_{\text{free}}(x, z) = C(x, z) I_{\text{raw}}(x, z) \tag{1}$$

where $C(x,z)$ is the correction coefficient for the detector pixel with coordinate (x,z). Generally, in order to obtain I_{free}, we need to find a suitable method to obtain the correction coefficient C as accurate as possible.

To find a suitable scale method was used in article [6] to obtain correction coefficient C, and the suitable scale is the quotient of ideal average raw data value along the scanning angle and the true average raw data value. In fan beam scanning, the response coefficient is a one-dimensional correction vector, whereas in cone beam scanning, the response coefficient is a two-dimensional correction matrix. However, through our experiments, we have discovered that, the received raw data values are vary greatly with the scanning view rotate when passing through different materials. Therefore, the response of the same detector pixel is difficult to describe by the same average value.Therefore, the correction coefficient C obtained by [6] is not accurate enough for the complicated situations.

Take negative log transform on both sides of equation (1), we have:

$$\begin{aligned}\mathcal{P} &\triangleq -\log\left(I_{\text{free}}\right) = -\log\left(CI_{\text{raw}}\right)\\ &= -\log\left(I_{\text{raw}}\right) - \log(C)\\ &\triangleq \mathcal{Y} - \mathcal{S}\end{aligned} \tag{2}$$

where $\mathcal{P} = -\log\left(I_{\text{free}}\right)$, $\mathcal{Y} = -\log\left(I_{\text{raw}}\right)$, and $\mathcal{S} = \log(C)$ are three-order tensors, and represent the ideal projections, the acquired projections, and the stripe artifacts in projection domain, respectively. Let R be the system matrix of the CBCT, $\mathcal{U}$ represents the target "clean" CT reconstructed image free from ring artifacts, then we have:

$$\mathcal{P} = R\mathcal{U} \tag{3}$$

Substitute equation (3) into equation (2), we have:

$$\mathcal{Y} = R\mathcal{U} + \mathcal{S} \tag{4}$$

Since both $\mathcal{U}$ and $\mathcal{S}$ are unknown, problem (4) is ill-conditioned. In order to address this issue, we convert it to an optimization model and incorporate prior information of the reconstructed image $\mathcal{U}$ and the stripe artifacts $\mathcal{S}$ as follows:

$$\begin{aligned}&\arg\min_{\mathcal{U},S} \alpha_1 J_1(\mathcal{U}) + \alpha_2 J_2(\mathcal{S})\\ &\text{s.t. } \mathcal{Y} = R\mathcal{U} + \mathcal{S}\end{aligned} \tag{5}$$

where $J_1(\mathcal{U})$ and $J_2(\mathcal{S})$ signify the prior constraints imposed on the "clean" reconstructed image and stripe artifacts, respectively. Parameters α_1 and α_2 determine the impact of the regularization terms.

For the strip tensor $\mathcal{S}$, it is assumed to be low-rank and have sparse gradient along the x and z direction. To address the low-rank property, we apply Tucker decomposition [16] on $\mathcal{S}$, as well as restrict the rank of the core tensor. The Tucker decomposition is:

$$\mathcal{S} = \mathcal{C} \times_1 \mathrm{A} \times_2 \mathrm{B} \times_3 \mathrm{C}$$

where $\mathrm{A} \in \mathbb{R}^{M\times r_1}$,$\mathrm{B} \in \mathbb{R}^{M\times r_2}$,$\mathrm{C} \in \mathbb{R}^{M\times r_3}$ are the factor matrices, respectively, and satisfies:

$$\begin{aligned}\mathrm{A}^T\mathrm{A} &= \mathrm{I}\\ \mathrm{B}^T\mathrm{B} &= \mathrm{I}\\ \mathrm{C}^T\mathrm{C} &= \mathrm{I}\end{aligned}$$

$\mathcal{C} \in \mathbb{R}^{r_1\times r_2\times r_3}$ is the core tensor, and the rank of Tucker decomposition is defined as a vector (r_1, r_2, r_3), which is the core tensor rank. Since it is difficult to directly find the appropriate rank (r_1, r_2, r_3), we relax the constraint and tries to minimize:

$$J_2(\mathcal{S})_{\text{Tucker}} = \|\mathcal{S} - \mathcal{C} \times_1 \mathrm{A} \times_2 \mathrm{B} \times_3 \mathrm{C}\|_F^2 \tag{6}$$

To address the gradient sparsity of $\mathcal{S}$, the $L_{2,1}$ norm (also known as group sparse) is applied on the partial derivatives if $\mathcal{S}$ along the x and z axes, i.e.

$$\begin{aligned}J_2(\mathcal{S})_{L_{2,1}} &= \|W \odot D\mathcal{S}\|_{2,1}\\ &= \sum_{n=0}^{N} W_x \|D_x\mathcal{S}\|_2 + \sum_{n=0}^{N} W_z \|D_z\mathcal{S}\|_2\\ &= \sum_{n=0}^{N}\sqrt{\sum_{m}^{M} D_x\mathcal{S}(:,n,l)} + \sum_{n=0}^{N}\sqrt{\sum_{l=0}^{L} D_z\mathcal{S}(m,n,:)}\end{aligned} \tag{7}$$

here, we define D_xS and $D_z\mathcal{S}$ as the first-order differences of stripe artifacts in the x and z direction, respectively. The variables M,N,L represent the number of pixels in the projection along the x,y, and z directions, $0 \le m,n,l \le M,N,L$. To account for the distinct sparseness in the x and z directions, we assign different weights W_x,W_z are given.

For the image tensor $\mathcal{U}$, the anisotropic TV constraint is adopted, i.e.

$$J_1(\mathcal{U}) = \|D\mathcal{U}\|_F = \|D_x\mathcal{U}\|_1 + \|D_y\mathcal{U}\|_1 + \|D_z\mathcal{U}\|_1 \tag{8}$$

where

$$\begin{aligned}D_x\mathcal{U} &= D_x\mathcal{U}(:,j,k) = \mathcal{U}(i+1,j,k) - \mathcal{U}(i,j,k)\\ D_y\mathcal{U} &= D_y\mathcal{U}(i,:,k) = \mathcal{U}(i,j+1,k) - \mathcal{U}(i,j,k)\\ D_z\mathcal{U} &= D_z\mathcal{U}(i,j,:) = \mathcal{U}(i,j,k+1) - \mathcal{U}(i,j,k)\end{aligned}$$

and I,J,K are the number of pixels in the x, y, and z directions of the reconstructed image respectively, $0 \le i,j,k \le I,J,K$.

To summarize, the proposed optimization model for ring artifact correction is as follows:

$$\begin{aligned}\underset{\mathcal{U},\mathcal{S}}{\operatorname{argmin}} &\|R\mathcal{U} + \mathcal{S} - \mathcal{Y}\|_F^2 + \lambda\|D\mathcal{U}\|_F + \mu\|W \odot D\mathcal{S}\|_{2,1}\\ &+v\,\|\mathcal{S} - \mathcal{C} \times_1 \mathrm{A} \times_2 \mathrm{B} \times_3 \mathrm{C}\|_F^2\\ \text{s.t. } &\mathrm{A}^T\mathrm{A} = \mathrm{I}\\ &\mathrm{B}^T\mathrm{B} = \mathrm{I}\\ &\mathrm{C}^T\mathrm{C} = \mathrm{I}\end{aligned} \tag{9}$$

where λ, μ, ν represent the parameters associated with the regularization terms.

B. The proposed optimization model

The variable separation approach is employed to solve the proposed optimization model (9). By decomposing model (9) into two independent sub-problems, we update the image tensor $\mathcal{U}$ and the ring tensor $\mathcal{S}$ alternately. The flowchart illustrating our method is depicted in Fig. 1.

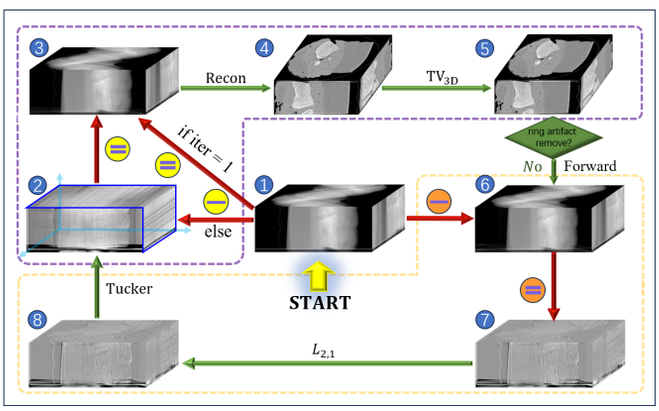


Fig. 1. Flow chart of TLT algorithm. ① is the projection containing ring artifact, ② is the stripe artifact, ③ is the remaining projection from ① minus ②, ④ is the reconstructed image of ③, ⑤ is the result of ④ after anisotropic TV, ⑥ is the corresponding projection of ⑤, ⑦ is the remaining projection from ① minus ⑥ contained stripe information ,⑧ is corresponding projection of ⑦ after $L_{2,1}$ operation, and the stripe information in ② is gathered through Tucker decomposition from ⑧.

1) Sub-problems for $\mathcal{U}$:

The reformulation leads model (9) to an equivalent optimization problem, which can be expressed as follows:

$$\arg\min_{\mathcal{U}} \|R\mathcal{U}+\mathcal{S}-\mathcal{Y}\|_F^2+\lambda\left(\|D_x\mathcal{U}\|_1+\|D_y\mathcal{U}\|_1+\|D_z\mathcal{U}\|_1\right) \tag{10}$$

And the solution process of $\mathcal{U}$ corresponds to the purple box of the flow chart. For the reconstruction process from ③ to ④, we utilize the Simultaneous Algebraic Reconstruction Technique (SART) algorithm.

2) Sub-problems for $\mathcal{S}$:

Fixed other variables, we obtain an equivalent optimization problem to model (9), represented as follows:

$$\begin{aligned}\arg\min_{\mathcal{S}} &\|\mathcal{P}_u+\mathcal{S}-\mathcal{Y}\|_F^2+\mu\|W\odot D\mathcal{S}\|_{2,1}\\ &+v\left\|\mathcal{S}-\mathcal{C}\times_1 \mathrm{A}\times_2 \mathrm{B}\times_3 \mathrm{C}\right\|_F^2\end{aligned} \tag{11}$$

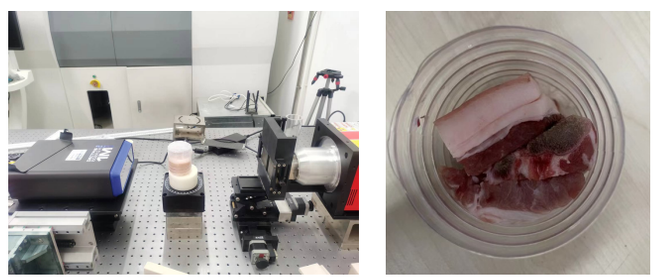
The solution process of Sub-problems 2 is the yellow box of the flow chart.

For the optimization process, an efficient Alternating Direction Method of Multipliers (ADMM) or Augmented Lagrangian Method (ALM) method can be designed to optimize the tensors $\mathcal{U}$ and $\mathcal{S}$ in the proposed model.

III. Experiments and Results

This section verifies the effectiveness of the proposed model and algorithm. In addition to the SART reconstruction algorithm, we compare our method with four state-of-the-art algorithms, i.e. FW method [8], the MTG method [6], the LRTD-SSTV method [9], and the RRRTV method [11].

In the real experiment, the scanning voltage is 100kV, and the scanning current is 0.3mA. In addition, the size of the EIGER2 1M-WR-DECTRIS detector pixel is 0.075mm. The raw data is acquired with an EIGER2 1M-WR-DECTRIS detector, as shown in Fig. 2(a), and the sample is pork-bone sandwich sample as shown in Fig. 2(b).



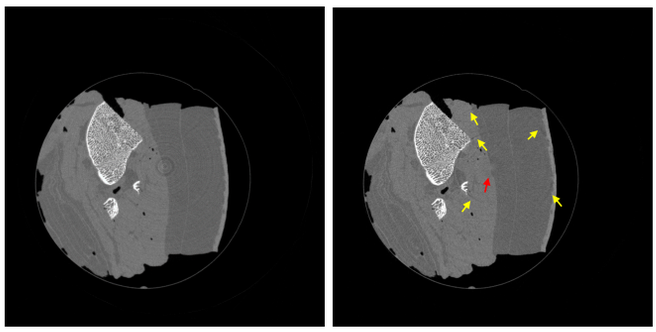
(a) (b)

Fig. 2. (a) The CT scanning system with an EIGER2 1M-W R-DECTRIS detector (b) The pork-bone sandwich sample.

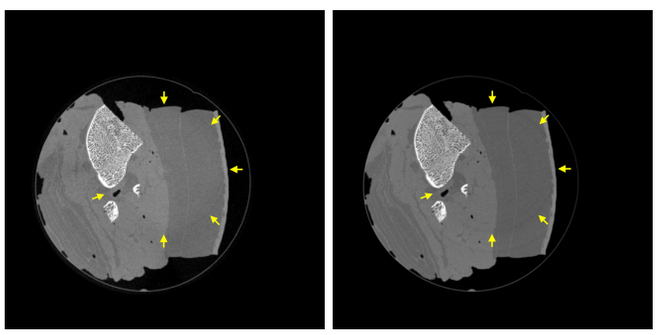
In this subsection, one representative case is selected for visual comparison, and the results are shown in Fig. 3. Subfigure (a) is the uncorrected reconstruction result from the SART algorithm, subfigures (b)-(e) present the reconstruction results of the comparison algorithm, and subfigure (f) shows the result of the proposed method. From subfigure (b), we observed that the results from the FW method exhibit flower-shaped artifacts near the central position, as indicated by the red arrow. Additionally, secondary artifacts are introduced at the location pointed by the yellow arrow. In subfigures (c) and (d), the MTG and RRRTV methods tend to produce bright or dark circle-shaped artifacts in areas where the sample is present, especially when the iteration number is large. Similar to the FW method, there are also significant secondary artifacts at the location indicated by the yellow arrow. The methods illustrated in subfigures (c)-(f), which utilize TV-type regularization terms, all have denoising effects to some extent. However, the LRTD-SSTV method introduces piecewise smooth constraints in the projection domain, which can induce blurring of the finer structures in the reconstructed image, as pointed out by the red arrow in subfigure (e). From subfigure (f), it can be seen that our algorithm is able to preserve the subtle structure more clearly and eliminate the ring artifact more effectively.

IV. Conclusions

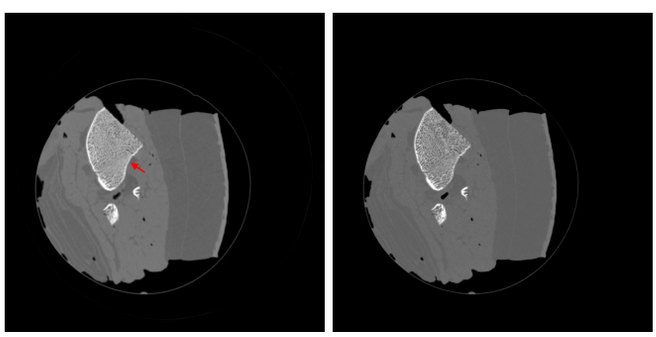
This paper introduces a novel dual-domain regularization method for removing ring artifacts in CBCT images. Our method offers two key innovations. Firstly, we observe and fully utilize the group sparsity pattern of stripe artifacts in 3D normalized scan data. This allows us to better identify and correct ring artifacts, leading to improved image quality. Secondly, we incorporate image domain constraints to stabilize the solution of the equation and maintain piecewise constant



(a) SART method (b) FW method



(c) MTG method (d) RRRTV method



(e) LRTD-SSTV method (f) TLT method

Fig. 3. The 255th slice in the reconstructed result in CBCT reconstructed results.The display window is set to [0, 0.06].

properties. This enables us to leverage prior information in both the dual domains, resulting in more accurate correction coefficients and successful removal of ring artifacts. Overall, our method provides a promising approach for enhancing the quality of CBCT images by effectively eliminating ring artifacts.

REFERENCES

[1] G. Davis and J. Elliott, "X-ray microtomography scanner using time-delay integration for elimination of ring artefacts in the reconstructed image," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, pp. 157–162, Jul 1997.

[2] D. M. Pelt and D. Y. Parkinson, "Ring artifact reduction in synchrotron x-ray tomography through helical acquisition," *Measurement Science and Technology*, p. 034002, Mar 2018.

[3] A. L. C. Kwan, J. A. Seibert, and J. M. Boone, "An improved method for flat-field correction of flat panel x-ray detector," *Medical Physics*, pp. 391–393, Feb 2006.

[4] R. A. Ketcham, "New algorithms for ring artifact removal," in *SPIE Proceedings,Developments in X-Ray Tomography V*, Aug 2006.

[5] S. Rudin, A. R. Podgorsak, S. V. Setlur Nagesh, D. R. Bednarek, and C. N. Ionita, "Use of a cmos-based micro-ct system to validate a ring artifact correction algorithm on low-dose image data," in *Medical Imaging 2018: Physics of Medical Imaging*, Mar 2018.

[6] K. An, J. Wang, R. Zhou, F. Liu, and W. Wu, "Ring-artifacts removal for photon-counting ct," *Optics Express*, p. 25180, Aug 2020.

[7] X. Guo, W. Wu, X. Duan, H. Yu, D. Chang, P. He, J. Wang, R. Zhou, Y. Du, and K. An, "Nonlinear deviation correction for ring-artifact removal with cone-beam computed tomography," *IEEE Transactions on Instrumentation and Measurement*, pp. 1–10, Jan 2022.

[8] B. MÃijnch, P. Trtik, F. Marone, and M. Stampanoni, "Stripe and ring artifact removal with combined wavelet-fourier filtering," *Optics Express*, p. 8567, May 2009.

[9] Y. Li, Y. Zhao, D. Ji, S. Han, M. Zheng, W. Lv, X. Xin, F. Li, X. Zhao, D. Liu, and C. Hu, "3d ring artifacts removal algorithm combined low-rank tensor decomposition with spatial-sequential total variation regularization and its application in phase-contrast microtomography," *Medical Physics*, pp. 393–410, Jan 2022.

[10] Q. Huo, J. Li, Y. Lu, and Z. Yan, "Removing ring artifacts in cbct images via l0 smoothing," *International Journal of Imaging Systems and Technology*, pp. 284–294, Nov 2016.

[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," *Medical Physics*, pp. 4803–4815, Nov 2019.

[12] K. Murata and K. Ogawa, "Ring-artifact correction with total-variation regularization for material images in photon-counting ct," *IEEE Transactions on Radiation and Plasma Medical Sciences*, pp. 568–577, Jul 2021.

[13] Y. Liu, C. Wei, and Q. Xu, "Detector shifting and deep learning based ring artifact correction method for low-dose ct," *Medical Physics*, vol. 50, pp. 4308–4324, Jul 2023.

[14] T. Fu, Y. Wang, K. Zhang, J. Zhang, S. Wang, W. Huang, Y. Wang, C. Yao, C. Zhou, and Q. Yuan, "Deep-learning-based ring artifact correction for tomographic reconstruction," *Journal of Synchrotron Radiation*, Aug 2022.

[15] Y. Liu, K. Weiss, N. Navab, C. Marr, J. Huisken, and T. Peng, "Destripe: A self2self spatio-spectral graph neural network with unfolded hessian for stripe artifact removal in light-sheet microscopy," *Medical Image Computing and Computer Assisted Intervention - MICCAI 2022*, Jun 2022.

[16] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, pp. 455–500, Aug 2009.

A CL scanning with a single spiral trajectory

Xiang Zou, Yuxiang Xing*

***Abstract*—Computed laminography (CL) technology is widely used for nondestructive testing of plate-shaped objects, such as printed circuit boards and composite panels. To improve the image reconstruction quality, a single spiral scanning trajectory for rotational CL was proposed, which can increase the incoherence of projection data, thereby help to improve reconstruction. A numerical experiment was carried out to examine the gain of the new trajectory. It was found that the RMSE of new trajectory is reduced compared with a circular trajectory in the numerical experiment. Finally, the influence of projection number, x-ray source translation displacement and noise level were analyzed and discussed.**

***Index Terms*—Computed laminography, X-ray imaging, Scanning trajectory, Image reconstruction**

I. Introduction

X-ray computed tomography (CT) is increasingly used in industry as a method for nondestructive testing and quality control. However, due to limitations in scanning space and radiation energy, CT is not suitable for plate-shaped objects, such as integrated circuits, ball grid arrays, and printed circuit boards. Meanwhile, computed laminography (CL) has recently received increasing attention due to it has unique advantages for the imaging plate-shaped objects.

According to the difference of scanning trajectory, CL can be divided into planar CL, swing CL, rotational CL and so on. In these methods, rotational CL is widely used due to its advantages of rich projection information and simple structure. Figure 1(a) shows the rotational CL geometry, where the x-ray source and detector rotate 360° around the measured object to collect projection information from different directions. Based on the obtained projection information, the volumetric image of the object can be obtained using reconstruction algorithms. Therefore, information in projections and reconstruction algorithms are the two major factors affecting the quality of reconstructed images. Meanwhile, the information in projection is mainly affected by the scanning trajectory.

Researches on rotational CL mainly focus on reconstruction algorithms, such as Yang et al. proposed an analytical reconstruction algorithm for rotational CL in 2010 [2]. A total variation minimization algorithm was used in rotational CL reconstruction, and the result showed that regularization processing can improve the quality of CL image [3]. However, there are relatively few researches on the rotational CL scanning trajectory. One study is by Abbas et al. who proposed a new CL scanning trajectory called spherical-sinusoidal scanning trajectory, and verified its effectiveness through numerical experiments [4]. However, this trajectory is relatively complex and has low feasibility in practical applications.

In this work, we explored a single spiral scanning trajectory for rotational CL, which has potential to lower the coherence between adjacent projection angles and improve the image quality of reconstruction. A detailed description of the trajectory was firstly provided. We examined the performance of such a trajectory with a simulation study, the impact of imaging parameters was analyzed afterwards.

II. A rotational CL with a Single spiral trajectory

Figure 1 (a) shows the schematic illustration of a rotational CL scan with a circular trajectory in source. The detector moves according to the source to cover a certain field of view. Figure 1(b) shows the corresponding Radon shell on the frontal plane. As we can see that the data of CL in Radon space is asymmetrically distributed but incomplete, which leads to well-known fact that CL scan produces images with artifacts, usually severe aliasing between slices. For sparse-view CL in on-line industrial applications, image quality further degrades.

It is interesting to study trajectories other than a commonly used circular trajectory. In this study, we explore a single spiral trajectory for a rotational CL scan. The single spiral trajectory of source can be realized by adding a uniform translation motion along the rotation axis to the x-ray source on the basis of a circular trajectory as shown in Figure 1(c). We set the starting height of the spiral to be z_0 that is same as the height of a circular trajectory. The source moves downwards, and the displacement is denoted as z_1. A frontal plane view of the Radon space for the single spiral CL scan is plotted in Figure 1 (d). We can see that the Radon shell covers most of the bottom part, but with bigger blank region than a circular trajectory. Moreover, Abbas et al. [2] found that increasing the distance between adjacent imaging positions in CL imaging can improve the incoherence between rays, thereby improving the accuracy of iterative reconstruction. In this way, we are hoping to see some different effect in reconstruction.

"This work was supported by National Key R&D Program of China (grant No.2022YFF0607800)."

The authors 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 (e-mail: xingyx@mail.tsinghua.edu.cn; zoux@mail.tsinghua.edu.cn).

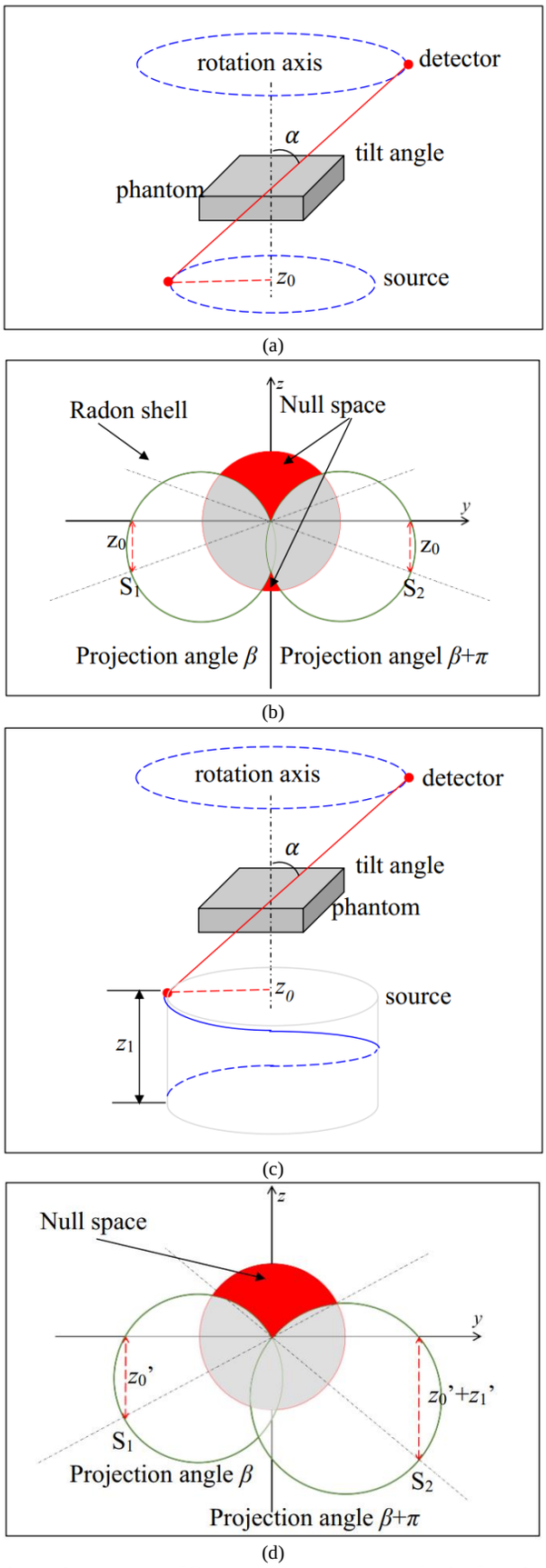


Fig. 1. Imaging geometry of CL scans: (a) scan geometry of a circular trajectory CL, (b) Illustration of the corresponding Radon shell by its frontal plane for the circular trajectory CL. (c) scan geometry of a spiral trajectory CL, (d) Illustration of the corresponding Radon shell by its frontal plane for the spiral trajectory CL.

To study the pure effect of trajectories, we do the reconstruction by the inverse of:

$$\mathbf{H}\boldsymbol{\mu} = \mathbf{p} \tag{1}$$

where $\boldsymbol{\mu}$ is a vector of linear attenuation coefficient of a 3D scanned object, $\mathbf{H}$ is the system matrix with each element $\mathbf{H}_{ij}$ denoting the contribution to i^{th} projection data from j^{th} voxel, $\mathbf{p}$ is a vector of the projection data. A simultaneous iterative reconstruction technique (SIRT) [4] was used to reconstruct the three-dimensional image of a scanned object.

III. Experimental Results

We did a simulation study on a spiral trajectory CL with a simulated printed circuit board (PCB) phantom as shown in Fig.2. The parameters for scan geometry are listed in Table 1. The results of circular trajectory were also given for comparison.

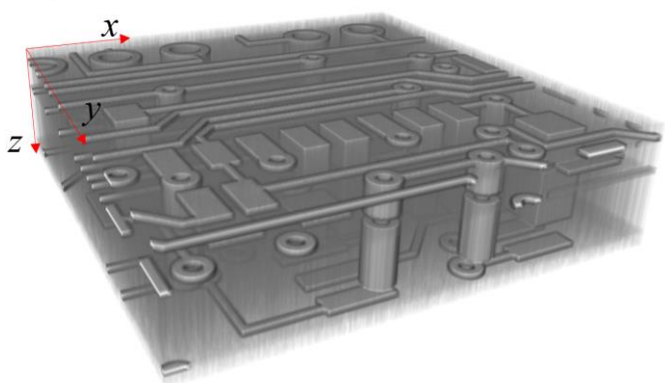


Fig. 2. PCB phantom, the size is 300×300×80 in *xyz* direction, the display window is [20,187].

TABLE 1
IMAGING PARAMETERS

Distance between source and object (mm)	45.790
Distance between source and detector (mm)	194.580
Size of detector (pixel)	2048×2048
Resolution of detector (mm)	0.17
Size of reconstructed image(voxel)	300×300×80
Resolution of reconstructed image (mm)	0.05
Tilt angle α (°)	60
Number of projection views	16
z_1 (mm)	40
Reconstruction iteration	200

Figure 3 shows the reconstructed images under two trajectories at the central slice (z=40). It can be seen that the main internal structures of the phantom can be reconstructed in both situations. However, due to the small number of projections and the incomplete projection data in rotational CL imaging, there are artifacts in both reconstruction results. Furthermore, it can also be noted that, the artifacts are mainly located at the edges of the internal structure of the phantom for the circular trajectory, while the aliasing in edges are reduced to a certain extent for the single spiral trajectory.

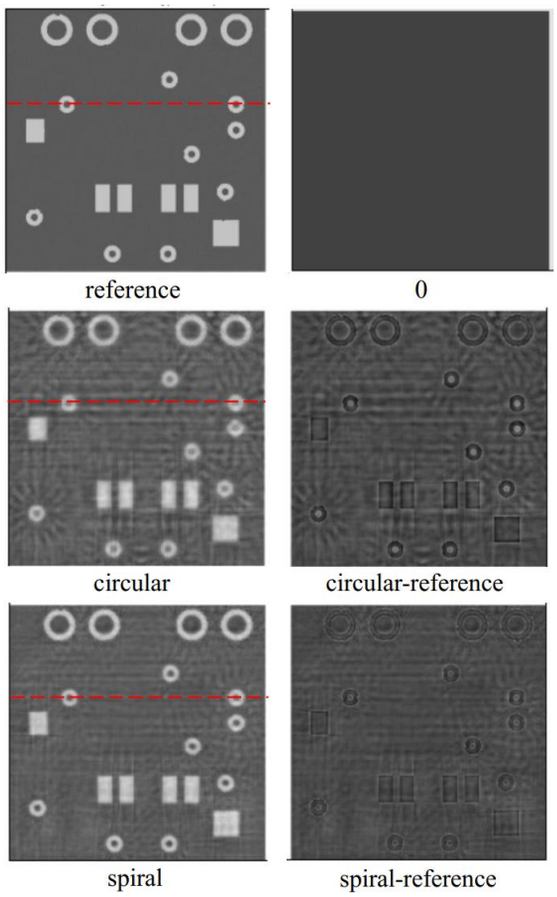
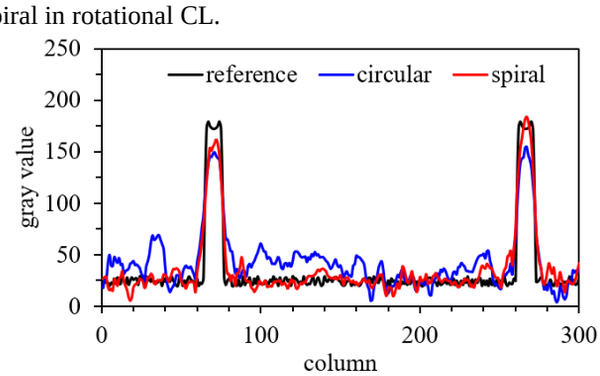
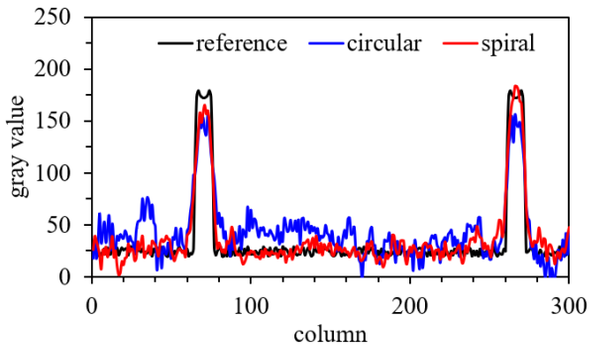


Fig. 3. Comparison of reconstructed and original images under two trajectories, the right column are the difference images and the display window is [-100, 255].

Figure 4 gives the horizontal profiles of the reconstruction along the line indicated in Figure 3. It can be seen that the fluctuation of circular trajectory is greater than that of the single spiral trajectory. The RMSE of the single spiral trajectory is smaller compared to the circular trajectory. The above results indicate that it is possible to improve the reconstruction quality by changing the x-ray source trajectory from circular to single spiral in rotational CL.



(a)



(b)

Fig. 4. Horizontal profiles along the red line in Fig.3: (a) without noise, (b) with noise

In a single spiral trajectory, the number of projections, the translation distance of the source along the rotation axis and noise level are important parameters that affect the image quality of reconstructions. Therefore, we also analyzed the influence of these three parameters through numerical experiments.

To analyze the influence of projection number, the reconstruction results with five different projection numbers are compared using numerical experiment, and other parameters are set according to Table. 1. Figure 5 shows the RMSEs of two trajectories as a function of projection number. As shown, when the number of projection views is small, the RMSE of reconstructions from the single spiral trajectory is smaller than that from the circular trajectory. However, when the number of projection views increases to a certain level, on the contrary, the RMSE of reconstruction from the circular trajectory is smaller. This is mainly because the single spiral trajectory increases the incoherence between x-ray projection equations. When the number of projection views is small, the improvement of incoherence between x-ray projection equations can significantly improve the reconstruction quality. As the number of projection views increases, the number of x-ray projection equations also increases sharply. At this time, the incoherence between adjacent x-ray projection equations is greatly reduced. Moreover, due to the more severe missing projection data of single spiral trajectories, their reconstruction errors begin to be greater than those of circular trajectories.

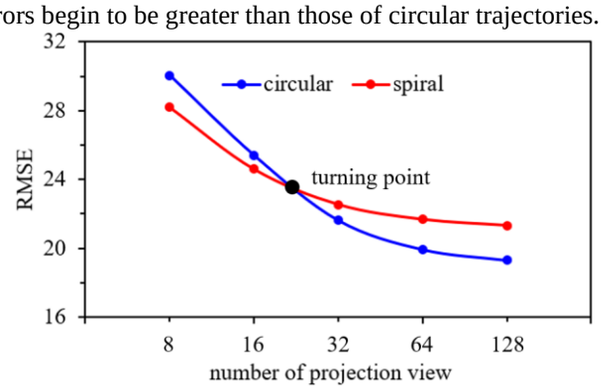


Fig. 5. Influence of projection number.

To analyze the influence of source displacement, the reconstruction results of five different displacements are compared using numerical experiment, and other parameters are set according to Table. 1. Figure 6 shows the RMSEs of single spiral trajectory under different displacement z_1. It can be

seen that, with the increase of z_1, RMSE shows a trend of first decreasing and then increasing. This is mainly due to the fact that when z_1 is small, the coherence reduction of the x-ray equation results in a greater improvement in imaging quality than the degradation of image quality caused by missing projection data, resulting in a decrease in RMSE. When z_1 increases to a certain extent, the incoherence effect of the x-ray equation becomes weaker than that of the missing projection data, so RMSE gradually increases.

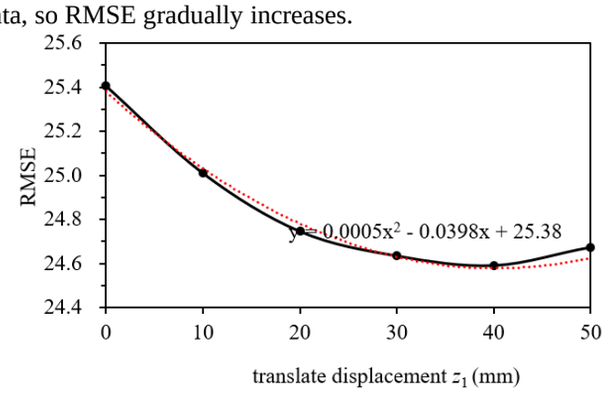


Fig. 6. Influence of translation displacement

To analyze the influence of noise, white Gauss noises with different standard deviations are added to the projections, and other parameters are set according to Table. 1. Figure 7 shows the RMSEs of two trajectories as a function of noise. It can be seen that, the RMSEs of reconstructions from two trajectories both increase with the increase of the noise. However, the RMSE of the reconstruction from the single spiral trajectory is smaller than that from the circular trajectory at each noise level.

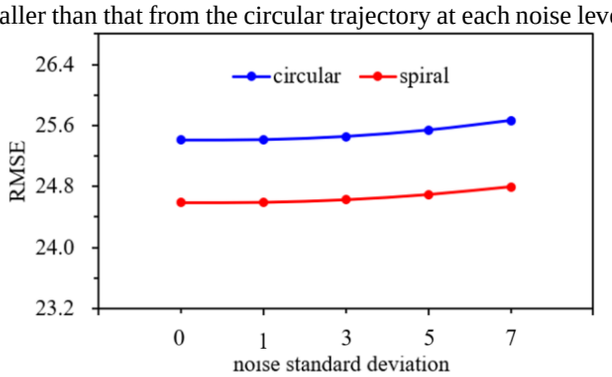


Fig. 7. Influence of noise.

IV. Discussion and Conclusion

In the single spiral trajectory, the distance between the x-ray source and the origin gradually increases during the imaging process. Therefore, the missing projection of the single spiral trajectory is more severe than that of circular trajectory, but the reconstruction results contradict this pattern. This is mainly due to the fact that although the downward movement of the x-ray source exacerbates the missing projection data, it also decreases the coherence of the projection data between different rays. At this time, the improvement in reconstruction results caused by the increase in incoherence of the projection data exceeds the increase in reconstruction error caused by the deterioration of data missing, thus improving the reconstruction quality when using iterative algorithms for reconstruction.

In summary, we proposed a single spiral scanning trajectory for rotational CL in this work, and the simulation results show that this trajectory may improve imaging quality when the number of projection views is limited. This study provides some insight for the design of CL scan trajectory for high-speed on-line inspection in industrial fields.

References

[1] M. Yang, G. Wang, and Y.Z. Liu, “New reconstruction method for x-ray testing of multilayer printed circuit board,” Opt. Eng., vol. 49, no. 5, pp. 56501, May. 2010.

[2] S. Abbas, J.P. Hong, J.R. Lee and S. Cho, “Effects of Data Correlation in a Sparse-View Compressive Sensing Based Image Reconstruction,” Int. J. Biometeorol, vol. 8, no. 9, pp. 582-585, Jul. 2014.

[3] S. Abbas, M. Park, J. Min, H.K. Kim and S. Cho, “Sparse-view computed laminography with a spherical sinusoidal scan for nondestructive testing,” Opt. Exp., vol. 22, no. 15, pp. 17745-17755, Jun. 2014.

[4] X.Y. Guo, L. Zhang and Y.X Xing, “Analytical covariance estimation for iterative CT reconstruction methods,” Biomed. Phys. Eng. Expr., vol. 8, no. 3, pp. 1-15, Feb. 2020.

Using a Patch-Wise Deep U-Net CNN for the Segmentation of Pores in Carbon Fiber Reinforced Polymers

Ankit Kumar, Johann Kastner and Christoph Heinzl

***Abstract*—This paper shows the usage of a deep convolutional neural network to segment the pores in an X-Ray computed tomography data acquired from carbon fiber reinforced polymer specimens. Deep learning in the recent years has been increasingly used in the CT domain with highly promising performance for both segmentation and classification tasks. In particular, the U-Net architecture has been widely adopted for this purpose, mainly in terms of the segmentation of defects such as pores and cracks. In this work we propose a U-Net architecture, which uses non-overlapping patch-wise image inputs as compared to the conventional U-Net architecture and which delivers similar results compared to the current state-of-the-art methods, but with the additional advantage of retention of more local information. The non-overlapping patches are fed to the proposed U-Net model along with their corresponding ground truth to train the network. The experimental results show that the proposed patch-wise U-Net model achieves a Dice similarity coefficient (DSC) score of 0.96.**

***Index Terms*—U-Net, segmentation, carbon fiber reinforced polymers, deep learning, X-ray computed tomography.**

I. Introduction

During the past few years, deep convolutional networks have performed exceptionally well with respect to the state-of-the-art in many visual recognition techniques [1]. Deep learning is getting increasingly popular in various application domains [2,3]. It has been widely used for segmentation and detailed analysis of interesting features, e.g., obtained in 3D datasets as generated by X-ray computed tomography (XCT). The trend of using deep learning CNN models started with the bio-medical domain research and yielded very efficient and accurate results for the segmentation of organs and tumors. Apparently, the CNN architectures are also of great use to the material science domain with the hindrances of in-depth characterization of material structures, requiring specialized and resource intensive and after all expensive software coupled to domain knowledge [4]. Furthermore, to explore, evaluate and present the tool's results, visualization techniques are of great use to support the domain experts in their material analyses and respective decision making [5]. Several evaluation tools and visualization techniques are available, e.g., in open_iA [6], VGStudio Max or other tools. However, there have been limitations of handling the larger dataset volumes for efficient feature segmentation. Linked to this issue, it also remains difficult to maintain local details because entire image needs to be fed into the network.

To overcome these problems, we propose a patch-wise deep CNN U-Net model for efficient segmentation of pores for the carbon fiber reinforced polymer generated by the XCT. As the partitioning with individual patches retains more of local information, the model can be trained better with local details in non-overlapping patches which are made with non-overlapping square partitioning, resulting in higher segmentation performance with efficient computational complexity unlike conventional U-Net. The bio-medical domain research also investigated a similar concept for patch-wise deep learning models [7] for the segmentation of cancer cells from the brain MRI images, which applies complex architectures and requires high amount of data sets and high computing time with powerful computing system. Our proposed method doesn't require high amount of data set and allows quick segmentation of pores in an efficient manner with less computing power applied to the XCT data. We demonstrate our results in contrast to conventional approaches with training and validation accuracy of 0.995.

II. Related Works

Extracting the features and its quantification is a very essential requirement for quantitative and visual analysis of defects like cracks, pores, and voids in the material science domain. Diverse characteristic properties are found in every feature extracted. This characteristic information may contain many properties depending on the feature. Several 3D volume segmentation approaches play a vital role in extraction of these features and more importantly for determination of detailed characteristics in particular areas. A categorization of

A. Kumar and J. Kastner are with CT Research Group, University of Applied Sciences Upper Austria, Stelzhamerstraße 15, Wels, Austria
C. Heinzl is with University of Passau, Innstraße 41, Passau, Germany
Corresponding author: Ankit Kumar (ankit952069@gmail.com)

segmentation methods can be done in 4 sub-groups: edge, pixel, region, and model-based approaches. One of the easiest and fastest approach is the Pixel based segmentation in which, it depends just on the grey value level of the image pixel. In the recent trends, deep learning has flourished quite efficiently into image and volume segmentation. It had been applied on medical science domain for medical image segmentation and now with it's success other domains like material science, automobiles and linguistics has also tried to implement it. Both 2D and 3D segmentations in deep learning has been an integral part for the image segmentation depending upon the usage. Several methods like image segmentation, transfer learning and instance segmentation approaches are vividly explained by Garcia. For XCT there are several approaches like image segmentation approach, intensity-based thresholding like Otsu or ISO50 are widely used in material science domain. These techniques help to attain a quick segmentation of defects like pores, inclusions, or matrix. More promising techniques must be used for observing the local information of data at the specific region in more details. In the recent trend, simplified neural network architecture-based approaches shows the potential to resolve this issue and segment the pores in the XCT image with more retention of local information.

III. METHOD

The SegNet [8] and U-Net [9] architectures are popularly used for different segmentation applications. Similar to the SegNet[8], the proposed deep U-Net consists of encoder and decoder parts, where the encoding part basically down samples the input image using multiple convolution matrices and max-pooling operations, and the decoding part subsequently upsamples the down sampled feature maps using the max-pooling indices and convolution operations. The Rectified linear unit (ReLU) activation function is used along the convolution layer to improve the computational speed of the training stage of the neural network. Additionally, the feature maps in the decoder part are concatenated to the feature maps corresponding in the encoder path during the upsampling process which exploits the better localization property to its optimal. The upsampling process also has a large number of feature channels. The sigmoid function is used as an activation function at the output layer.

For the preprocessing, the input volumetric image is normalized, then padded with 5 voxels mirror-padding to the XCT scan image which results in the size of 132×132×132 volumetric image. This image is then split into optimal sub-volumes possible of dimension 128×128×128, which could undergo into further sub-divisions of patches and also is used for creating the training data samples. The labelled image was also split into sub-volumes as shown in Figure [1]. For further training of our neural network architecture network, we use 80% of sub-volumes for training and remaining 20% for test. This input image is then divided into four non-overlapping patches and fed to the network as shown in Figure 1.

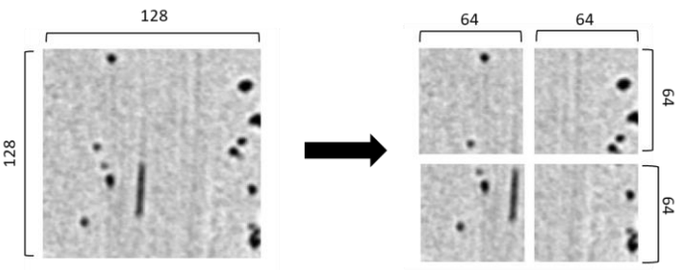


Figure 1: Splitting the input image into non-overlapping patch inputs.

For the deep patch-wise U-net architecture, the feature maps present in the decoding path are concatenated with corresponding feature maps in the encoding path during upsampling. This approach helps for better localization and the patch-wise inputs as shown in Figure 2, focuses to address the retention of local information in small patches while training the network. With the larger input image size, relatively more GPU memory would be consumed to train the network. One of the concerns is that when the input image is fed to the network, it might miss the certain regions details of the input image. To overcome this issue the patch-wise inputs help to cope up and the main advantage achieved is the better localization during the network training as it focuses more on spatial local details. The overall architecture of the proposed deep convolutional neural network U-net model is illustrated below in Figure 2.

The Sørensen-Dice coefficient (DSC) method is a statistical approach, which measures the overlapping of the predicted label with the label in the reference as shown mathematically in equation (1). If the result of the method is one, it explains that the prediction and the reference are the same. Whereas, if the result is zero, it indicates there is no similarity in between the prediction and the reference image.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \tag{1}$$

The optimizers used in the CNN training try to find the smallest possible value for the loss. Therefore, the dice coefficient has to be inverted to be usable as loss. This can be achieved by subtracting the DSC from equation (1) and is mathematically expressed in form of equation (2).

$$Dice\ loss = 1 - DSC \tag{2}$$

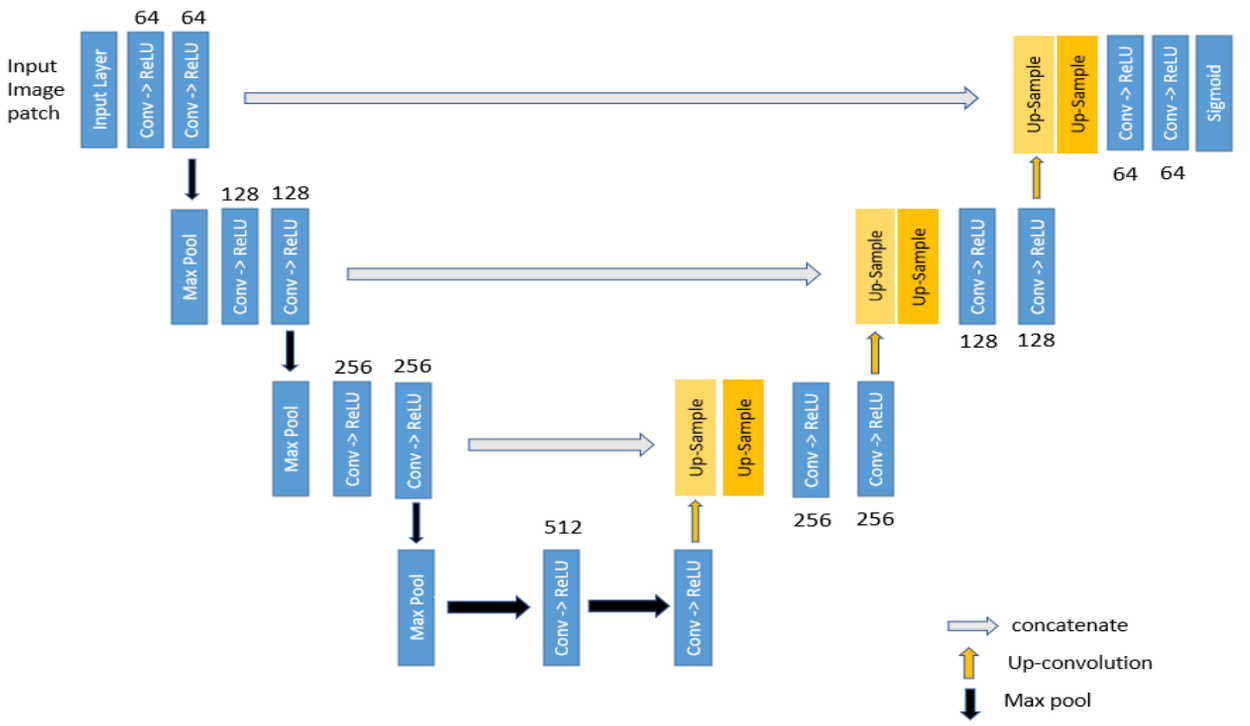


Figure 2. Overall architecture of the proposed deep U-net.

IV. RESULTS

For the demonstration of our results, we scanned a CFRP specimen using a GE phoenix X-ray Nanotom XCT with a voltage of 60kV and voxel size of (2.5µm)2. For the training, an Intel® X® W-2295 CPU @ 3.00GHz workstation with Nvidia RTX A6000 was used. The training dataset consisted of 560 samples created from the sub-volumes and was splitted into 80% of samples to be used for training and 20% for testing. The Adam [10] optimizer is applied with the optimal learning rate of 4e-5 and trained the network for 25 epochs to get better segmentation results. The Dice coefficient loss function was used, and we achieved DSC score [11] of 0.9647 with training accuracy of 0.9951 which is better than the conventional U-Net model for the segmentation of pores which has the DSC score of 0.9356, especially with the main advantage of retention of the local information exploited patch-wise. This approach is better for smaller resolution dataset where the retention of the local information is achieved better than the conventional method with less parameters. The validation dice coefficient score was 0.9558 with validation accuracy of 0.9952.

V. CONCLUSION

This paper has shown a novel patch-wise deep U-net CNN architecture that has been used to segment the pores in carbon reinforced fiber polymer XCT dataset. To setup a proper pipeline for getting reasonable and reusable training data is equally important. The resulting neural network made it possible to produce and achieve reasonable segmentation results of XCT data. Also, it is impressive that the prediction on the real world XCT dataset using the latest hardware, gives us the possibility to use this approach of segmentation for real industry applications. Although, the quality and size of reference data used for training has always been a concern and limits the achievable quality of CNN segmentation. The proposed network also handles the big datasets in combination with multi-GPU systems for fast loading and processing the data for training. Also, this approach is better for smaller resolution datasets where the retention of the local information is achieved better than the conventional method with less parameters. Our future work will be focused on usage of improved reference data for training and achieve more better segmentation results on the real world XCT applications.

ACKNOWLEDGEMNTS

The research leading to these results has received funding by research subsidies granted by the European Union's Horizon 2020 research and innovation program under grant no. 956172, project "xCTing".

REFERENCES

[1] 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 (MICCAI)], 234–241, Springer, Cham (2015), doi: 10.1007/978-3-319-24574-4_28.

[2] Y. Jing, S. Li, Z. Wang, H. Dong, J. Wang, and S. Tang, “Using deep learning to detect defects inmanufacturing: A comprehensive survey and current challenges,” Materials 13, 5755 (12 2020), doi: 10.3390/ma13245755.

[3] S. Minaee, Y. Y. Boykov, F. Porikli, A. J. Plaza, N. Kehtarnavaz, and D. Terzopoulos, “Image segmentation using deep learning: A survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 1–1 (2021), doi: 10.1109/TPAMI.2021.3059968.

[4] P. Weinberger, B. Plank, B. Fröhler, J. Kastner, C. Heinzl, "Segmentation of multiple features in carbon fiber reinforced polymers using a convolutional neural network", Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XVI, SPIE, 2022, doi: 10.1117/12.2612895.

[5] C. Heinzl and S. Stappen, “STAR: Visual computing in materials science,” Computer Graphics Forum, vol. 36, no. 3, pp. 647–666, 2017, doi: 10.1111/cgf.13214

[6] B. Fröhler, J. Weissenböck, M. Schiwarth, J. Kastner, C. Heinzl, open_iA : A tool for processing and visual analysis of industrial computed tomography datasets, Journal of Open Source Software, 4 (35), 1185, 2019, doi: 10.21105/joss.01185.

[7] B. Lee, Yamanakkanavar N, and J.Y. Choi, "Automatic segmentation of brain MRI using a novel patch-wise U-Net deep architecture" PLoS ONE 15(8): e0236493, doi: 10.1371/journal.pone.0236493.

[8] V. Badrinarayanan, A. Kendall, and R. Cipolla. "Segnet: A deep convolutional encoder-decoder architecture for image segmentation." IEEE transactions on pattern analysis and machine intelligence 39, no. 12 (2017): 2481-2495, doi: 10.1109/TPAMI.2016.2644615.

[9] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation, In Proc. Int. Conf. Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2016, doi: 10.1007/978-3-319-46723-8_49.

[10] D. P. Kingma, J. Ba, “Adam: A method for stochastic optimization,” in [International Conference for Learning Representations], (2017), doi: 10.48550/arXiv.1412.6980.

[11] T. Sørensen, “A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on danish commons,” Kongelige Danske Videnskabernes Selskab, Biologiske Skrifter, 1–34 (1948), doi: 10.1234/12345678.

Double bowtie design for high sensitivity pediatric spectral CT

Olivia F. Sandvold[1,2], Yinglin Ge[1,2], Roland Proksa[1], Peter B. Noël[1]

***Abstract*—Despite the evident benefits of spectral computed tomography (CT) in delivering qualitative imaging superior to that of conventional CT in adults, its application in pediatric diagnostic imaging is still relatively limited due to various reasons, including design limitations and radiation dose considerations. The use of specialized K-edge filters, in conjunction with other spectral technologies, has been demonstrated to improve spectral quantification accuracy. X-ray flux limitations generally pose challenges in these concepts when applied to adults. However, such limitations are not present in pediatric imaging, allowing the full exploitation of K-edge filters to improve performance. To facilitate the adoption of spectral CT's benefits, as seen in the adult population, into pediatric settings, we introduce an innovative double bowtie filter design. This design incorporates a K-edge material coupled with Teflon and is integrated with rapid kVp-switching technology. A Python simulation was built to model a rapid kVp-switching x-ray tube and to estimate Cramer-Rao lower bound (CRLB) noise in photoelectric and Compton scatter basis domains. We estimate a conventional bowtie filter and corresponding reference patient dose before optimizing double bowtie configurations to contain the highest obtainable spectral signal-to-noise content for the specified phantom. Our findings indicate that an optimal combination of holmium and Teflon in the filter geometry can increase spectral SNR up to twofold the conventional estimates, while still maintaining low radiation dose exposure. This study broadens the scope for pediatric patients to fully benefit from the capabilities of spectral CT.**

***Index Terms*—Spectral CT, pediatric imaging, x-ray beam filtration, quantitative imaging**

I. Introduction

SPECTRAL computed tomography (CT) has the capability to provide increased diagnostic image quality and quantitative measurement of contrast agents, particularly iodine, to maximize the benefit of CT usage in the pediatric population. Despite strict radiation exposure dose requirements, CT remains a first-line pediatric imaging modality due to both short scan times and excellent spatial resolution [1]. Compared to conventional CT, spectral CT can generate energy- and material-specific results to mitigate effects of beam hardening, presence of metal artifacts, and rectify mis-timed bolus tracking scans without the need for a second scan [2, 3]. In adult patients, spectral CT has been shown to have superior spatial and temporal resolution with greater detectability of small, low-contrast lesions [4, 5]. Fewer studies have validated the performance of spectral CT in pediatric cases. CT systems are typically designed to optimize outcomes for a specific patient diameter, representative of an average adult-sized individual. This design, however, is suboptimal for pediatric patients. Moreover, the variation in preferred scanning tube voltages between adult and pediatric patients poses a challenge in adopting spectral CT scanning protocols for pediatric applications, due to reduced energy separation at the lower tube voltages used for children [6]. Therefore, a technical gap exists in the development and translation of spectral CT for pediatric diagnostic imaging. By tailoring advanced spectral hardware specifically for pediatric patients by utilizing current gantry concepts and maintaining a focus on low radiation doses, clinicians can benefit not only from the qualitative imaging enhancements of spectral CT but also gain quantitative CT biomarkers that aid in the diagnosis of pediatric diseases.

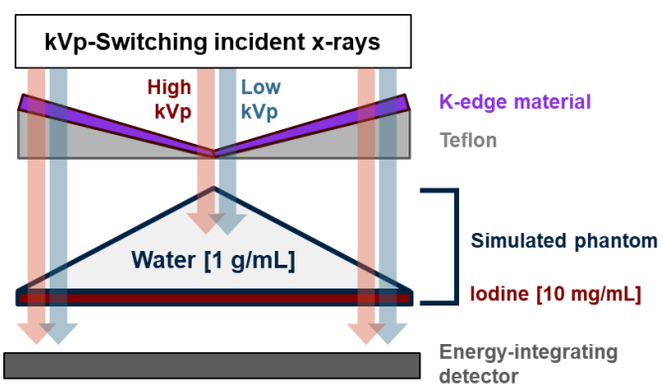


Fig. 1. Schematic depicting double bowtie shape attenuating kVp-switching x-rays for a simple water/iodine pediatric phantom.

We propose a double bowtie filter (**Fig. 1**) designed to optimize spectral hardware capabilities in a rapid kVp-switching (kVp-s) CT. Specialized filters to shape spectral response of detector hardware have been studied in several iterations of CT technology. These include silver and tin filtration as well as investigation of a multi-material spatial-spectral filter device to generate spectral beamlets [7-9]. Crucially, photon fluence constraints from x-ray generator power limitations are not approached in pediatric imaging. This allows for additional filtration while maintaining low patient doses and avoiding photon starvation. In previous simulation work, we demonstrated that a single K-edge material reduces the Cramer-Rao lower bound (CRLB) estimate of variance in the iodine domain using a dual-layer detector and varying pediatric sized phantoms [10]. We demonstrated that holmium,

We recognize support from the National Institutes of Health (R01EB030494). O.F. Sandvold, Y. Ge, R. Proksa, and P.B. Noël are with the [1]Department of Radiology, Perelman School of Medicine, Philadelphia, PA USA. O.F. Sandvold and Y. Ge are additionally with the [2]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA.

erbium, and ytterbium materials using 100 and 120 kVp tube voltage in 10 to 20 cm diameter, pediatric-sized, water phantoms had up to 11% noise difference in iodine estimation. In this abstract, we extend the idea of the single layer K-edge filter to a task-specific double bowtie design that includes the traditional Teflon material bowtie on a rapid kVp-s spectral CT system. We then optimize filter thicknesses to obtain an optimal spectral decomposition noise across the bowtie profile.

II. METHODS

To simulate the spectral performance of a rapid kVp-switching spectral CT scanner, two central x-ray spectra, denoted $\varphi_{kVp}(E)$, were modeled using manufacturer specified photon count/keV/mA/s/sr from 10 to 150 keV in single keV photon energy intervals for 140 and 80 kVp tube voltages at 100 and 73 mA tube currents, respectively, approximating for the field related emission drop from high to low. A total cycle time of 1000 ms was used with a duty cycle of 33/67, or amount of time split between high and low kV spectra respectively. The corresponding duty cycle ratio was 0.33. An ideal unit step function transition period between high and low kVp was assumed. A single layer energy-integrating detector response, $D(E)$, was modeled.

The pediatric-sized phantom comprised of water [1 g/mL] was simulated by defining an array of water path lengths from 0-150 mm in 0.5 mm increments. A simple Teflon bowtie filter was then designed to produce a homogeneous detector signal-to-noise ratio (SNR) for each water path length, x, relative to the SNR at the maximum path length where we assumed 0 mm Teflon filtration. The detector SNR for each x was estimated using $\mu(E)$, the mass attenuation [mm^{-1}] of k material in M set of basis functions={photoelectric, Compton scatter} (1). An estimated patient radiation dose exposure per sr was given by (2) where M={Teflon} using corresponding Teflon thicknesses x_k for each given water path length x. The patient radiation doses for the Teflon bowtie were used as the reference doses for double bowtie configurations.

$$SNR_{Detector}(x) = \frac{\int E\cdot\varphi_{140}(E)\cdot D(E)\cdot e^{-\sum_k^M \mu_k(E)\cdot x_k}\,dE + \int E\cdot\varphi_{80}(E)\cdot D(E)\cdot e^{-\sum_k^M \mu_k(E)\cdot x_k}\,dE}{\int E^2\cdot\varphi_{140}(E)\cdot D(E)\cdot e^{-\sum_k^M \mu_k(E)\cdot x_k}\,dE + \int E^2\cdot\varphi_{80}(E)\cdot D(E)\cdot e^{-\sum_k^M \mu_k(E)\cdot x_k}\,dE} \quad (1)$$

$$Dose(x) = \sum_{kVp}^{140,80} E\cdot\varphi_{kVp}(E)\cdot e^{-\sum_k^M \mu_k(E)\cdot x_k}\,dE \quad (2)$$

The estimated noise in the photoelectric/Compton scatter material decomposition domains was estimated with the Cramer-Rao lower bound of variance. This was accomplished by assuming each spectrum $\varphi_{kVp}(E)$ to be independent signals with Gaussian noise on each measurement, m_j, or the total energy for $j=1, \ldots, N$ measurements with distribution mean (θ) and variance (σ^2) for M={filter material(s), photoelectric, Compton scatter} (3-4). Equation 5 describes the negative log likelihood, ℓ, of obtaining said measurement under the stated conditions.

$$\theta_{kVp} = \int E\cdot\varphi_{kVp}(E)\cdot D(E)\cdot e^{-\sum_k^M \mu_k(E)\cdot x_k}\,dE \quad (3)$$

$$\sigma^2_{kVp} = \int E^2\cdot\varphi_{kVp}(E)\cdot D(E)\cdot e^{-\sum_k^M \mu_k(E)\cdot x_k}\,dE \quad (4)$$

$$\ell\left(m_1,\ldots,m_N\middle|\theta_{kVp},\sigma^2_{kVp}\right) = \frac{1}{2}\ln(2\pi) + \sum_{j=1}^{N}\left(\frac{1}{2}\ln\left(\sigma^2_{kVp}\right) + \frac{(m_j-\theta_{kVp})^2}{2\sigma^2_{kVp}}\right) \quad (5)$$

Given ℓ is dependent on input photo/scatter path lengths, and assuming an unbiased estimator, we can estimate the CRLB by computing the expectation of negative ℓ partial derivatives w.r.t to the phantom basis functions. The lower bound variance terms for each of these basis functions, σ^2_{photo} and $\sigma^2_{scatter}$, are then represented by the diagonal of the Fisher information matrix.

The computation of the Fisher matrix was then used to define the estimated variance at a single monoenergetic (monoE) keV level given the covariance term along the anti-diagonal (6).

$$\sigma^2_{MonoE}(E) = \sigma^2_{photo}\cdot\mu_{photo}(E) + \sigma^2_{scatter}\cdot\mu_{scatter}(E) + 2\cdot Covariance_{photo,scatter} \quad (6)$$

The approximate SNR at each monoE level could then be defined as the inverse of the square root of the variance, $(\sigma^2_{MonoE}(E))^{-1/2}$. An overall noise quality value termed "area under mono-energy curve", or AUMC, was set equivalent to the sum of these monoE SNR estimates from 35 to 120 keV (**Fig. 2**). This AUMC value represented the estimated total SNR across the photon energy spectrum.

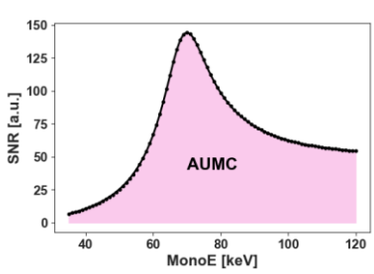


Fig. 2. Plot depicting monoenergetic SNR values and the AUMC integral area for an arbitrary kVp-switching system.

Three K-edge materials were selected as potential candidates: holmium (Ho), erbium (Er), and ytterbium (Yb). First, we used the CRLB to estimate the ideal K-edge material filter thicknesses (0.01-0.5 mm) for the central ray where no Teflon filtration was assumed. We located the optimum amount of each K-edge filter with a Scipy minimization function on the negation of AUMC for the thickness and material pair. This yielded the starting thickness of the K-edge filter for the designed double bowtie with the highest obtainable AUMC.

To account for the decrease in photon fluence through the K-edge filter, an iterative process was used to create adjustments in the duty cycle ratio from 0.1 up to 0.6 and increase of tube current to a maximum tube current of 571 mA as necessary. The reference dose computed with the conventional Teflon bowtie served as the target dose. This enabled realistic simulation scan parameters with the K-edge filter material in place in order to match the Teflon conventional bowtie reference dose.

Finally, for each path length outside of the central ray, an optimization procedure was implemented to determine the amount of K-edge filter material and Teflon material for a given

path length, *x*, so that the combination was caused equivalent *Dose(x)* reference value and to minimize the computed negation of the AUMC for that ray. The adjusted tube parameters were held constant over the path lengths using the duty cycle and tube current found for the central ray. We then had optimal filter thicknesses for the double bowtie which maximized spectral SNR across the photon energy spectrum with identical dose profiles to the reference profile.

III. Results

A. *Teflon filter thickness for homogeneous detector SNR*

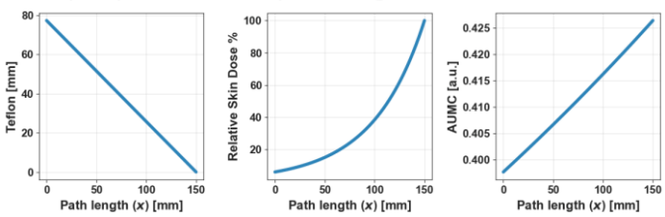


Fig. 3. *From left to right:* the designed conventional Teflon bowtie filter to generate a homogeneous detector SNR at each path length *x*, the skin dose per sr relative to the dose at the central ray, and the computed AUMC values for the conventional Teflon-only bowtie filter.

For the patient thickness, or path length *x*, approaching 0 mm, we estimate a maximum Teflon filter thickness of 77.3 mm (**Fig. 3**). Skin dose per sr peaks for the central ray, where no Teflon attenuates the incident x-ray beams, and declines to 6.0% of the maximum dose near the edges of the simulated phantom. The AUMC for the conventional bowtie filter is linear and ranges from 0.40 to 0.43 with increasing noise as path length increases.

B. *K-edge filter AUMC maximization at center ray*

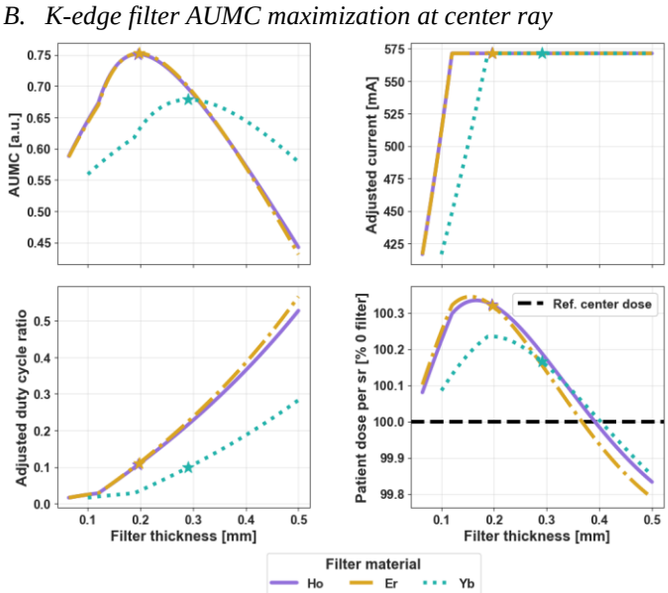


Fig. 4. *Top left:* For the three tested K-edge filter materials, the AUMC given varying filter thicknesses assuming no Teflon filtration at the center ray. The highest AUMC obtainable denoted by a star. *Top right:* The adjusted tube current to ensure necessary photon fluence through the filter. *Bottom left:* Duty cycle ratio adjusted for increasing filter thickness. *Bottom right:* Patient dose per sr for each filter thickness with adjusted tube parameters relative to the central reference dose (black dashed line).

Fig. 4 shows the results at the central x-ray where no assumed Teflon filtration occurs, and the longest path length (150 mm) is encountered. Between the three materials, holmium and erbium perform similarly by obtaining AUMC maxima at approximately 0.75 for 0.20 mm filter thickness. Ytterbium using 0.3 mm thickness reaches a maximum AUMC of 0.68. These AUMC values correspond to improvement of 1.76x and 1.59x SNR increase relative to the estimated AUMC in the conventional bowtie. Regardless of filter material, the adjusted tube current is maximized at 571 mA for the ideal center K-edge filter and the adjusted duty cycle ratio is set at approximately 0.1 or 10/90. The peak AUMC filter thicknesses in conjunction with the updated scan parameters produce reference patient dose levels per sr within 0.4% of the target reference dose.

C. *Double bowtie design to maximize AUMC*

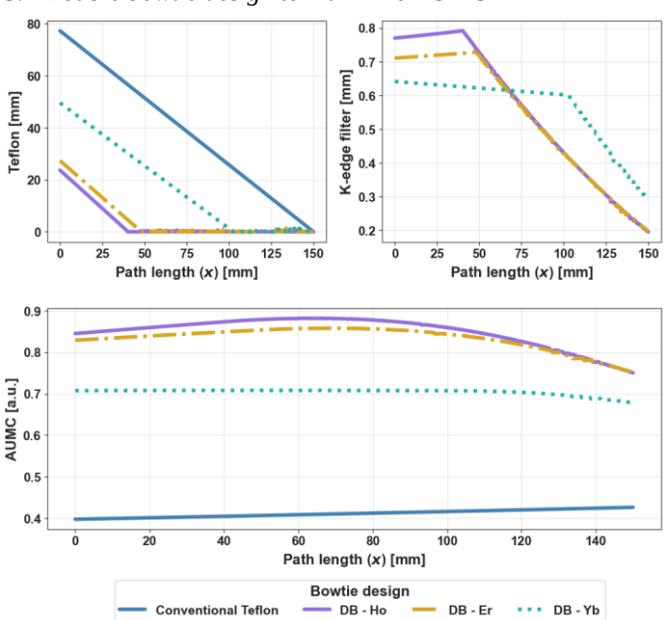


Fig. 5. *Top left:* Teflon filter of the conventional bowtie design compared to the **Teflon component of the double bowtie** (DB) with K-edge material configuration. *Top right:* K-edge filter thickness of the double bowtie along the path length *x* where AUMC is maximized, and patient dose is matched to the reference value. *Bottom:* The estimated AUMC along the bowtie and patient lengths for the optimal bowtie configurations.

Fig. 5 contains the double bowtie filter design compared to the conventional Teflon-only bowtie (solid blue line). For all tested K-edge filter materials, the total Teflon thickness in the double bowtie is less than the original reference conventional bowtie design. For ytterbium double bowtie, the optimal Teflon thickness goes from 50 mm to 0 mm, staying at 0 mm for path lengths greater than 100 mm. In comparison, the holmium and erbium double bowtie optimizations result in an optimal Teflon thickness of around 23 mm and go to 0 mm for path lengths greater than 50 mm. Before these 0 mm Teflon points, the K-edge material and Teflon provide the necessary attenuation to match the reference patient dose profile together. Generally, the optimal K-edge thickness is held constant in this range as Teflon thickness decreases. After the 0 mm Teflon point, the K-edge filter provides the necessary attenuation alone. The maximum K-edge filter thickness is 0.8, 0.7, and 0.65 mm for holmium, erbium, and ytterbium respectively.

The AUMC along every path length *x* is greater in the double bowtie construction for each K-edge filter compared to the

AUMC estimated with the conventional bowtie. Holmium produces the highest AUMC value of 0.88, or approximately 2x the same AUMC value for a water path length of 64 mm. All three of the double bowtie filters decline slightly in AUMC values as path lengths approach the maximum 150 mm. On average, across all tested path lengths, holmium has an increased AUMC SNR measurement of 2.06x ± 0.11 of the conventional bowtie, erbium is slightly less at 2.03x ± 0.10 AUMC, and ytterbium is 1.71x ± 0.05 the average AUMC for every tested phantom thickness.

For every path length *x*, each double bowtie matches the reference patient dose within 0.1%. This requirement ensures that no additional dose is given to the patient as the tube current and duty cycle were updated to accommodate the additional filter material attenuation.

IV. Discussion

This study introduced a task-specific approach for adapting spectral CT technology to pediatric patients. Building on the single K-edge filter design, we aim to tailor this technology for pediatric imaging. This clinical domain typically involves fewer spectral CT scans, and the evidence for quantitative imaging's diagnostic benefits in children is less established compared to adults. At the same time, it is absolutely essential to provide the same level of imaging to younger patients. Our research demonstrates that using a double bowtie filter composed of holmium and Teflon, optimized through a whole energy spectrum analysis, could enhance spectral signal-to-noise ratio (SNR) in smaller patients while adhering to CT tube power limitations.

However, this study has limitations, including the use of pencil-beam geometry and the absence of kVp-switching transition periods modeling in the dual input spectra. Future work will involve using a fan beam spectrum and refining transition time estimates based on tube current and voltage settings. Additionally, we plan to conduct experimental validations using a CT benchtop with a rapid kVp-switching x-ray tube, focusing on the proposed double bowtie's effectiveness. Exploring various tube voltage pairs and their impact on reducing radiation exposure will also be a key aspect of our ongoing research.

V. Conclusion

Our introduction of a double bowtie design, tailored for specific tasks, facilitates its integration into current CT technologies, eliminating the necessity to create a dedicated pediatric CT gantry. Subsequent clinical assessments will illuminate the design's efficacy across various clinical applications.

VI. References

[1] R. V. Gottumukkala, M. K. Kalra, A. Tabari, A. Otrakji, and M. S. Gee, "Advanced CT Techniques for Decreasing Radiation Dose, Reducing Sedation Requirements, and Optimizing Image Quality in Children," *Radiographics*, vol. 39, no. 3, pp. 709-726, May-Jun 2019, doi: 10.1148/rg.2019180082.

[2] C. H. McCollough *et al.*, "Principles and applications of multienergy CT: Report of AAPM Task Group 291," *Medical Physics*, vol. 47, no. 7, 2020-07-01 2020, doi: 10.1002/mp.14157.

[3] J. B. Rapp, D. M. Biko, A. M. White, K. I. Ramirez-Suarez, and H. J. Otero, "Spectral imaging in the pediatric chest: past, present and future," *Pediatric Radiology*, vol. 52, no. 10, pp. 1910-1920, 2022-09-01 2022, doi: 10.1007/s00247-022-05404-9.

[4] P. Rajiah, S. Abbara, and S. S. Halliburton, "Spectral detector CT for cardiovascular applications," *Diagn Interv Radiol*, vol. 23, no. 3, pp. 187-193, May-Jun 2017, doi: 10.5152/dir.2016.16255.

[5] Y. S. Kim, S. H. Kim, H. S. Ryu, and J. K. Han, "Iodine Quantification on Spectral Detector-Based Dual-Energy CT Enterography: Correlation with Crohn's Disease Activity Index and External Validation," *Korean Journal of Radiology*, vol. 19, no. 6, p. 1077, 2018-01-01 2018, doi: 10.3348/kjr.2018.19.6.1077.

[6] L. Yu, McCollough, C, "Optimal Tube Potential for Radiation Dose Reduction in Pediatric CT: Principles, Clinical Implementations, and Pitfalls," *RadioGraphics*, no. 31, pp. 835-848, 2011.

[7] A. N. Primak, J. C. Ramirez Giraldo, X. Liu, L. Yu, and C. H. McCollough, "Improved dual-energy material discrimination for dual-source CT by means of additional spectral filtration," *Med Phys*, vol. 36, no. 4, pp. 1359-69, Apr 2009, doi: 10.1118/1.3083567.

[8] M. Tivnan, W. Wang, and J. W. Stayman, "Multi-Contrast CT Imaging with a Prototype Spatial-Spectral Filter," *arXiv*, 15 Oct 2020, doi: arXiv:2010.07840v1.

[9] K. Kawamoto, H. Sato, and Y. Kogure, "Usefulness of Ag Additional Filter on Image Quality and Radiation Dose for Low-Dose Chest Computed Tomography," *Journal of Computer Assisted Tomography*, 2023, doi: 10.1097/rct.0000000000001538.

[10] O. F. Sandvold, Liu, L.P., Shapira, N., Perkins, A.E., Stayman, J.W., Gang, G.J., Proksa, R., Noel, P.B., "Combining spectral CT technologies to improve iodine quantification in pediatric imaging," presented at the Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, 2023.

Feasibility Study of a Rectangular CT Geometry for High Throughput Total-Body PET-CT

Boris Vervenne, Jens Maebe, Stefaan Vandenberghe
MEDISIP, Department of Electronics and Information Systems
Ghent University, Ghent, Belgium

***Abstract*—To increase the efficiency and reduce the cost of Total-Body PET imaging, our group is developing a new system based on flat-panel detectors in between which patients are scanned in a standing position. Its unconventional geometry however complicates the integration of a CT component. Novel carbon nanotube (CNT) sources may be used to create flexible CT geometries, in which alternating activation of sources obviates the need for source and detector rotation. We propose a rectangular CT geometry which obtains 3D tomographic images during a vertical translation movement. In this study, the effects of several geometrical factors on sinogram filling and image quality are studied in order to choose an optimal source-detector arrangement. A raytracing simulation package is developed, and a data processing chain is established to convert the rectangular measurements to sinogram data. Filtered backprojection (FBP) and iterative gradient descent (GD) reconstruction are used to study the image artifacts associated with the different types of gaps in the sinogram. Three geometric variations (varying source spacing, gaps at the source array edges, and only using the detector opposing the source) introduce different types of gaps in the sinogram, and artifacts in the image. GD is able to strongly reduce the artifacts present in the FBP image, but further improvements are required to achieve images of diagnostic quality.**

***Index Terms*—PET-CT, non-conventional geometries, CNT sources, limited angle tomography, simulation**

I. Introduction

A. Total-Body PET

Positron Emission Tomography (PET) is a common functional imaging study used for diagnosis, staging, and monitoring of disease in various clinical fields. However, the short axial detector coverage of current PET scanners limits their sensitivity for the injected tracer molecules, which negatively impacts the scan time, injected dose and/or signal to noise ratio, as many of the isotropically emitted gamma photons escape the system without being detected. Recently, Total-Body (TB)-PET systems with a significantly extended AFOV have been introduced to the commercial market. While these enable high-quality PET imaging at shorter scan times (<1 min) and reasonable dose levels, their cost poses barriers to adoption in a standard nuclear medicine department. Additionally, the throughput of TB-PET systems remains limited by the need to position patients on and off the scanner bed [1].

B. Walk-Through Total-Body PET

To eliminate the bottlenecks of current TB-PET, our group is developing an alternative concept based on two vertical flat detector panels (Fig. 1); patients simply walk into the so-called Walk-Through PET (WTPET) scanner and position themselves, after which a fast (30 second) image of the torso is obtained. No bed positioning is needed, reducing the scan time and staff involvement, and thus also the cost per scan. Further cost reductions are enabled by the use of affordable detector technology and the smaller required detector surface to cover the same solid angle. More details on the characteristics and expected (cost) benefits of the system can be found in [2]. A GATE Monte Carlo simulation study was published as well, in which the performance of the WTPET was shown to be competitive with a conventional, annular TB-PET system, at a much lower cost [3].

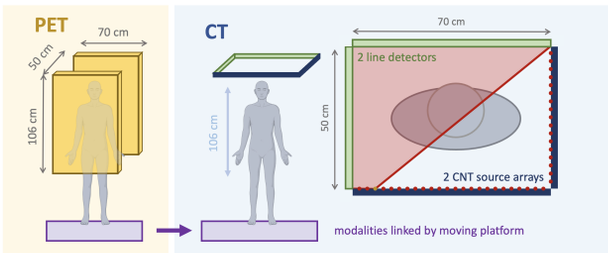


Fig. 1. Schematic of the proposed WTPET system, extended with the moving patient platform and rectangular CT prototype.

C. WTPET-CT integration

All recent commercial PET systems include a CT imaging component, since the functional images are commonly co-read with anatomical images for accurate localization of the activity. Additionally, the standard method to correct PET emission data for attenuation and scatter makes use of an attenuation map derived from a patient-specific CT image [4]. Alternative methods exist for attenuation correction, using iterative and DL-based reconstruction or transmission scans [5] [6] [7], but their lower reliability and convenience make them less desirable for use in clinical context than the standard CT method. Since there also is a strong preference among physicians to have a good quality anatomical image, the WTPET will be extended with a CT component. The unconventional geometry and upright patient position of the WTPET however impose several constraints that limit the applicability of conventional circular source-detector geometries using bulky X-ray sources. An alternative was found in a novel type of X-ray source based on carbon nanotube (CNT) technology which produces X-rays through field emission [8]. Multiple individually controllable

CNT sources can be arranged into flexible CT geometries; when sequentially turning the distributed sources on and off, projections are obtained at different angles around the object without physically moving the source and detector. A good fit with the PET component is achieved by combining an L-shaped source and detector array into a planar rectangular geometry (50x70 cm), which obtains 3D tomographic images slice-by-slice during a vertical translation movement. It is expected that most scattered photons will not hit the thin (6 mm) line detector, obviating the need for an anti-scatter grid. This assumption will be verified using Monte Carlo simulation. A downside of this non-conventional geometry is that the sampling of projections may be sparse (depending on the number of sources) and limited at certain angles [9]. Other studies have however experimentally demonstrated the ability of a similar rectangular geometry to produce tomographic images for baggage inspection purposes by making use of Total Variation (TV) regularized iterative reconstruction [10]. In this study, several variations of the proposed rectangular CT geometry are simulated and the resulting sinogram filling and image quality after reconstruction are evaluated. The obtained insights will serve as a basis to decide on the optimal design of the WTPET-CT system.

II. Materials and Methods

A. Rectangular CT simulation software

During the design phase of the PET component, Monte Carlo simulation was extensively used to investigate the effect of several geometrical parameters. Such simulations are however very computationally expensive for CT system prototyping due to the high photon counts in X-ray physics. As an alternative approach, deterministic methods based on raytracing and Beer's law of photon attenuation have shown to sufficiently mimic the global behavior of CT systems when disregarding the effects of noise and scatter [11]. Since existing raytracing simulation packages were found to be too rigid to model the unconventional rectangular geometry of the WTPET-CT and in order to allow seamless integration with reconstruction software, a new simulation package was developed in the high-performance Julia programming language [12]. Before starting a simulation, users define the positions of the CNT sources and the detector arrays (with a given pixel size). A voxelized phantom is defined as well, and the phantom used in this study is shown in Fig. 2. Note that all simulations are performed using mono-energetic sources of 80 keV.

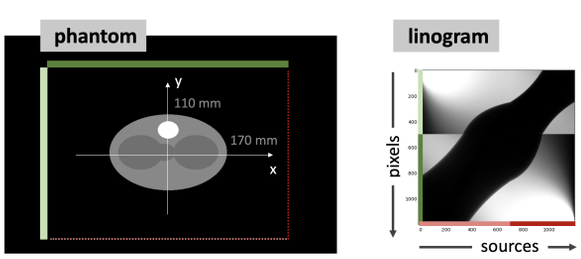


Fig. 2. Visualization of the phantom under study and its intensity linogram.

During the simulation, a user-defined number of rays is sampled by random selection of a source and direction in order to model uncertainty. For each simulated ray, raytracing is performed using a fast implementation of Siddon's algorithm [13], and the value of $I_0 \exp\left(-\sum_i \mu_i \cdot l_i\right)$ is calculated by summing over all voxels i that are crossed by the ray. Note that the rectangular geometry does not directly generate sinogram-style data; instead, the simulation output is a matrix that represents the measured intensity per source-detector pixel pair. This data matrix is referred to as a linogram due to the more linear appearance of features compared to the sinogram. Finally, note that the current version of the software still contains multiple simplifications: for instance, simulations are modelled as fully 2D and mono-energetic, and perfect quantum efficiency of the detectors is assumed. Implementation of these extra features is planned, but the current simple model suffices for the purpose of this study.

B. Considered geometrical parameters

The overall geometry measures 70 cm by 50 cm as shown in shown in Fig. 1. All parameters related to the detectors are kept constant for all simulations, with the pixel width set to 1 mm. First, an idealized geometry is studied with a source spacing of only 1 mm and no gaps at the edges in order to demonstrate the effects of the rectangular geometry without including the effects due to source sparsity. All simulated variations start from this idealized geometry, modifying only the mentioned parameters. To investigate the effect of source sparsity, systems where the sources are respectively 10 mm and 25 mm apart are simulated. In practice, a source spacing of 10 mm is possible according to our manufacturer, but increasing the spacing is desirable to reduce component cost. Secondly, the effect of gaps (due to housing etc.) between the source arrays is studied by adding a source-free margin of respectively 10 mm and 30 mm on both ends. Lastly, the angular range of the sources is altered: where all previous simulations modelled fan beams covering the full L-shaped detector surface, sources in this geometry are collimated to only cover the opposite detector. When scatter towards the perpendicular detector is low enough, this would allow the horizontal and vertical sources to operate in parallel, increasing the acquisition speed.

C. Linogram to sinogram

Conventional reconstruction techniques cannot readily be applied on the simulation output, as the linogram data needs to be converted to (parallel-beam) sinogram data first. The general processing pipeline is illustrated in Figure 3. Before rebinning, the user needs to define the desired parallel-beam detector configuration and set of angles. In this study, 1 mm wide detector pixels ranging from -450 to 450 mm, and 1000 angles in a 360 degree interval were used. During rebinning, all linogram intensity values are added to the corresponding position in sinogram space, and normalization is applied to account for the number of measurements per sinogram bin. An optional linear interpolation step may be applied on the

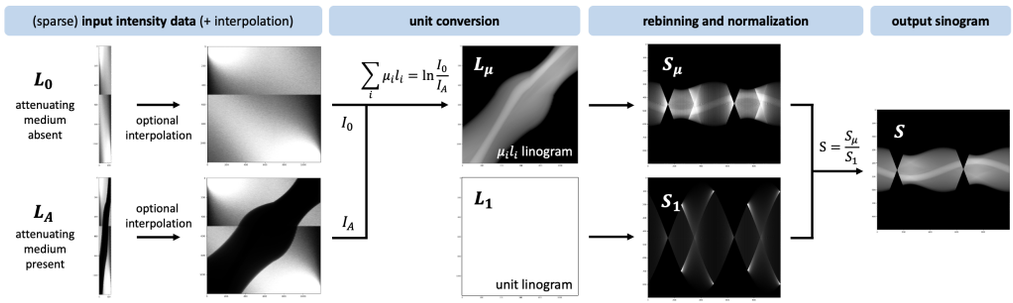


Fig. 3. Overview of the processing steps used to convert intensity data measured by the rectangular CT scanner to the corresponding sinogram.

input linogram data to account for the sparsity of the input data as discussed in section III-B.

D. Image reconstruction

Various types of gaps are observed in the resulting sinograms that lead to artifacts when using filtered backprojection (FBP). To remedy this, simple sinogram inpainting methods were tested such as the constrained Fourier space method [14]. However, these did not yield good results. Instead of filling the gaps, focus shifted to the use of iterative algorithms which are known to better handle sparse data. A standard gradient descent (GD) reconstruction algorithm with L2 regularization and a non-negativity constraint was implemented based on the following update equation, in which $\hat{f}$ represents the image estimate, k the iteration number, g the (sparse) input data, H the system matrix, and λ the regularization parameter.

$$\hat{f}^{k+1} = \hat{f}^k - 2H^t[H\hat{f}^k - g] - \lambda\hat{f}^k \geq 0 \qquad \text{(GD)}$$

All reconstructions were performed with a λ value of 0.001, using 8 subsets and a total number of 400 iterations. The next section shows the resulting images and corresponding line profiles. The latter allows analysis of the quantitative performance, which is important in the context of attenuation correction of PET data.

III. Results and Discussion

A. Idealized geometry

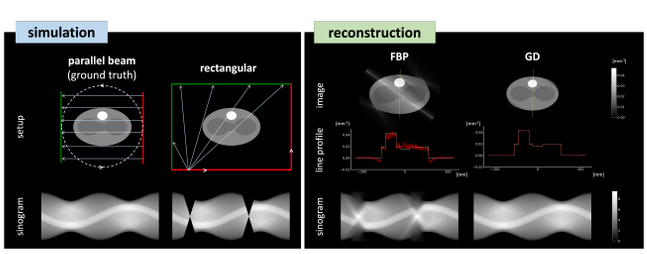


Fig. 4. The left side visualizes the ground truth and rectangular sinogram, which contains triangular gaps. The right side shows the FBP and GD reconstructions of the rectangular data, illustrating that GD leads to better image quality. This is also clear from the line profiles which are drawn along the yellow line. Forward projections of the reconstructions are shown to illustrate the gap filling.

Fig. 4 shows that two opposing triangular gaps are introduced in the sinogram due to the rectangular geometry. The fact that the rest of the sinogram corresponds well to the ground truth demonstrates that the simulation and data processing are implemented correctly. Many artifacts are observed in the FBP reconstruction: a bright diagonal line is superimposed on the image, the different areas of the phantom are inhomogeneous, and the ellipses seem to have sharp corners. GD performs much better, as illustrated by the absence of the line and an overall much smoother appearance of the image. However, the corners and shadowing effects are still present to a certain extent. This is reflected in the sinogram, where the gaps have been filled much more accurately by GD than FBP. For the remainder of this study, only GD is used due to its superior performance.

B. Effect of source spacing

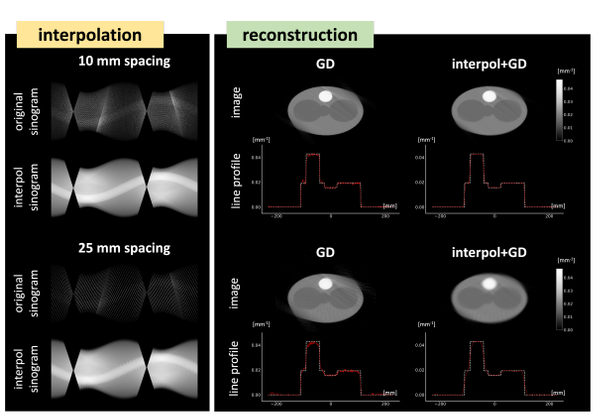


Fig. 5. The left side shows that increasing the spacing between sources makes the sinograms increasingly sparse, and that linear interpolation in the linogram space may be used to fill the gaps. The right side compares the image quality for different source spacings and interpolation uses.

Fig. 5 shows that an increased source spacing leads to gaps over the entire angular and position range, giving the sinogram a sparse appearance. Artifacts in the GD reconstructions increase as the distance between sources gets larger. Alternatively, linear interpolation may be performed between the columns in the linogram (see Fig. 3) to obtain data as if measured using 1 mm source spacing. This reduces

the sparsity in the sinogram, although a somewhat stepped behavior is introduced (especially for 25 mm spacing). In general, interpolation of the linogram data results in more homogeneous images, which however have more blurred-out edges compared to non-interpolated reconstructions.

C. Effect of gaps

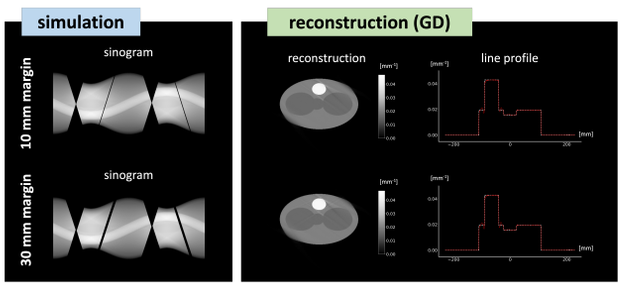


Fig. 6. Linear gaps are introduced when adding source-free margins on the sides of the source arrays. Their effect on image quality is shown on the right.

Introduction of dead space on both sides of the source arrays leads to diagonal line gaps with a thickness depending on the size of the gaps. Fig. 6 shows that GD handles these gaps well if they are narrow. For the 30 mm gap, shadowing effects along a second diagonal direction appear. Gaps on the sides of the detectors lead to similar effects, but with the direction of the line mirrored around the horizontal axis.

D. Effect of only using opposite detector

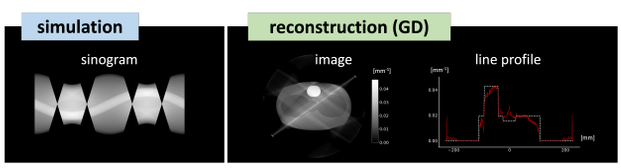


Fig. 7. The left side shows that additional gaps are introduced when only the opposite detector data is used, causing strong artifacts in the reconstructed image. Conventional GD is not sufficient to obtain good image quality.

When only the detector array opposing the source is used, two additional opposing triangular gaps are introduced. Fig 7 shows that the GD reconstruction is severely affected by artifacts, such as a bright diagonal line and shadowing effects. The outer ellipse now shows four corners.

E. Future work

For this feasibility study, several simplifications were made both in terms of simulation and reconstruction. Current work focuses on the development of more specialized reconstruction code that minimizes TV to further improve the quality of the images. Depending on its performance, also alternative (DL-based) sinogram inpainting methods might be explored to minimize artifacts in the images. Secondly, the simulation code will be extended to more accurately model a real-world system, such as including support for poly-energetic X-ray spectra and finite focal spot sizes. The improved simulations and reconstruction code will be used to decide on the geometry of the final prototype.

IV. Conclusion

In this work, it was shown that a rectangular geometry based on CNT sources is a promising method for extending the WTPET system with a CT component. The potential gaps in the sinogram due to relevant geometric variations were identified, and the associated artifacts in the images were studied using conventional FBP and GD reconstruction. It is expected that the use of more specialized algorithms will further reduce the presence of artifacts in the images, making it possible to obtain tomographic images of diagnostic quality.

Acknowledgment

This work was funded by BV's UGent BOF grant

References

[1] S. Vandenberghe, P. Moskal, and J. S. Karp, "State of the art in total body pet," *EJNMMI physics*, vol. 7, pp. 1–33, 2020.

[2] S. Vandenberghe, F. M. Muller, N. Withofs, M. Dadgar, J. Maebe, B. Vervenne, M. Abi Akl, S. Xue, K. Shi, G. Sportelli *et al.*, "Walk-through flat panel total-body pet: a patient-centered design for high throughput imaging at lower cost using doi-capable high-resolution monolithic detectors," *EJNMMI*, vol. 50, no. 12, pp. 3558–3571, 2023.

[3] M. Dadgar, J. Maebe, M. Abi Akl, B. Vervenne, and S. Vandenberghe, "A simulation study of the system characteristics for a long axial fov pet design based on monolithic bgo flat panels compared with a pixelated lso cylindrical design," *EJNMMI Physics*, vol. 10, no. 1, 2023.

[4] G. K. von Schulthess, H. C. Steinert, and T. F. Hany, "Integrated pet/ct: current applications and future directions," *Radiology*, vol. 238, no. 2, pp. 405–422, 2006.

[5] A. Rezaei, M. Defrise, G. Bal, C. Michel, M. Conti, C. Watson, and J. Nuyts, "Simultaneous reconstruction of activity and attenuation in time-of-flight pet," *IEEE transactions on medical imaging*, vol. 31, no. 12, pp. 2224–2233, 2012.

[6] J. S. Lee, "A review of deep-learning-based approaches for attenuation correction in positron emission tomography," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 2, pp. 160–184, 2020.

[7] S. R. Meikle, D. L. Bailey, P. K. Hooper, S. Eberl, B. F. Hutton, W. F. Jones, R. R. Fulton, and M. J. Fulham, "Simultaneous emission and transmission measurements for attenuation correction in whole-body pet," *JNM*, vol. 36, no. 9, pp. 1680–1688, 1995.

[8] G. Yue, Q. Qiu, B. Gao, Y. Cheng, J. Zhang, H. Shimoda, S. Chang, J. Lu, and O. Zhou, "Generation of continuous and pulsed diagnostic imaging x-ray radiation using a carbon-nanotube-based field-emission cathode," *Applied Physics Letters*, vol. 81, no. 2, pp. 355–357, 2002.

[9] E. M. Quan and D. S. Lalush, "Three-dimensional imaging properties of rotation-free square and hexagonal micro-ct systems," *IEEE transactions on medical imaging*, vol. 29, no. 3, pp. 916–923, 2010.

[10] B. Gonzales, D. Spronk, Y. Cheng, A. W. Tucker, M. Beckman, O. Zhou, and J. Lu, "Rectangular fixed-gantry ct prototype: combining cnt x-ray sources and accelerated compressed sensing-based reconstruction," *IEEE Access*, vol. 2, pp. 971–981, 2014.

[11] Q. Gong, R.-I. Stoian, D. S. Coccarelli, J. A. Greenberg, E. Vera, and M. E. Gehm, "Rapid simulation of x-ray transmission imaging for baggage inspection via gpu-based ray-tracing," *Nuclear Instruments and Methods in Physics*, vol. 415, pp. 100–109, 2018.

[12] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM review*, vol. 59, no. 1, pp. 65–98, 2017.

[13] H. Zhao and A. J. Reader, "Fast ray-tracing technique to calculate line integral paths in voxel arrays," in *2003 IEEE NSS*, vol. 4. IEEE, 2003, pp. 2808–2812.

[14] J. Karp, G. Muehllehner, and R. Lewitt, "Constrained fourier space method for compensation of missing data in emission computed tomography," *IEEE transactions on medical imaging*, vol. 7, no. 1, pp. 21–25, 1988.

Spectral Cone-Beam Artifact Reduction for Dual-Energy Cone-Beam CT

Zhilei Wang, Hewei Gao*,

***Abstract*—Flat-panel detector based cone-beam CT (CBCT) has been widely used in medical applications including image guidance for radiotherapy and interventional surgery. The FDK algorithm is an effective and the most popular reconstruction method for circular CBCT. Cone-beam artifacts associated with the FDK algorithm, however, will appear as low-intensity drop and/or streaks with increasing cone angle because the circular trajectory scan does not satisfy the data sufficiency condition. With the demand for improving the performance of circular CBCT, multiple spectral imaging technologies have been extended from diagnostic CT to CBCT. The material-specific images and a series of virtual monoenergetic images can be generated by material decomposition. Unfortunately, these images still suffer from cone-beam artifacts. In this work, we extend the well known two-pass cone-beam artifact reduction approach to spectral CBCT to correct for the cone-beam artifacts. we perform the two-pass approach to the two virtual monoenergetic images at two different energies firstly. And then basis material images can be obtained by matrix inversion and virtual monoenergetic image at any desired energy can be generated by linear combination of the two basis material images. Our experiment results demonstrate that our method can effectively correct the cone-beam artifacts in the spectral CBCT imaging.**

***Index Terms*—cone-beam CT, spectral imaging, material decomposition, cone-beam artifact reduction.**

I. Introduction

With the development of detector technologies, the large-area flat-panel detector based circular trajectory cone-beam CT (CBCT) has been widely used in medical applications including image guidance for radiotherapy and interventional surgery[1]. A 3D volume can be reconstructed from the cone-beam projections with single circle scan, as shown in Fig. 1. The FDK algorithm extended from filtered backprojection approach is an effective and the most popular method to reconstruct the 3D image from the cone-beam data[2]. However, the cone-beam artifacts associated with the FDK algorithm will appear as low-intensity drop and/or streaks with increasing cone angle because the circular trajectory does not satisfy the data sufficiency condition. There are several FDK-type methods to achieve cone-beam artifact reduction (CBAR) while maintain the efficiency of FDK[3, 4]. Another well known method without departing from the FDK algorithm is the so-called two-pass approach[5], which uses a second-pass forward projection and reconstruction to estimate cone-beam artifact and subsequently remove them from the original image.

With the demand for improving the performance of CBCT, the multiple spectral imaging technologies, such as dual source, fast-kV switching and dual-layer detector, have been extended from diagnostic CT to CBCT[6]. The spectral CBCT has many advantages compared with conventional imaging [6, 7]. For example, it can achieve bone removal, iodine quantification and virtual non-contrast imaging by basis material decomposition. In addition, a series of virtual monoenergetic images (VMIs) at different energies generated from linear combination of basis material images are very useful in many imaging tasks such as low-intensity contrast enhancement and metal artifact reduction. Unfortunately, the spectral CBCT technologies mentioned above based on the circular trajectory have no help in solving missing data problem. As a result, these images including basis material images and VMIs still suffer from cone-beam artifacts. In this work, we extend the two-pass approach to spectral CBCT to correct the cone-beam artifacts according to the correlation between the basis material images and the VMIs. Our experiment results demonstrate that our method can effectively correct the cone-beam artifacts in the spectral CBCT imaging.

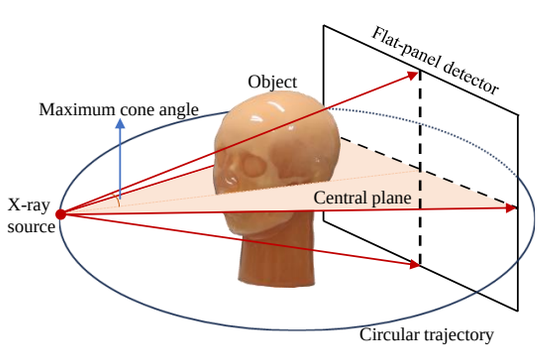


Fig. 1. A diagram of the circular trajectory CBCT geometry with a flat-panel detector .

II. Method

A. Spectral CBCT Imaging and Reconstrcution

The spectral CBCT imaging can be achieved by collecting low-energy (LE) and high-energy (HE) projection data of object at two spectrum with a moderate energy separation.

This work was supported in part by the National Natural Science Foundation of China (No.U20A20169 and No.12075130), and the National Key R&D Program of China (No.2022YFE0131100).

Zhilei Wang and Hewei Gao* are with the Key Laboratory of Particle and Radiation Imaging (Tsinghua University), Ministry of Education, China and Department of Engineering Physics, Tsinghua University, Beijing, 100084, China. (* indicates corresponding author: hwgao@tsinghua.edu.cn)

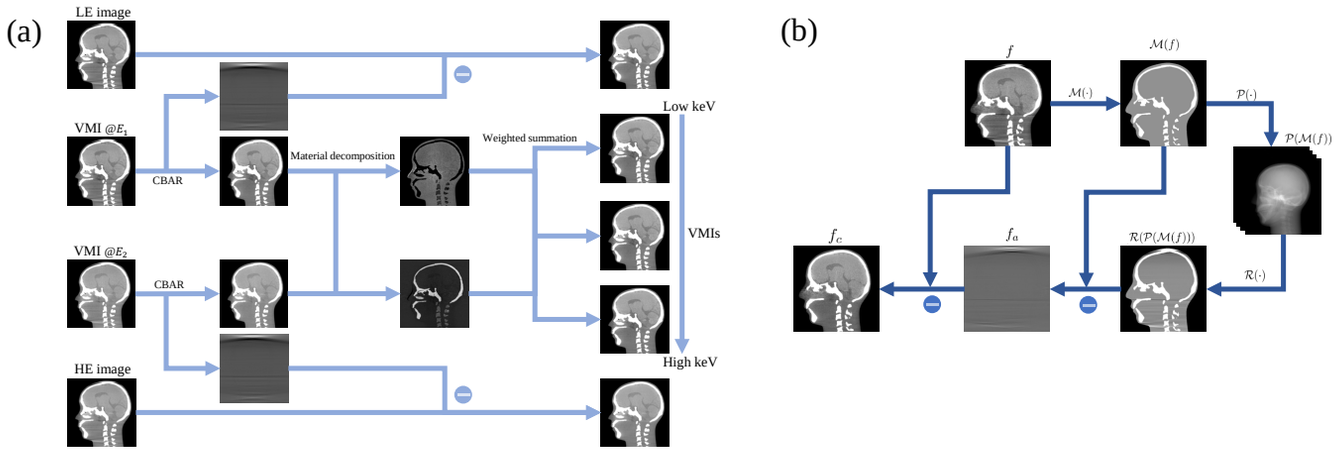


Fig. 2. (a) The workflow of the spectral cone-beam artifact reduction (CBAR) and (b) the processing steps of the CBAR for each VMI.

Mathematically, the dual-energy projections can be expressed as

$$\begin{aligned} P_{\mathrm{L}} &= -\ln\left(\int S_{\mathrm{L}}(E)e^{-\mu_1(E)\int\rho_1 \mathrm{d}l-\mu_2(E)\int\rho_2 \mathrm{d}l}\mathrm{dE}\right), \\ P_{\mathrm{H}} &= -\ln\left(\int S_{\mathrm{H}}(E)e^{-\mu_1(E)\int\rho_1 \mathrm{d}l-\mu_2(E)\int\rho_2 \mathrm{d}l}\mathrm{dE}\right). \end{aligned} \tag{1}$$

Here, P_{L} and P_{H} denote the low- and high-energy projections,respectively. $S_{\mathrm{L}}(E)$ and $S_{\mathrm{H}}(E)$ denote the normalized effective spectra at low and high energy, respectively. E denotes photon energy; μ_i and ρ_i denote the mass attenuation coefficient and density of basis material i;

Give that the energy dependence of different materials' attenuation is distinct, the line integral of density of the two selected basis materials can be solved from the dual-energy projections, from which the density map of the two material can be reconstructed. It is the so-called projection-domain material decomposition, which is easy to be achieved if the dual-energy projections are spatially aligned. The VMI can be generated from the linear combination of the basis material images following

$$\mu(E,r) = \mu_1(E)\rho_1(r) + \mu_2(E)\rho_2(r), \tag{2}$$

where, $\mu(E,r)$ denotes the linear attenuation sufficient of VMI at energy E and location r.

Scatter is ignored in Eq. (1), but it is a great challenge for quantitative imaging in CBCT. Therefore, scatter correction is necessary for accurate material decomposition. In this work, we use the CBCT Software Tools (CST 2.3, Varex Imaging) to estimate scatter by solving the linear Boltzmann transport equation following the proposed method by Maslowski et al [8]. In addition, to improve the robustness of material decomposition, we perform a hybrid material decomposition which is proposed in our previous work[9]. The FDK algorithm is used to reconstruct 3D images from the 2D cone-beam projections.

B. Spectral Cone-Beam Artifact Reduction

For the circular CBCT, the FDK algorithm is widely used in the medical and industrial applications. However, due to the missing data problem, the cone-beam artifacts will occur in the slices away from the central plane which could not be relieved by dual-energy scan. The basis material images and a set of VMIs will suffer from the artifacts, leading to reduced image quality. In this study, we tend to choose the well-known two-pass approach to correct the cone-beam artifacts. According to the correlation between the basis material images and the VMIs, it is not necessary to perform a correction process to each image because the two corrected images can refine the other images. For example, we could apply the two-pass approach to the two basis material images for once respectively, and the cone-beam artifacts in the VMIs synthesized following Eq. (2) will be corrected synchronously. Equivalently, we can also apply the approach to the two VMIs at two different energies and subsequently refine the basis material images and other VMIs. We choose the latter and take the effective energies of LE and HE spectrum as the selected energies so that the estimated cone-beam artifacts can also be used to correct the LE and HE images reconstructed from LE and HE projections, respectively. Fig. 2(a) shows the workflow of the spectral CBAR and the detailed processing steps of two-pass approach for each VMI are shown in Fig. 2(b). $\mathcal{M}(f)$ denotes a modeling operator including segmenting the first-pass reconstruction f into bone-tissue-air map image and replacing values in soft tissue and air by two constants to remove the most cone-beam artifacts in$f$. Subsequently, a forward projection $\mathcal{P}(\cdot)$ and a subsequent second-pass reconstruction $\mathcal{R}(\cdot)$ are conducted on $\mathcal{M}(f)$ to reproduce the cone-beam artifacts f_a.

$$f_a = \mathcal{R}(\mathcal{P}(\mathcal{M}(f))) - \mathcal{S}(\mathcal{M}(f)) \tag{3}$$

Here, a smoothing filtering operator $\mathcal{S}(\cdot)$ is usually applied in order to match the loss of spatial resolution from the second-pass reconstruction. The reproduced cone-beam artifact is considered similar to that of the initial volume f and can be subtracted from the image f to obtain cone-beam artifact reduced CT images (denoted as f_c).

$$f_c = f - \alpha f_a \tag{4}$$

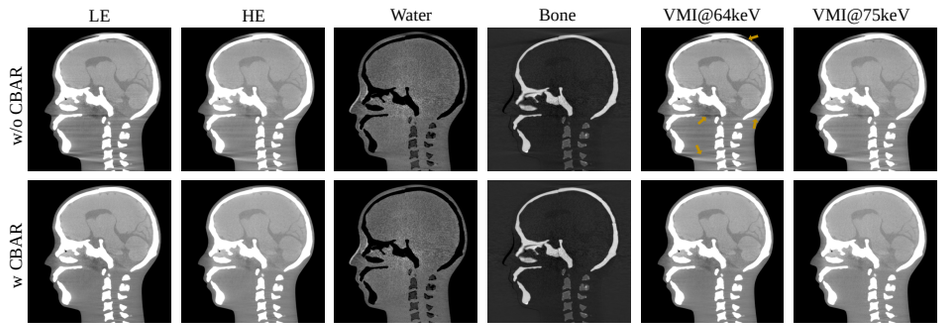


Fig. 3. The reconstructed images of LE and HE projections, decomposed basis materials and VMIs without and with the cone-beam artifact reduction. Display window: [500,2000] mg/cc for the water images; [-500,2500] mg/cc for the bone images; [-200,200] HU for the other images.

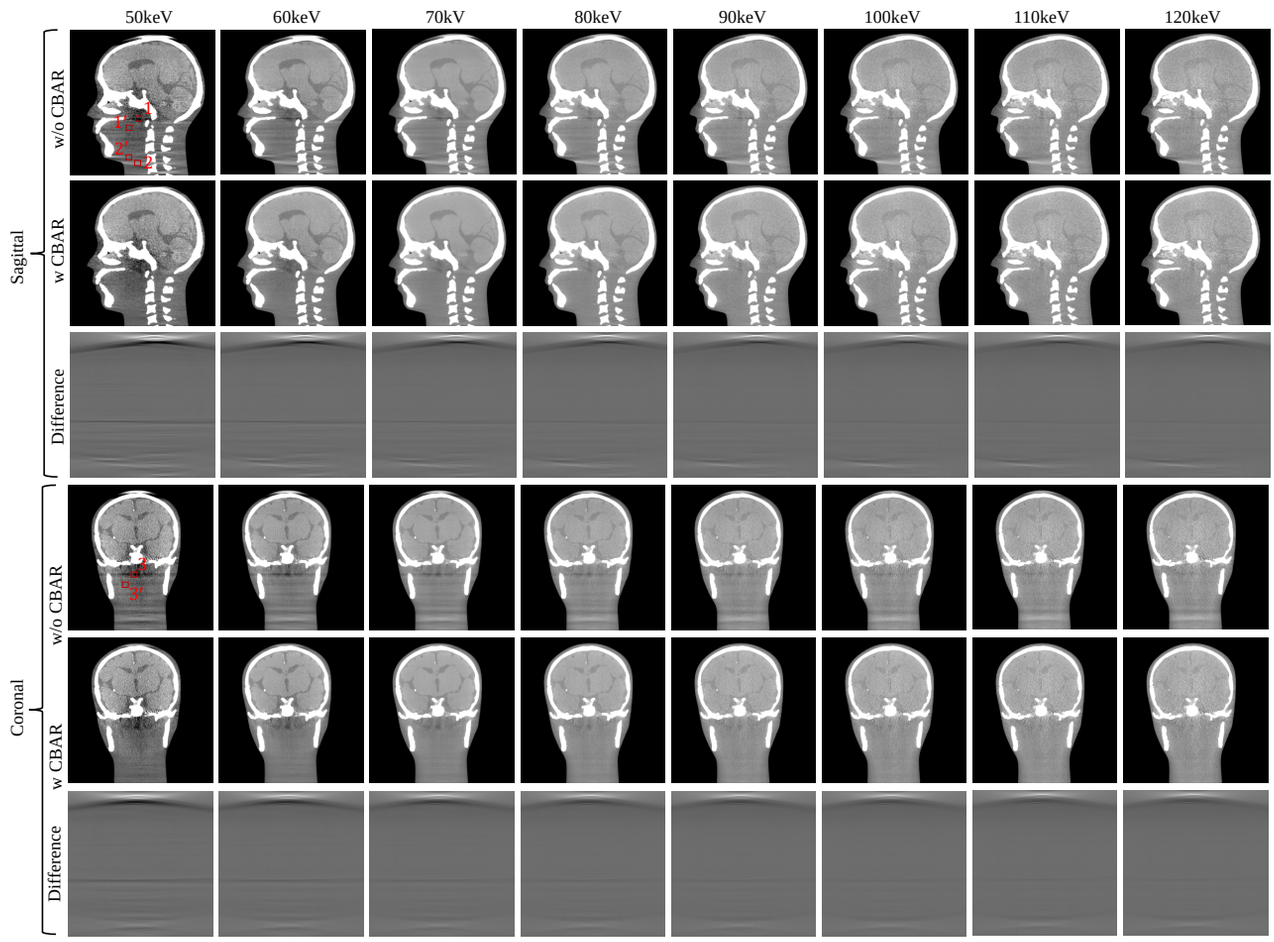


Fig. 4. The VMIs at different energies from 50 to 120 keV without and with the CBAR and the responding difference images. Display window: [-200,200] HU for VMIs; [-300,400] HU for difference images.

Here, α is a scaling factor that is close to 1 in practical applications.

To quantitatively evaluate the performance of the used spectral CBAR method, we select two adjacent ROIs in the homogeneous area and measure the difference of the means of the ROIs as the nonuniformity caused by the cone-beam artifacts. In the following analysis, we select multiple ROI pairs and measure the averaged nonuniformity to quantify the artifacts.

$$\Delta_{HU} = \frac{\sum_i^N |u_i - u_{i'}|}{N}. \tag{5}$$

Here, u_i denotes the mean value of the i_{th} ROI.

C. Experimental System Setup

To valid the proposed method for spectral cone-beam artifact correction, we acquired the dual-energy projections of a head phantom through two separate scans at two different kV settings based on a bench-top CBCT system. The low-energy data is acquired at 80 kVp with 0.1 mm copper as prefilter and the high-energy data is acquired at 120 kVp with 0.7 mm copper. The current and pulse width are 80mA and 16.6 ms for

low energy, 80 mA and 13.7 ms for high energy. The detector has a 1440 $\times$ 1440 matrix with a pixel size of 0.3 mm $\times$ 0.3 mm. 600 views per rotation over 360 degrees are collected for each energy setting. The two separate scans start at the same angle to confirm that the low- and high-energy projections are spatially aligned. The reconstructed image has a matrix size of 512 $\times$ 512 $\times$ 512 matrix size and a voxel size of 0.5 mm using the FDK algorithm.

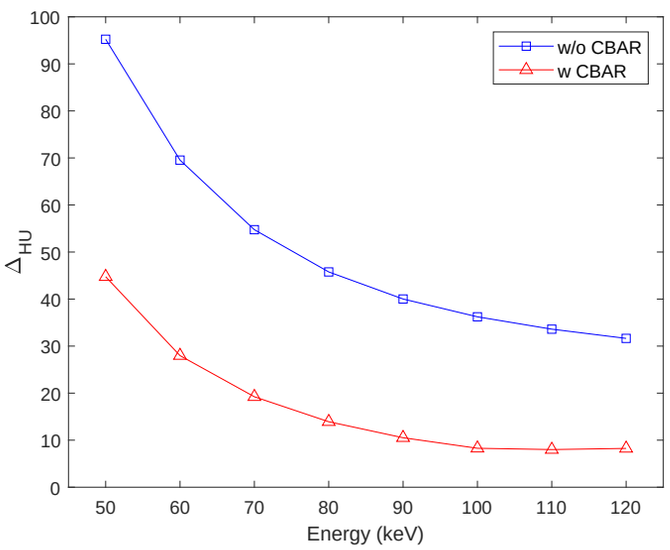


Fig. 5. The measured nonuniformity of VMIs at energies from 50 to 120 keV with and without CBAR. The selected ROIs for measurement are marked in Fig. 4.

III. RESULTS

Figure 3 shows the results of LE and HE reconstruction and material decomposition.The effective energies for LE and HE spectrum are measured as 64 keV and 75 keV, respectively. Therefore, the CBAR is applied to the VMIs at 64 and 75 keV, respectively. The yellow arrows indicate some of the locations of the noticeable cone-beam artifacts. The two corrected VMIs are used to refine the two basis material images. The results show that the streak artifacts in the basis material images could be greatly suppressed by CBAR. The LE and HE images are corrected by subtracting the estimated artifacts from the VMI at 64 keV and 75 keV, respectively. We can see that most of cone-beam artifacts in LE and HE images are removed. The comparison of VMIs at different energies without and with CBAR is shown in Fig. 4. It shows that the cone-beam artifacts in the VMIs at different energies could be all reduced obviously. The three selected ROI pairs are marked by red rectangle in Fig. 4 and the measured Δ_{HU} in VMIs from 50 to 120 keV is plotted in Fig. 5. Consistently reduced nonuniformity illustrates the effectiveness of our approach.

IV. CONCLUSION

Because spectral CBCT imaging dose not help in meeting the data sufficiency condition, the cone-beam artifacts still remain in the basis material images and VMIs obtained by dual-energy CBCT. In this work, we extend the two-pass CBAR approach to spectral CBCT to correct for the cone-beam artifacts. By applying the two-pass approach to the two VMIs at two distinct energies for once, respectively, the basis material images and VMIs at other desired energies can also be corrected correspondingly. Our experiment results on a head phantom demonstrate that our method can effectively correct the cone-beam artifacts in the spectral CBCT imaging. In the future work, we will aim at the scatter correction for the spectral CBCT.

REFERENCES

[1] Jeffrey H. Siewerdsen. Cone-beam CT with a flat-panel detector: From image science to image-guided surgery. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 648:S241–S250, 2011.

[2] Feldkamp L. A., L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *Medical Physics*, 1984.

[3] M Grass, Th Köhler, and R Proksa. 3d cone-beam CT reconstruction for circular trajectories. *Phys Med Biol*, 45(2):329, feb 2000.

[4] Xiangyang Tang, Jiang Hsieh, Akira Hagiwara, Roy A Nilsen, Jean-Baptiste Thibault, and Evgeny Drapkin. A three-dimensional weighted cone beam filtered backprojection (cb-fbp) algorithm for image reconstruction in volumetric ct under a circular source trajectory. *Phys Med Biol*, 50(16):3889, aug 2005.

[5] Jiang Hsieh. Two-pass algorithm for cone-beam reconstruction. In Kenneth M. Hanson, editor, *Medical Imaging 2000: Image Processing*, volume 3979, pages 533 – 540. International Society for Optics and Photonics, SPIE, 2000.

[6] Cynthia H. McCollough, Kirsten Boedeker, Dianna Cody, Xinhui Duan, Thomas Flohr, Sandra S. Halliburton, Jiang Hsieh, Rick R. Layman, and Norbert J. Pelc. Principles and applications of multienergy CT: Report of aapm task group 291. *Medical Physics*, 47(7):e881–e912, 2020.

[7] Cynthia H. McCollough, Shuai Leng, Lifeng Yu, and Joel G. Fletcher. Dual- and multi-energy CT: Principles, technical approaches, and clinical applications. *Radiology*, 276(3):637–653, 2015.

[8] Alexander Maslowski, Adam Wang, Mingshan Sun, Todd Wareing, Ian Davis, and Josh Star-Lack. Acuros cts: A fast, linear boltzmann transport equation solver for computed tomography scatter – part i: Core algorithms and validation. *Medical Physics*, 45(5):1899–1913, 2018.

[9] Zhilei Wang, Hao Zhou, Shan Gu, Yingxian Xia, Haiyue Liao, Yifan Deng, and Hewei 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.

Reconstruction Methods for Scans of X-ray Micro Tomography with a Rotation Stage Offset

Faguo Yang, Andriy Andreyev, and Matthew Andrew

***Abstract*—X-ray micro tomography systems provide high-resolution, non-destructive imaging of internal structures in samples. However the field of view (FoV) is limited, because of its high resolution. One way to enlarge the FoV is to offset the rotation stage to one side to collect one-side truncated projections. In literature there are many reconstruction methods for scans with detector offsets. In this paper, we study reconstruction methods for scans with rotation stage offsets.**

***Index Terms*—Image Reconstruction, Rotation Stage Offset, Detector Offset.**

I. Introduction

X-RAY micro tomography systems (XRM) provide high-resolution, non-destructive imaging of internal structures of a sample. The systems are utilized in a variety of industrial and research applications, such as mining, manufacturing, materials science, clinical research, and failure analysis. The systems provide the ability to visualize features in samples without the need to cut and slice the samples. The spatial resolution of high end XRMs could reach 0.5um, which results in a relatively small field of view. A very commonly used technique to extend the in-plane field of view is to offset detectors [1], [2], [3], [4]. However, for some X-ray micro tomography systems, it is difficult to shift the detector due to hardware constrains and easy to shift rotation stages. The differences between systems shifting the detectors and those shifting the rotation stages to one side can be seen from Fig. 1. For detector-shifted systems, the line connecting the source and rotation axis in the central plane still perpendicular to the detectors, which is not true for rotation stage shifted systems.

To enlarge FoV by shifting detectors, X-ray projections are usually truncated at one side of the detectors and cover a little bit more than half of the trans-axial field of view, so it requires a 360-degree full scan to acquire complete information to reconstruct the scanned samples. However, for the central overlap region, the angular coverage is 360 degree and for the non-overlap region, it is 180 degree. In [1], [5], pre-weighting the x-ray projections is deployed to reduce redundant information among the projections. In [5], there is also a method that uses line integrals from opposite projections to estimate missing line integrals of a projection when performing ramp filtering and then performs redundancy weighting on the ramp-filtered projections. Iterative reconstructions [2], [3] are also applied to detector-shifted scans, which are time-consuming.

In this paper, we put forward reconstruction methods for a rotation-stage shifted XRM system. With an introduction of a virtual detector concept, the reconstruction problem of a scan with rotation stage offset can be converted to a reconstruction problem of a scan with a detector-offset.

F. Yang, A. Andreyev and M. Andrew are with the Advanced Development and Design Department Carl Zeiss X-ray Microscopy, Inc, Dublin CA 94568.

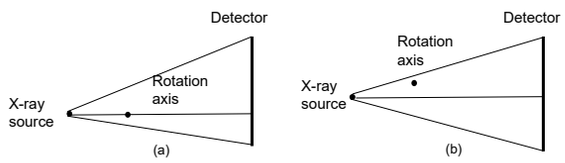


Fig. 1: Top view of system geometry with detectors or rotation stages shifted to one side. (a) Shifted detectors; (b) Shifted rotation stages.

II. Method

A. Geometry Representation

In an XRM system, usually the X-ray sources and detectors are kept motionless during a tomography scan, while the scanned sample is mounted on a rotation stage and rotates to collect a set of projections. When performing reconstruction, we can imagine that the sample is still, and the X-ray source and detector rotate relatively. With this equivalent scanned-sample-motionless system, the scan geometry of each projection can be described by four 3-element vectors (s_x, s_y, s_z), (d_{ox}, d_{oy}, d_{oz}), (d_{ux}, d_{uy}, d_{uz}) and (d_{vx}, d_{vy}, d_{vz}), which are the locations of the X-ray source, the center of the first detector pixel, detector column direction and detector row direction in a motion-less sample coordinates, respectively. We denote this geometry description as an SOUV geometry representation of a projection.

In reconstruction algorithms, a voxel-driven back-projection is often used [7], which requires computation of the location on the detector of a voxel in the reconstruction volume. The SOUV geometry representation of a projection can be converted to a 3 by 4 forward projection matrix M [6]. The mapping of a voxel with coordinate (x_v, y_v, z_v) in the reconstruction volume to a detector pixel denoted by (u, v) can be described by the following formula:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} p_u/w \\ p_v/w \end{bmatrix} \tag{1}$$

where p_u, p_v and w are computed by

$$\begin{bmatrix} p_u \\ p_v \\ w \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \begin{bmatrix} x_v \\ y_v \\ z_v \\ 1 \end{bmatrix} \tag{2}$$

If (s_x, s_y, s_z), (d_{ox}, d_{oy}, d_{oz}), and (x_v, y_v, z_v) are in same units (e.g. millimeter); the length of the detector orientation vectors (d_{ux}, d_{uy}, d_{uz}) and (d_{vx}, d_{vy}, d_{vz}) are the detector pixel size in the same unit as the source and detector origin (e. g. millimeter), then (u, v) is in unit of detector pixels.

B. Conversion to Detector Offset Geometry

In this section, we describe a method to convert the rotation-stage-offset system to a detector-offset system by definition of a virtual detector.

1) Virtual Detector Definition: As in Fig. 2, the source is denoted by S; the rotation axis is denoted by R; the cross point of the line connecting the source(S) to the rotation axis(R) (SR) and the real detector is denoted by C. The virtual detector passes through point C and perpendicular to the line SR. With the introduction of the virtual detector, the rotation stage offset system is converted to a detector offset system composed of the X-ray source (S), the Rotation axis(R) and the virtual detector.

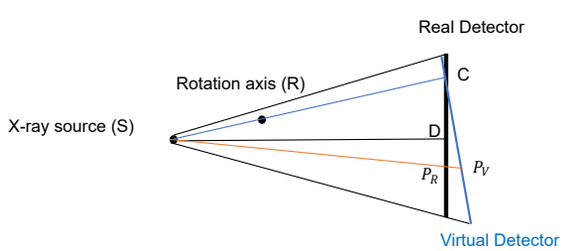


Fig. 2: Definition of a virtual detector in the top view of a rotation stage offset system. The real detector is in black and the virtual detector is in blue.

2) Computation of Pixel Values on Virtual Detector: From Fig. 2, we can see that the relationship between the real and virtual detectors is a rotation around point C. From the real detector representation (d_{ox}, d_{oy}, d_{oz}), (d_{ux}, d_{uy}, d_{uz}) and (d_{vx}, d_{vy}, d_{vz}), we can compute the virtual detector description, which are denoted by (i_{ox}, i_{oy}, i_{oz}), (i_{ux}, i_{uy}, i_{uz}) and (i_{vx}, i_{vy}, i_{vz}). For a pixel (α_u, α_v) on the virtual detector, as P_v shown in Fig. 2, its coordinates in the scanned-sample coordinate system can be computed by:

$$\begin{bmatrix} x_{(\alpha_u,\alpha_v)} \\ y_{(\alpha_u,\alpha_v)} \\ z_{(\alpha_u,\alpha_v)} \end{bmatrix} = \begin{bmatrix} i_{ox} \\ i_{oy} \\ i_{oz} \end{bmatrix} + \alpha_u \begin{bmatrix} i_{ux} \\ i_{uy} \\ i_{uz} \end{bmatrix} + \alpha_v \begin{bmatrix} i_{vx} \\ i_{vy} \\ i_{vz} \end{bmatrix} \tag{3}$$

After we know the coordinates in the sample coordinate system of the pixel (α_u, α_v) on the virtual detector, we can then use formula (1) to map it to the real detector, as P_R in Fig. 2. The value of a pixel (α_u, α_v) on the virtual detector can be computed by a bi-linear or other high order interpolation methods using neighboring pixel values of the mapped point on the real detector.

3) Maximum Rotation Axis Shift Value: With the introduction of the virtual detector concept, we can also determine the maximum rotation axis shift. From Fig. 3 (a), we can see that when we shift the rotation axis to the line connecting the source and the edge of the detector, we can still compute values for all pixels on half of the virtual detector from the real detector. Therefore there is enough raw data to perform a reconstruction for a 360-degree scan with a detector-shifted system using the method described in section II-C or II-D. However, as demonstrated in Fig. 3 (b), if we move the rotation axis beyond the line connecting the source and the edge of the detector, we can not find values on some part of the half virtual detector (e.g. P_V in Fig. 3 (b)). In practice, we usually do not move the rotation axis to the line connecting the source and the edge of the detector to give some overlap of the projections to make the system robust to small mechanical shifts during a tomography scan.

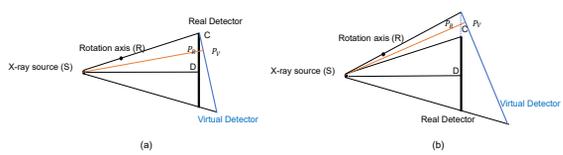


Fig. 3: Maximum rotation axis shifts. (a) The geometry with maximum rotation axis shift that still has enough raw data to perform a reconstruction; (b) The geometry with a rotation axis shift that does not have enough raw data to perform a reconstruction.

C. Analytical Reconstruction of Scans from Detector Offset Systems

In [1], [5], a step to pre-weight the raw data are added prior to the ramp-filtering step to properly deal with the redundant line integrals in a 360-degree full scan of a detector offset system in the FDK reconstruction workflow [9]. One weighting function form is as follows:

$$w(x) = \begin{cases} 1.0 + sin(\frac{\pi x}{4\sigma}) & 0 \le x \le 2\sigma \\ 2.0 & otherwise \end{cases} \tag{4}$$

where σ is defined as in Fig. 4.

The weighting function is a smooth function, which reduces truncation artifact in the reconstructed volume.

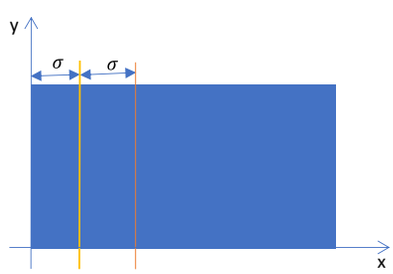


Fig. 4: A detector-offset scan system. X and Y axes are the detector column and row direction, respectively; The yellow line is the projection of the rotation axis onto the detector plane.

In our study, we find out that pre-weighting the raw data does not work for all our datasets, as shown in Fig. 10. So we adopt a different strategy: 1) Pad each row of the projection data using the first and last pixel values on the left and right, respectively in the ramp-filtering step to reduce truncation artifacts in the reconstruction volume [8]; 2) Perform

the weighting step using formula (4) after ramp-filtering the projection data, then perform a voxel-driven back-projection.

D. An Alternative Reconstruction Method without Performing Raw Projection Data Interpolation

In the above introduced reconstruction method, we have to perform interpolation when computing pixel values on the virtual detector, which could reduce the resolution of the reconstructed volume. From Fig. 2, we can see that the back-projection of P_V on the virtual detector and P_R on the real vector actually go through the same ray. To avoid interpolation of the raw data, we perform the following FDK [9] steps on the real detector instead of the virtual detector: 1) Cone-beam weighing; 2) Ramp filtering; 3) Redundancy weighting using formula 4, and 4) voxel-driven back-projection.

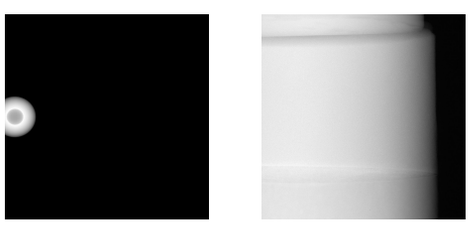
In the cone-beam weighing step, the weight should be the cosine value of the angle between the ray (SP_R in Fig. 2) and the line connecting X-ray source and rotation axis (SR in Fig. 2) instead of the angle between the ray and line connecting X-ray source and detector center (SD in Fig.2).

To make the intensity scale correct due to the principle ray (SR in Fig. 2) is not perpendicular to the real detector, we have to scale the final reconstruction volume by $cos(\alpha)$, where α is the angle between the real and virtual detectors.

III. Experimental Results

A. Datasets

1) Simulated Datasets: we use a hollow ball with an attenuation coefficient $\mu = 0.01mm^{-1}$ as the scan sample. The reason that we use a hollow ball is that the line integrals of the attenuation coefficient of each projection can be precisely computed out given the radius of the hollow ball and the scan geometry. The ideal scan geometry without rotation axis shift is defined as: the detector size is 1000 mm by 1000 mm with a pixel size 1mm. The line connecting the source and detector center is perpendicular to the detector plane and intersect the rotation axis. The rotation axis is parallel to the detector column direction. The X-ray source to the detector center and the rotation axis distance in the central plane is 1500mm and 750mm, respectively. To simulate a rotation axis shift system, we shift the rotation axis to the left of the detector by 225mm. One projection is shown in Fig. 5a



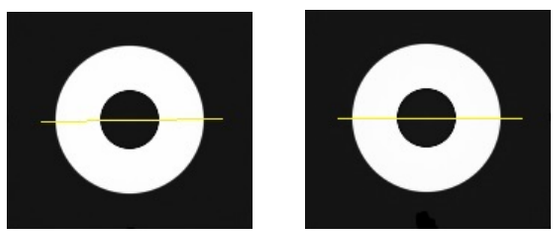
(a) (b)

Fig. 5: A projection of the simulated (a) and real datasets (b).

2) Real Datasets: The real datasets were acquired by a Zeiss Versa System. The scanned samples are rocks. One projection of a scan is shown in Fig. 5b.

B. Reconstruction Results of the Simulated Dataset

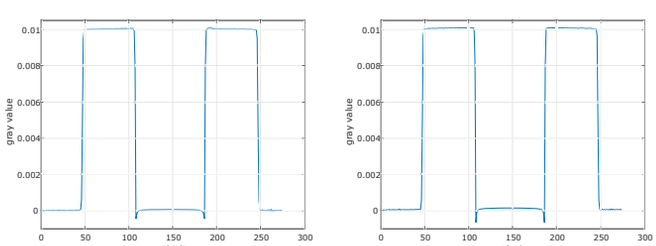
One slice of the reconstruction results of the simulated dataset is shown in Fig. 6. Fig. 6a is one slice of the reconstruction results by mapping the projection data from real detectors to virtual detectors, which is described in detail in sections II-B2 and II-C; Fig. 6b is the same slice as in Fig. 6a using the reconstruction method detailed in II-D. From Fig. 6a and Fig. 6b, we can not see significant image quality differences between the two reconstruction methods.



(a) (b)

Fig. 6: Reconstruction results of the simulated dataset. (a) Reconstruction results by mapping real detectors to virtual detectors; (b) Reconstruction results without mapping real detectors to virtual detectors.

Profiles corresponding to the yellow lines in Fig. 6a and Fig. 6b are drawn in Fig. 7a and Fig. 7b, respectively. As described in section III-A1, the simulated dataset has an attenuation coefficient of 0.01 and 0.0 mm^{-1} in foreground and background, respectively. Fig. 7 demonstrates that the reconstruction method that map real detectors to virtual detectors detailed in II-B2 recovers the attenuation coefficients better than performing the reconstruction in the real detector space described in section II-D, which may be because the ramp-filtering is more accurate in virtual detector space than that in real detector space.

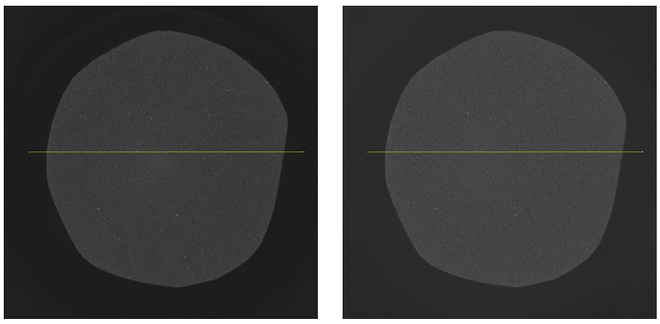


(a) (b)

Fig. 7: Profiles along the yellow lines in Fig.6a (a) and Fig.6b (b), respectively. The foreground value is closer to 0.01 mm^{-1} and the background value is closer to 0 mm^{-1} in (a) than in (b).

C. Reconstruction Results of the Real Dataset

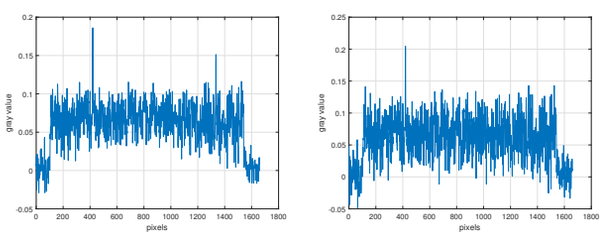
In Fig. 8, a slice of the reconstruction results of a real dataset with and without mapping real detectors to virtual detectors is shown in Fig. 8a and 8b, respectively. We can not see significant image quality differences from Fig. 8.



(a) (b)

Fig. 8: Reconstruction results of the real dataset. (a) Reconstruction results by mapping real detectors to virtual detectors; (b) Reconstruction results without mapping real detectors to virtual detectors.

Line profiles corresponding to those yellow lines in Fig. 8a and Fig. 8b are shown in Fig. 9a and Fig. 9b, respectively. Though we do not know the true attenuation coefficients of the real dataset, based on the results of the simulated dataset, the intensity value in Fig. 9a should be more accurate.



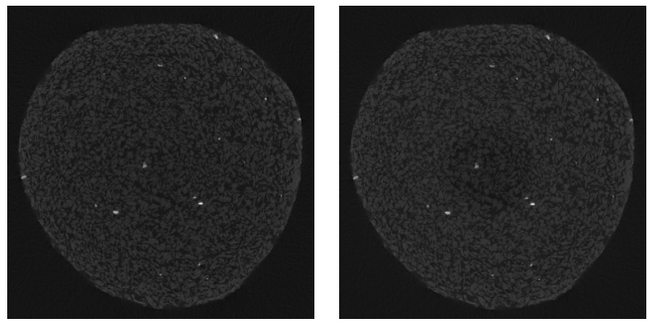
(a) (b)

Fig. 9: Profiles along the yellow lines in Fig.8a (a) and Fig.8b (b), respectively

Fig. 10 is the reconstruction results of another real dataset using the method described in section II-D. In Fig. 10a, we show the reconstruction results of the method performing redundant weighting after ramp-filtering step using formula 4; Fig. 10b shows the reconstruction results of the method pre-weighting the raw projection data. At the center of Fig.10b, we can see some 'shading' image artifact, which does not exist in Fig. 10a.

IV. Discussion

In this paper, we put forward two reconstruction methods for scans from a rotation-stage-offset XRM system. By introduction of a virtual detector concept, a rotation-stage-offset scan could be converted to a detector-offset scan, for which there are published reconstruction methods. For detector-offset scans, we also adapt the method to apply redundant line integral weighting after the ramp filtering step, which is critical for the algorithm to work for all our datasets. Bi-linear or other high order interpolations are involved when mapping raw projection pixel values onto the virtual detector, which may affect the resolution of the final reconstructed volume. So we also introduce another reconstruction method without



(a) (b)

Fig. 10: Reconstruction results of redundancy weighting on ramp-filtered projection data (a) and acquired projection data before any processing steps in the FDK reconstruction workflow (b), respectively.

performing mapping real detectors to virtual detectors, which reduces one interpolation step and an intensity scale correction step has to be deployed to make the attenuation coefficient scale correct. The reconstruction results of the simulated dataset demonstrate that the intensity scale is more accurate if we do the reconstruction using the virtual detectors, which may be due to that the ramp filtering step is more accurate in virtual detector space than real detector space. However, reconstructed images do not show significant differences for both simulated and real datasets.

References

[1] G. Wang, *X-ray micro-CT with a displaced detector array*, vol. 29, Med Phys, 2002.

[2] E. Hansis, J. Bredno, D. Sowards-Emmerd, L. Shao, *Iterative reconstruction for circular cone-beam CT with an offset flat-panel detector*, Nuclear Science Symposium Conference Record(NSS/MIC), IEEE, New York, 2010.

[3] K.S. Sharma, H. Gong, O. Ghasemalizadeh, H. Yu and G. Wang, *Interior micro-CT with an offset detector*, vol. 41, Med Phys, 2014.

[4] H.W. Tseng, A. Karellas, S. Vedantham, *Cone-beam breast CT using an offset detector: effect of detector offset and image reconstruction algorithm*, vol. 67, Phys Med Biol., 2022.

[5] P.S. Cho, A.D. Rudd, R.H. Johnson, *Cone-beam CT from width-truncated projections*, vol. 20, pp.49-57, Comput Med Imaging Graph, 1996.

[6] M. Kachelrieβ, M. Knaup, and O. Bockenbach, *Hyper-fast parallel-beam and cone-beam back-projection using the cell general purpose hardware*, vol. 34, Med Phys, 2007.

[7] Y. Du, G. Yu, X. Xiang, X. Wang, *GPU accelerated voxel-driven forward projection for iterative reconstruction of cone-beam CT*, vol. 16, BioMed Eng OnLine, 2017.

[8] M.M. Seger, *Ramp filter implementation on truncated projection data. Application to 3D linear tomography for logs*, Symposium on Image Analysis, Laund, Sweden, 2002.

[9] L.A. Feldkamp, L.C. Davis, J.W. Kress, *Practical Cone-beam Algorithm*, vol. 1, Journal of the Optical Society of America A, 1984.

Faster and More Realistic CT Projection Data Generation Using Digital Phantoms and CT Simulation Software

Jannik Dettmer, George S.K. Fung, W. Paul Segars and Karl Stierstorfer

Abstract*—This research investigates the speed improvement of CT simulations using 4D eXtended CArdiac-Torso (XCAT) phantoms in a triangle mesh format compared to Non-Uniform Rational B-Splines (NURBS). It also explores a way to make images from CT simulation more realistic by making objects inhomogeneous. CT simulations are done with the Deterministic RAdiological SIMulation (DRASIM) CT projector software (Siemens Healthineers) using realistic digital human phantoms (XCAT). The simulation times of phantoms defined by NURBS and triangular mesh at different resolutions are compared. By using triangular mesh objects, simulation times can be reduced to less than half without substantially losing object fidelity. To make objects inhomogeneous, a varying attenuation coefficient is defined mathematically. The attenuation integral for rays passing through an object with such areas is solved. With this, inhomogeneous looking materials and organic looking structures inside objects can be simulated using DRASIM.

I. Introduction

MUCH effort has been put into developing more realistic phantoms for the simulation of CT imaging data. With the development of phantoms such as the XCAT [1], realistic human anatomical structures defined by Non-Uniform Rational B-Spline (NURBS) [2] objects can be used for simulation purposes. Calculating line integrals through these NURBS-based phantoms, however, is generally computationally intensive. To improve the simulation speed, triangular mesh objects as an approximation of NURBS are explored instead.

In addition to improving simulation speed, it is desirable to generate more realistic looking images from the simulated data to better mimic real patients or human subjects. So far, each object of a phantom has been assigned a single material, making the attenuation (HU value) a constant inside each object. We propose a method to add mathematically defined but statistically distributed inhomogeneities to objects within phantom anatomies.

II. Methods

A. CT Projection Simulator

The CT projection simulator DRASIM [3] (Siemens Healthineers) is used to generate projection data from geometrically defined digital phantoms. Basically, DRASIM shoots a ray from the source to the detector and calculates the intersection points with the objects along its path. With this, the path length of the ray in each object can be obtained and the attenuation calculated. By default, DRASIM can only do intersection calculations on simple geometric objects. Previously in [4], we developed the functionality within DRASIM to enable direct simulation of the XCAT NURBS objects. In this work, the proposed speedup converts the NURBS objects into triangular meshes on the fly based on the specific requirements of the simulation. More efficient intersection calculations are then performed on the meshes.

B. Defining Surfaces with NURBS and Triangular Meshes

NURBS surfaces are defined by a set of control points and knot vectors with which all the points on the surface of the object can be calculated. The surface of such objects is perfectly smooth and can be zoomed into without losing resolution. To calculate the intersection point of a ray with the surface, the surface is subdivided into patches. Patches that are intersected by the ray are identified and tested for flatness. If not flat, a patch is subdivided into four sub-patches. Each sub-patch is then tested for intersection and flatness. The process iterates in this fashion until an intersecting sub-patch is found that meets the flatness criterion. The identified sub-patch is then treated as a plane and the intersection point with the ray is calculated. This is a computationally intensive process.

On the other hand, triangular mesh surfaces are defined by the coordinates of the vertices of the triangles comprising the surface mesh. Zooming in reveals sharp straight-line edges and corners. Intersection calculations on the other hand are much more efficient. Once the triangle that is intersected by the ray is identified, the intersection point of the ray with the triangle can be calculated immediately.

NURBS surfaces can be easily converted/approximated into triangle meshes. Users can define a resolution parameter to set the number of triangles the NURBS surface is approximated to.

C. Definition of Inhomogeneities

To create inhomogeneities, many inhomogeneous spherical objects can be placed within an organ or structure of the phantom. These can be hyperdense or hypodense. Each sphere has a minimum/maximum attenuation coefficient at the center and smoothly increasing/decreasing attenuation coefficient towards its surface, vanishing at the surface. When many of

J. Dettmer and K. Stierstorfer are with Siemens Healthineers, Forchheim, Germany (email: jannik.dettmer@siemens-healthineers.com).

G.S.K. Fung is with Siemens Healthineers, Baltimore, MD, USA

W.P. Segars is with the Department of Radiology, Duke University, Durham, NC, USA

these inhomogeneities overlap, organic looking structures are formed.

1) Definition of a Single Sphere / Inhomogeneity

Each inhomogeneity is defined by a polynomial function:

$$f(\vec{x},\ \vec{\lambda},\ R) = \begin{cases} \frac{\left(\vec{\lambda}-\vec{x}\right)^4}{R^4} - \frac{2\left(\vec{\lambda}-\vec{x}\right)^2}{R^2} + 1, & \left|\vec{\lambda}-\vec{x}\right| \le R \\ 0, & else \end{cases} \quad (1)$$

$$f(\vec{x},\ \vec{\lambda},\ R) \in [0;1]$$

R is the radius of the inhomogeneity. $\vec{\lambda}$ is its center. In Fig. 1, a plot of (1) showing a cross-section of an inhomogeneity is shown.

2) Attenuation Coefficient of an Object Containing Inhomogeneities

The attenuation coefficient of an object that contains one inhomogeneity at position $\vec{x}$ is defined as follows:

$$\mu(\vec{x}) = \mu_{material} \cdot \left(1 + \alpha \cdot f(\vec{x},\ \vec{\lambda},\ R)\right) \quad (2)$$

$\mu_{material}$ is the attenuation coefficient of the object, $\alpha \in \mathbb{R}$ is the parameter that changes the value range of f from $[0;1]$ to $[0;\alpha]$, thus changing how much the attenuation coefficient of the center of the inhomogeneity differs from $\mu_{material}$.

The attenuation coefficient of an object at a position $\vec{x}$ that contains many inhomogeneities is defined as follows:

$$\mu(\vec{x}) = \mu_{material} \cdot \left(1 + \sum_{i=1}^{N} \alpha_i \cdot f(\vec{x},\ \vec{\lambda_i},\ R_i)\right) \quad (3)$$

N is the number of inhomogeneities inside of the object. Each inhomogeneity can have its own parameters α_i, $\vec{\lambda_i}$ and R_i.

D. Attenuation of a Ray Passing an Inhomogeneity

Fig. 2 shows a ray passing through an inhomogeneity. $a + b$ is the length of the ray inside the inhomogeneity. In the case of Fig. 2 (a), $a = b = \sqrt{R^2 - d^2}$, where d is the distance of the ray

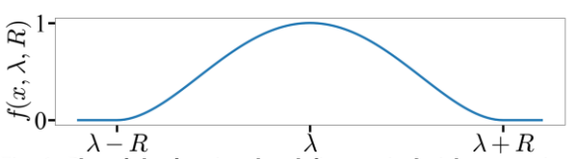


Fig. 1: Plot of the function that defines a single inhomogeneity in one dimension.

to the center of the inhomogeneity. In Fig. 2 (a), the inhomogeneity is entirely inside the object. The case where an inhomogeneity is clipped by the edge of an object is shown in Fig. 2 (b).

A point on the ray can be expressed as $\vec{x}(s) = \vec{0} + s \cdot \vec{v}$, $\vec{0}$ being the origin of the ray and $\vec{v}$ its normalized direction. The attenuation integral of a ray hitting one inhomogeneity can be written as:

$$\int_{m-a}^{m+b} \mu(\vec{x}(s))\ ds \quad (4)$$

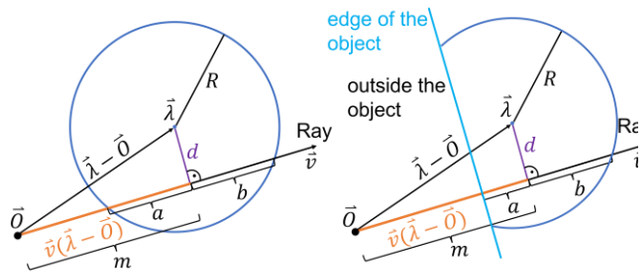


Fig. 2: A single inhomogeneity hit by a ray passing through it completely. $a = b$, (b) a single inhomogeneity that is cut off by the edge of the enclosing object and is hit by a ray. The ray does not pass through the entire inhomogeneity, $a \neq b$.

In the case of Fig. 2 (a), the integral can be simplified by setting $a = b = \sqrt{R^2 - d^2}$ such that it only depends on d and R.

E. Distribution of Inhomogeneities in an Object

For each inhomogeneity, its radius and amplitude (α) are randomly selected within the limits set by the user. The number of inhomogeneities (n) to be placed can be selected as well. By default, the desired number of inhomogeneities are randomly placed in the object's volume. Two other useful rules for distributing inhomogeneities have been developed:

1) Cluster Formation

First, a desired number of m inhomogeneities are equally distributed in the object, then the following $n - m$ inhomogeneities are placed such that they overlap with already existing ones. Thus the $n - m$ inhomogeneities "grow" around the m cluster seeds. Thus, organic looking structures can be "grown".

2) No Overlap

All inhomogeneities are placed inside the object avoiding overlapping with other inhomogeneities.

F. Image Reconstruction

Image reconstruction of noise-free simulated project data was done by using the offline reconstruction prototype software ReconCT (Siemens Healthineers). All images were reconstructed with a medium smooth Br40f kernel and a resolution of 512x512.

III. Results

A. NURBS vs. Triangle Mesh Performance

The following section is presenting results for simulation runtime and image quality. A simulation of an abdominal CT scan was performed using NURBS and triangle mesh representations of an XCAT phantom. Two different resolutions were used for the triangle mesh: the lower resolution used a number of triangles equal to twice the number of control points (liver: 3900), the higher resolution used twice as many triangles as the lower resolution.

All results are obtained from the simulation of a short spiral scan in the abdominal area of the phantom. The parameters of a clinical CT scanner (SOMATOM X.cite, Siemens Healthineers) were used. Simulation parameters included: 840 channels, 64 rows, 3x2 detector oversampling, total number of readings 4728, 1152 readings per rotation, pitch of 0.5.

1) NURBS

The simulation of the abdominal scan took 21 h 4 min. The contour of the chest surface in Fig. 3 is smooth, no straight-line edges are visible.

2) Triangle Mesh, Lower Resolution

The simulation of the abdominal scan took 7 h 59 min. The contour of the chest surface in Fig. 4 is not perfectly smooth. At first glance it looks very much the same as the NURBS version in Fig. 3, but when taking a close look some straight-line edges and corners are visible.

3) Triangle Mesh, Higher Resolution

The simulation of the abdominal scan took 12h 3min. The contour of the chest surface in Fig. 5 looks smooth. Even when taking a closer look, no straight-line edges or corners can be seen, at least not at a reconstruction resolution of 512x512.
Fig. 6 shows a difference image of the images in Fig. 3 and Fig. 5. Notable differences between the images are observed only at the edges of objects.

4) Overview

The simulation times for NURBS and triangle mesh at the two different resolutions are shown in Table I. By using triangle mesh at the lower resolution instead of NURBS, one can cut simulation time by a factor of 0.38 and still get decent object quality in the reconstructed images. Going to the higher resolution renders reconstructed images from NURBS and triangle mesh simulations almost indistinguishable while still cutting simulation time by a factor of 0.57 compared to NURBS. Increasing the number of triangles by factor of four does not increase simulation time by the same factor.

Some things to note are some artifacts in the reconstructed images near the object surface (especially the one with air). It is more prominent than typical patient CT images due to the homogenous objects, noise-free simulation and the kernel used. Also, there are some streak artifacts due to beam hardening on the bones (spines and ribs) which is normal.

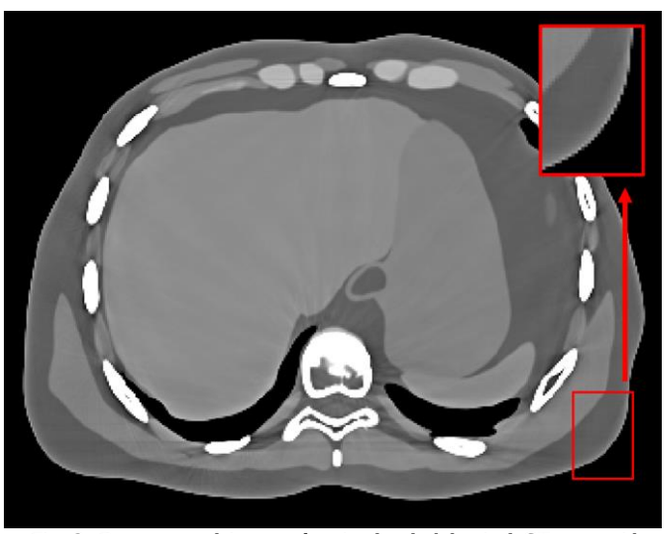


Fig. 3: Reconstructed image of a simulated abdominal CT scan with a NURBS XCAT phantom.

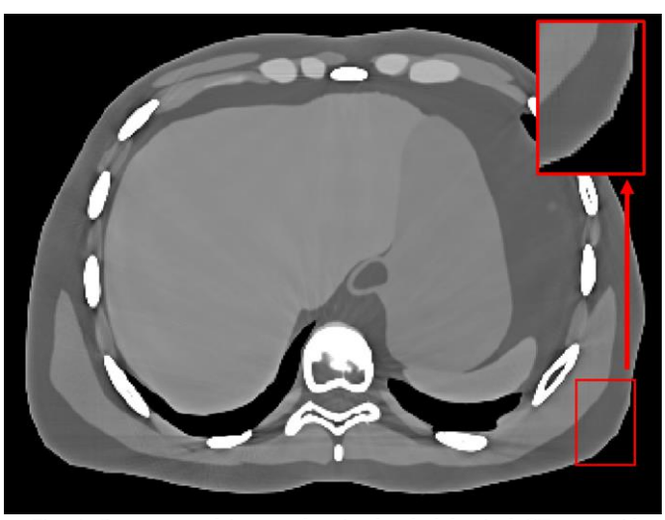


Fig. 4: Reconstructed image of a simulated abdominal CT scan with a triangle mesh XCAT phantom at the lower resolution.

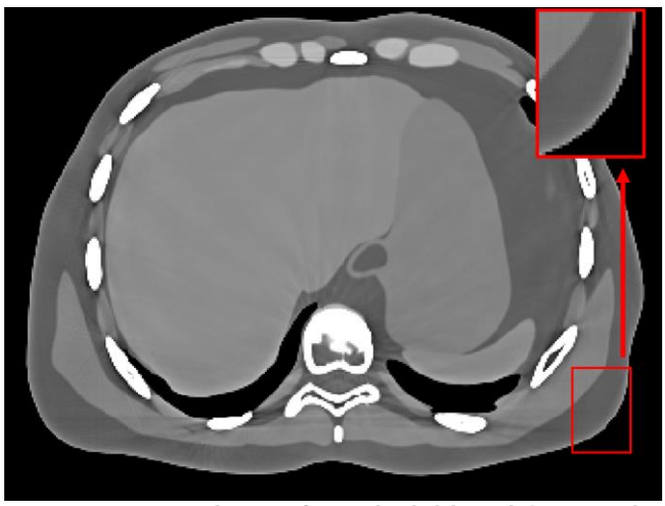


Fig. 5: Reconstructed image of a simulated abdominal CT scan with a triangle mesh XCAT phantom at the higher resolution.

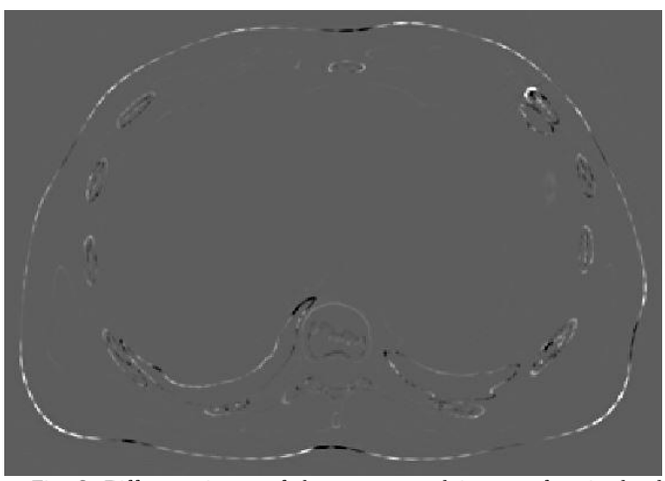


Fig. 6: Difference image of the reconstructed images of a simulated abdominal CT scan with a NURBS XCAT phantom and a triangle mesh XCAT phantom at the higher resolution (Fig. 3 – Fig. 5).

Phantom Type	Simulation Time
NURBS	21h 4min
Mesh, lower resolution	7h 59min
Mesh, higher resolution	12h 3min

Table I: Simulation times for an abdominal CT scan of the same XCAT phantom for three different phantom definition types.

B. Inhomogeneous Objects

Several reconstructed images from CT scan simulations of objects containing inhomogeneities produced with different settings are shown.

Fig. 7 shows reconstructed images of a spherical water phantom. The phantom in Fig. 7 (a) contains 1200 equally distributed hypo- and hyperdense inhomogeneities. Each inhomogeneity has a radius between 0.8 and 1.2 cm and α between -0.02 and 0.02. The phantom in Fig. 7 (b) contains 1200 non-overlapping inhomogeneities. Each inhomogeneity has a radius between 0.4 and 0.8 cm and α=0.01.

Fig. 8 shows a reconstructed image of an abdominal CT simulation scan. The liver object contains 500 hyperdense inhomogeneities clustered at 45 cluster seeds. Each inhomogeneity has a radius between 0.4 and 0.8 cm and α between 0.01 and 0.02.

Fig. 9 shows a reconstructed image of an abdominal CT simulation scan. The liver object contains 600 hypodense inhomogeneities clustered at 65 cluster seeds. Each inhomogeneity has a radius between 0.6 and 1.2 cm and α between -0.02 and -0.01.

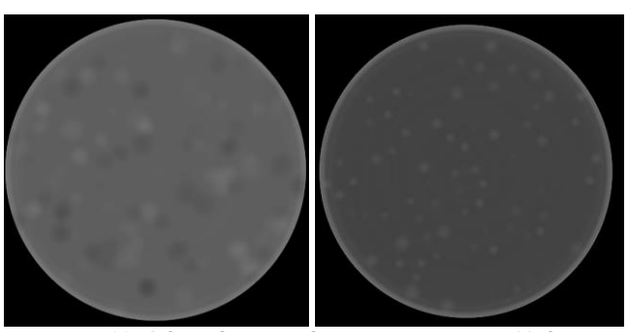


Fig. 7: (a) Spherical water phantom containing 1200 hypo- and hyperdense equally distributed inhomogeneities. (b) Spherical water phantom containing 1200 hyperdense non-overlapping inhomogeneities. (b) is windowed differently than (a).



Fig. 8: Image reconstructed from a simulated abdominal CT scan with 45 cluster seeds and a total of 500 hyperdense inhomogeneities in the liver.



Fig. 9: Image reconstructed from a simulated abdominal CT scan with 65 cluster seeds and a total of 600 hypodense inhomogeneities in the liver.

IV. CONCLUSION

Using triangle mesh instead of NURBS surfaces for CT simulations can substantially shorten simulation time. Even at the lower triangle mesh resolution, anatomical structures in reconstructed CT images appear realistic; corners and straight-line edges are only visible under close examination. Increasing the resolution makes NURBS and triangle mesh objects almost indistinguishable.

Random inhomogeneities added to otherwise piecewise homogeneous simulation phantoms create more realistic looking CT sections which could be used e.g., for model observer investigations. [5]

REFERENCES

[1] Segars, W. P., Mahesh, M., Beck, T.J., Frey, E. C., and Tsui, B. M. W., "Realistic CT simulation using the 4D XCAT phantom," Medical Physics, 35(8), 3800-3808 (2008).

[2] Cohen, E., et al., "Discrete B-splines and subdivision techniques in computer-aided geometric design and computer graphics," Comput. Grap. Image Process., 14, 87-111 (1980).

[3] Stierstorfer, K., "DRASIM: A CT-simulation tool," Internal Report, Siemens Healthcare.

[4] G.S.K. Fung, K. Stierstorfer, K. Taguchi, W.P. Segars, T.G. Flohr, B.M.W. Tsui, "A Realistic CT/Human-Model Simulation Package", Proc. 2nd Int. Conf. Image Format. X-ray Comput. Tomogr., pp. 66-69, 2012-Jun

[5] Kupinski, M. A., Clarkson, E., Hoppin, J. W., Chen, L., & Barrett, H. H., "Experimental determination of object statistics from noisy images" JOSA A, 20(3), 421-429, (2003)

Initial Results for Joint Reconstruction/ Decomposition of Spectral Cone-Beam Computed Tomography Data

David Leibold, Jacco van Leenen, Dennis R. Schaart and Marlies C. Goorden

***Abstract*—The use of cone-beam computed tomography (CBCT) for dose calculation in adaptive radiotherapy is intriguing but currently hindered by the insufficient image quality of CBCT. Spectral CBCT might offer a way to solve some of the shortcomings of CBCT. In order to optimize the amount of information that can be utilized from spectral CBCT, we implemented a statistical joint reconstruction/decomposition algorithm to directly acquire material maps and stopping power ratios from projection data. The algorithm was adapted from Mechlem *et al.* (2018 *IEEE T. Med. Imaging* 37 68–80) and Mory *et al.* (2018 *Phys. Med. Biol.* 63 235001) who applied it to fan-beam CT. It uses a polychromatic forward model and minimizes a cost function incorporating the negative Poisson likelihood and a Huber regularization term. Convergence was accelerated using Ordered Subsets and Nesterov Acceleration. We show preliminary results in a simulation study and outline the challenges ahead pertaining to CBCT data.**

***Index Terms*—Computed tomography, cone-beam CT, decomposition, joint reconstruction & decomposition, one-step algorithm, one-step inversion, reconstruction, radiography, X-ray imaging.**

I. Introduction

Cone-beam computed tomography (CBCT) based on flat-panel detectors is widely available for image guidance in radiotherapy. Currently, the volumetric information provided by CBCT is mostly restricted to checking patient alignment and positioning on the patient couch prior to a radiotherapy treatment session. The possibility to utilize this information for dose calculation in replanning for adaptive radiotherapy is intriguing, yet currently hindered by the inferior image quality of CBCT compared to either fan-beam CT or magnetic resonance imaging (MRI). As a consequence, radiotherapy treatment plans are commonly based on a dedicated scan acquired prior to the whole treatment series instead of utilizing the CBCT setup fitted inside the treatment gantry. However, the planning scan acquired prior to treatment might not reflect the patient's anatomy on the day of treatment.

The inferior image quality of CBCT compared to fan-beam CT is due to increased contribution of scatter (scatter that would not be detected in fan-beam CT ends up being detected by the flat panel detector), motion artifacts (longer scan times), beam hardening (longer path lengths due to cone-beam geometry) and intrinsic limitations of flat panel detectors (e.g., image lag). These effects impact both soft tissue contrast, important for (re-)delineation of contours, as well as Hounsfield unit (HU) accuracy, required for accurate dose calculation.

The implementation of adaptive radiotherapy based directly on CBCT does not only require strongly improved image quality, but also depends on the development of accelerated automatic software for auto-contouring and treatment plan calculation which are the topic of ongoing investigations [1], [2], [3]. Here we focus on improving the image quality of CBCT scans as the first step in the adaptive radiotherapy workflow and the basis for all subsequent steps.

Some of the shortcomings of CBCT may be resolved by applying the principle of spectrally-resolved imaging, which refers to the measurement of an object's attenuation at different energies. While single-energy CT allows only for the extraction of a single physical quantity, spectrally-resolved CBCT can be used to determine as many physical parameters as there are measurements at different energies (without using additional constraints). In its simplest form, images are acquired at two different energies which allows to extract two independent physical quantities from the data, e.g., the electron density and effective atomic number or the concentration of two different materials in each voxel in image space (material maps). The increased amount of information on material composition of tissues provided by spectrally-resolved CT may aid in delineation of anatomical structures and enable more accurate dose calculations [4], [5], [6]. Furthermore, the spectral information might help with the reduction of beam hardening artefacts.

Spectrally-resolved CBCT can be realized via, e.g., a voltage-switching source combined with an energy-integrating detector, a dual-source/dual-energy-integrating detector setup, or by combining a source at a fixed voltage with either a photon-counting detector (PCD) or a dual-layer detector [7].

D. Leibold is sponsored by Varian, A Siemens Healthineers Company, grant number 2018016.

D. Leibold, J. van Leenen and M. C. Goorden are with the Department of Radiation Science and Technology, Delft University of Technology, Mekelweg 15, 2629 JB Delft, The Netherlands.

D. R. Schaart is with the Department of Radiation Science and Technology, Delft University of Technology, Mekelweg 15, 2629 JB Delft, The Netherlands, and with HollandPTC, Huismansingel 4, 2629 JH Delft, The Netherlands.

Corresponding author: M. C. Goorden, e-mail: m.c.goorden@tudelft.nl.

Generally, the sinograms acquired at different energies in spectrally-resolved CBCT can be converted into material maps via three possible pathways: 1) The sinograms are independently reconstructed to create images (i.e., attenuation maps) at different energies, which are then decomposed to yield materials maps; 2) the sinograms are first decomposed into material sinograms, and these are then reconstructed to yield material maps in image space; or 3) the reconstruction and decomposition steps are performed simultaneously to directly convert sinograms into material maps.

In the case of the first two pathways both utilizing two-step approaches, any loss of information or introduction of artefacts in the first step cannot be recovered or corrected for in the second step. Since the last mentioned joint, one-step inversion approach performs reconstruction and decomposition simultaneously, it can incorporate all available information and is therefore expected to perform better than a two-step approach, as has been demonstrated for fan-beam CT [8].

In this work, we report initial results on the implementation of a joint reconstruction/decomposition algorithm for CBCT for extracting materials maps and stopping power ratios (SPRs). We built upon the work by Mory *et al.* [9], who compared five one-step reconstruction algorithms for spectral fan-beam CT. While all of the five algorithms compared by Mory *et al.* performed similarly in terms of results, i.e., the reconstructed concentration in the material maps and their visual appearance were comparable, they differed greatly in convergence speed. Of the five algorithms in the comparison, the statistical iterative algorithm by Mechlem *et al.* [10] converges one to three orders of magnitude faster than the other approaches, which can in parts be attributed to its use of Nesterov Acceleration and the use of Ordered Subsets. Since the reconstruction of CBCT data is even more computationally demanding than that of fan-beam CT data, we decided to implement the algorithm by Mechlem *et al.* for CBCT. Next to its importance in clinical practice, a shorter computation time has the additional advantage that it allows for a wider range of tuning of the regularization parameters.

The initial results presented here are based on the simulated acquisition of a small, simple phantom consisting of water and bone. Section II. outlines the simulated setup and the phantom as well as the core components of the algorithm, and Section III presents and discusses the results.

II. Methods

A. Adaption of work by Mory/Mechlem et al.

Mory *et al.* implemented five one-step reconstruction/ decomposition algorithms, including the one by Mechlem *et al.*, in Matlab and published the code open source on GitHub (https://github.com/SimonRit/OneStepSpectralCT). Our work is based on an adaptation of this code to a cone-beam geometry. The main components of the algorithm are detailed in the following; for a more extensive description we refer to the pseudo-code presented in Appendix A.4 in Mory *et al.*

B. Material Decomposition

It is assumed that the linear attenuation coefficient $\tilde{\mu}_{ve}$ in each voxel v at energy e can be expressed as a linear combination of N_m basis materials with linear attenuation coefficients μ_{me}:

$$\tilde{\mu}_{ve} = \sum_{m=1}^{N_m} \mu_{me} x_{vm} \quad , \tag{1}$$

with x_{vm} the volume fraction of material m in voxel v.

C. Forward Model

We use a polychromatic forward model that calculates the expected number of photons $\bar{y}_{pb}$ in each detector pixel p via

$$\bar{y}_{pb} = \sum_{e=1}^{N_e} s_{be} \exp\left(- \sum_{v=1}^{N_v} a_{pv} \sum_{m=1}^{N_m} \mu_{me} x_{vm} \right) \quad . \tag{2}$$

The expression in the exponent is the discretized line integral, that is, the sum over the linear attenuation coefficient in each voxel (see Eq. (1)) times the path length a_{pv} in that voxel. Furthermore, s_{be} specifies the effective spectrum of each energy bin of a PCD, which includes the X-ray tube spectrum, filtering and detector response, and therefore performs the binning of the photons into b energy bins. The binning of the number of photons is thus specific to the detector.

D. Cost Function

The cost function that is to be minimized consists of a data attachment and regularization term. The data attachment term $\mathcal{D}$ is the Poisson negative log-likelihood $\mathcal{D} = \sum_{b,p} \bar{y}_{bp} - y_{bp} \log \bar{y}_{bp}$ (excluding constant terms), where y_{bp} denotes the measured number of counts in energy bin b and detector pixel p. The regularization term $\mathcal{R}$ uses a Huber regularization:

$$\mathcal{R} = \sum_{m=1}^{N_m} w_m \sum_{v=1}^{N_v} \sum_{\tilde{v} \in \mathcal{N}} \phi^{\text{Huber}}(x_v - x_{\tilde{v}}, \delta_m) \quad , \tag{3}$$

with $\mathcal{N}$ being the set of neighborhood voxels $\tilde{v}$ around voxel v, w_m the regularization parameters, and the Huber function serving as a twice-differentiable approximation of the $\ell 1$-norm:

$$\phi^{\text{Huber}}(x, \delta) = \begin{cases} x^2 & \text{if } |x| < \delta, \\ 2\delta|x| - \delta^2 & \text{else.} \end{cases} \tag{4}$$

Hence, the Huber function is quadratic for differences smaller than the Huber parameter δ, and linear otherwise. This form of regularization function tries to suppress noise, e.g. small differences between neighboring pixels, while preserving large differences, e.g., large contrast due to different anatomical structures.

The total cost is then $\mathcal{C} = \mathcal{D} + \mathcal{R}$. Note that the parameters δ_m and w_m have to be optimized for each basis material.

E. Ordered Subsets

The set of all projection angles is divided into subsets such that the subsets are pairwise disjoint. Data in each subset is used as a sub-iteration of the algorithm, which consists of forward

projection, comparison with the measured data and calculation of the update of the estimated material maps. A cycle through all subsets then constitutes one iteration.

In our implementation we divide the number of projections into 4 subsets.

F. Nesterov Acceleration

Nesterov Acceleration [11] uses the momentum of previously calculated gradient steps during iteration in subsequent iteration steps. Along with the usage of Ordered Subsets, Mechlem *et al.* were able to show that this leads to a substantial improvement in convergence speed compared to not using these techniques (algorithm by Weidinger *et al.* [12]).

G. Measurement data

The measurement data, that is, the input data that the joint reconstruction/decomposition algorithm is tasked to process, is generated in a simulation and created by forward projecting the ground truth of the material maps and applying Poisson sampling to the expected number of photons. It is assumed that each detector pixel receives about $5.4 \cdot 10^4$ photons if no phantom is inserted.

H. Separable Quadratic Surrogates

While Mechlem *et al.* use Separable Quadratic Surrogates for the minimization of the cost function, Mory *et al.* did not apply this in their implementation of the algorithm (contrary to the statement in their paper), most likely since the Huber function is already a quadratic function for values smaller than the parameter δ.

I. Simulation setup and geometry

1) Phantom

The phantom consists of a square box made of water (mass density of 1 g/cm^3) with a box-shaped insert of cortical bone (material definition taken from White *et al.* [13], Table A1, "Skeleton-cortical bone, Adult"; mass density of 1.92 g/cm^3), since human tissues can be well represented by a combination of these two materials [14]. The phantom is voxelized and covers a volume of 32 × 32 × 32 voxels, with voxels of 1 × 1 × 1 mm^3 each.

We do not use a, e.g., two-step method for initializing the phantom during reconstruction, and instead initialize it with zeros.

2) Geometry

The geometry we simulate is as follows: The source-isocenter distance is set to 2300 mm, the source-detector distance to 3200 mm, and the detector has an extension of 64 × 64 pixels with a pixel pitch of 1 mm in both dimensions. The restriction to such a small detector, and, correspondingly, small phantom, is due to the large memory usage of our algorithm in its current implementation.

3) Source spectrum

While Mechlem *et al.* incorporate an estimation step for the effective spectrum (i.e., the product of the source spectrum and detector response), we assume perfect knowledge of it. Following Mory *et al.*, the source spectrum is taken from Ducros [15] which is based on a 120 kVp spectrum filtered by 1.2 mm Al. We used the spectrum definition included in the code published by Mory *et al.*

4) Detector

A photon-counting detector is assumed with five energy bins spanning 30–50 keV, 51–61 keV, 62–71 keV, 72–82 keV and 83–150 keV. We use the same bin edges as Mory *et al.*, however, we note that these could be optimized in future work. The detector response is taken from Mory *et al.*, who in turn base it on Schlomka *et al.* 2008 [16]. Except for the detector response, an ideal photon-counting detector is assumed, i.e., effects such as pixel cross-talk (charge sharing, X-ray scatter etc.), pulse pile-up, non-homogeneous pixel sensitivity etc. are not incorporated.

III. Results And Discussion

Figure 1 shows the ground truth material maps for the volume fractions of water and bone (Fig. 1a, and c) in the phantom. For reconstruction with our one-step algorithm, first the regularization parameters were optimized based on visual quality of the reconstructed material maps. We chose $w_m = 50$, 100 and $\delta = 0.05$, 0.01 (please refer to Eq. (3) and (4)) as the optimal regularization parameters for water and bone, respectively (Fig. 1b and d). The results were obtained after 10 iterations.

The mean volume fraction extracted from the reconstructed material maps is 0.98 ± 0.03 for the water and 0.939 ± 0.009 for the bone volume (both eroded by one voxel), which agrees very well with the ground truth.

Constructing an SPR map based on the volume fraction maps and comparing it to the ground truth yielded an average estimation error (following Zhang *et al.* [17], Eq. (23)) of 0.042.

We would like to note that the measured data y, which the reconstruction algorithm is tasked to reconstruct, is created by the same forward model that is used during reconstruction, the only difference being the application of Poisson sampling to the expected number of photons. Consequently, the forward model in our algorithm covers all aspects of the measured data, which is normally not the case in a realistic setting. Hence, this might lead to an overestimation of the performance of our algorithm.

IV. Conclusions and Outlook

In this work, we presented the implementation of a joint reconstruction/decomposition algorithm for CBCT applied to simulated data of a simple water/bone phantom, and we determined the accuracy of the material maps and SPR map.

Future work will concern improvements on the memory usage of the algorithm in order to make it applicable to realistic

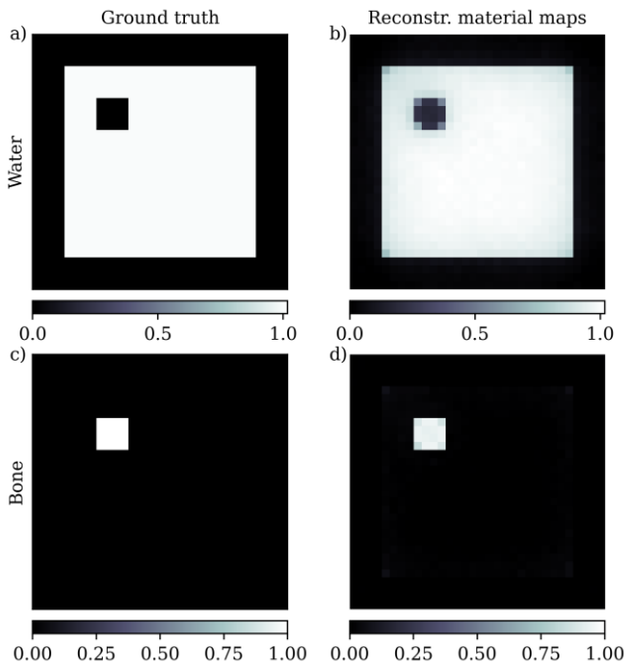


Fig. 1: Volume fraction maps for water and bone. The first column shows the ground truth volume fraction maps for water (a) and bone (c), the second column shows the reconstructed volume fraction maps obtained with our algorithm (b, d) after 10 iterations. The regularization parameters are $w_m = 50$, 100 and $\delta = 0.05$, 0.01 for water and bone, respectively. Please note that the value range for the color scale is specific for each material.

data, since the current implementation uses a large amount of memory for a small reconstructed volume (approximately 8 GB for 3 material maps containing 32×32×32 voxels each, 400 projections and a detector of 64×44 pixels).

Furthermore, we will compare the performance of our one-step algorithm to a two-step, image-based method (based on reconstruction of images for different energy bins followed by decomposition into material maps).

One of the main reasons for the inferior image quality of CBCT compared to fan-beam CT is the larger susceptibility of the detector to scatter. We will investigate the performance of our one-step algorithm on data that includes a realistic scatter model via a full Monte Carlo simulation, which the forward model of our one-step algorithm currently lacks.

Last but not least, Zhang *et al.* have demonstrated subpercentage accuracy for SPR prediction [17] whereas the error on our results is about one order of magnitude larger. This will be the subject of future investigations, as will be the numerical stability of Nesterov acceleration when facing real data [9] which may e.g. suffer from image truncation.

Acknowledgment

The authors would like to acknowledge the high-performance-computing clusters provided by the ICT department of TU Delft. The authors would also like to express their gratitude towards the authors of the paper by Mory *et al.* for making their code open-source.

References

[1] G. Heilemann *et al.*, "Clinical Implementation and Evaluation of Auto-Segmentation Tools for Multi-Site Contouring in Radiotherapy," *Physics and Imaging in Radiation Oncology*, vol. 28, p. 100515, Oct. 2023, doi: 10.1016/j.phro.2023.100515.

[2] M. Oud *et al.*, "An online adaptive plan library approach for intensity modulated proton therapy for head and neck cancer," *Radiotherapy and Oncology*, vol. 176, pp. 68–75, Nov. 2022, doi: 10.1016/j.radonc.2022.09.011.

[3] O. Pastor-Serrano and Z. Perkó, "Millisecond speed deep learning based proton dose calculation with Monte Carlo accuracy," *Phys. Med. Biol.*, vol. 67, no. 10, p. 105006, May 2022, doi: 10.1088/1361-6560/ac692e.

[4] V. T. Taasti *et al.*, "Validation of proton stopping power ratio estimation based on dual energy CT using fresh tissue samples," *Phys. Med. Biol.*, vol. 63, no. 1, p. 015012, Dec. 2017, doi: 10.1088/1361-6560/aa952f.

[5] M. Bazalova, J.-F. Carrier, L. Beaulieu, and F. Verhaegen, "Dual-energy CT-based material extraction for tissue segmentation in Monte Carlo dose calculations," *Phys. Med. Biol.*, vol. 53, no. 9, pp. 2439–2456, May 2008, doi: 10.1088/0031-9155/53/9/015.

[6] G. Landry *et al.*, "Simulation study on potential accuracy gains from dual energy CT tissue segmentation for low-energy brachytherapy Monte Carlo dose calculations," *Phys. Med. Biol.*, vol. 56, no. 19, pp. 6257–6278, Oct. 2011, doi: 10.1088/0031-9155/56/19/007.

[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] S. Zhang *et al.*, "Experimental implementation of a joint statistical image reconstruction method for proton stopping power mapping from dual-energy CT data," *Med. Phys.*, vol. 46, no. 1, pp. 273–285, Jan. 2019, doi: 10.1002/mp.13287.

[9] C. Mory, B. Sixou, S. Si-Mohamed, L. Boussel, and S. Rit, "Comparison of five one-step reconstruction algorithms for spectral CT," *Phys. Med. Biol.*, vol. 63, no. 23, p. 235001, Nov. 2018, doi: 10.1088/1361-6560/aaeaf2.

[10] K. Mechlem *et al.*, "Joint Statistical Iterative Material Image Reconstruction for Spectral Computed Tomography Using a Semi-Empirical Forward Model," *IEEE Transactions on Medical Imaging*, vol. 37, no. 1, pp. 68–80, Jan. 2018, doi: 10.1109/TMI.2017.2726687.

[11] Y. Nesterov, "A method for unconstrained convex minimization problem with the rate of convergence O(1/k^2)," in *Dokl. Akad. Nauk. SSSR*, in 3, vol. 269. 1983, p. 543.

[12] T. Weidinger, T. M. Buzug, T. Flohr, S. Kappler, and K. Stierstorfer, "Polychromatic Iterative Statistical Material Image Reconstruction for Photon-Counting Computed Tomography," *International Journal of Biomedical Imaging*, vol. 2016, pp. 1–15, 2016, doi: 10.1155/2016/5871604.

[13] ICRU, "Report 46," *Journal of the International Commission on Radiation Units and Measurements*, vol. os24, no. 1, Feb. 1992, doi: 10.1093/jicru/os24.1.Report46.

[14] K. Taguchi, "Multi-energy inter-pixel coincidence counters for charge sharing correction and compensation in photon counting detectors," *Med. Phys.*, vol. 47, no. 5, pp. 2085–2098, May 2020, doi: 10.1002/mp.14047.

[15] N. Ducros, J. F. P. Abascal, B. Sixou, S. Rit, and F. Peyrin, "Regularization of nonlinear decomposition of spectral x-ray projection images," *Medical Physics*, vol. 44, no. 9, Sep. 2017, doi: 10.1002/mp.12283.

[16] 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, pp. 4031–4047, Aug. 2008, doi: 10.1088/0031-9155/53/15/002.

[17] S. Zhang, D. Han, D. G. Politte, J. F. Williamson, and J. A. O'Sullivan, "Impact of joint statistical dual-energy CT reconstruction of proton stopping power images: Comparison to image- and sinogram-domain material decomposition approaches," *Med. Phys.*, vol. 45, no. 5, pp. 2129–2142, May 2018, doi: 10.1002/mp.12875.

Spectral CT reconstruction with ProMeSCT: results on real data from PIXSCAN.

Y. Boursier[1], F. Cannet[1], M. Dupont[1], S. Anthoine[2], C. Morel[1]

I. Introduction

We recently proposed the ProMeSCT algorithm, a one-step method to simultaneously separate and reconstruct the physical components of an object observed in the context of spectral Computed Tomography [1]. This task is done by solving a non-convex ill-posed inverse problem with a variable metric proximal algorithm that encompasses the underlying polychromatic model of the X-ray beam and the positivity of the quantities of interest. In this paper, we assess ProMeSCT on real data obtained on the spectral Photon Counting (PC) micro-CT PIXSCAN equiped with the XPAD3 hybrid pixels camera. Quantitative results are shown and detailed on 3D real data processed with ProMeSCT. For more details on the convergence analysis of ProMeSCT and its performances on simulated data including a comparison with the state-of-the-art, we refer to [1].

II. Results with ProMeSCT on real data

A. Material

The experimental datasets have been acquired with the PC-CT prototype PIXSCAN-FLI (Fig. 1) developed at CPPM. The PIXSCAN-FLI is made of three main blocks whose characteristics are:

- the X-ray source block embeds an X-ray tube either with a tungsten anode (Microfocus 90 KV TGT, Oxford Instruments, Scotts Valley CA, USA) with a 30 μm × 30 μm emission spot, or with a molybden anode (MCBM65 B-50, RTW, Berlin, Germany) with a 50 μm × 50 μm emission spot. A rotating wheel filter with 12 slots for metallic or liquid filters is placed in front of the X-ray tube outlet in order to harden the spectrum if need be.
- the animal block embeds a rotative motor with a $0.1°$ precision for performing tomographic scans. It can be also translated along the three axis with an 10 μm accuracy .
- the detector block has two slots for two different X-ray cameras that can be placed in the X-ray beam according to the user's choice thanks to translation motorized platform with a 10 μm accuracy.

The PIXSCAN-FLI is using a single threshold hybrid pixel detector XPAD3, either with 500 μm thick silicon pixels (referred to as XPAD3/Si pixels), either with 750 μm thick cadmium telluride (CdTe) pixels (referred to as XPAD3/CdTe pixels) or with 1000 μm thick gallium arsenide prorotype (referred to as XPAD3/GaAs pixels). The two latter have an increased detection efficiency with respect to XPAD3/Si, typically at energy above 30 keV, which allows to access and identify materials with a K-edge signature higher than 33.2 keV for iodine, *e.g.* 37.4 keV for barium or 80.7 keV for gold.

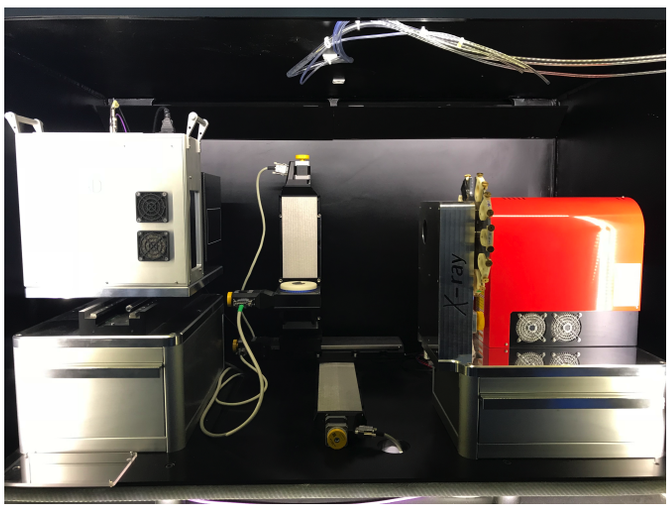


Fig. 1. The spectral micro Photon-Counting CT PIXSCAN-FLI with the three main parts, from left to right: the detector block, the animal block, the X-ray source block.

The XPAD3/GaAs detector is a single chip prototype that comprises $9,600$ (80×120) pixels of 130 μm × 130 μm. The two other XPAD3 detector are composed of modules of 7 readout chips hybridized to the pixels with bump bonds. The XPAD3/CdTe detector has only one module of about $70,000$ CdTe pixels for a detection surface of 7.5 cm × 1.5 cm. The XPAD3/Si detector comprises 8 modules tiled one above the other to avoid gaps between the modules. It ensures a global detection surface of 7.5 cm × 11.4 cm, which is sufficient to scan a whole mouse chest without dead areas.

Depending on the experiment, each scan consists of 360 or 720 projections around a $360°$ angular span followed by flat and dark field projections (typically between 25 and 100).

We suggest [2], [3] for a detailed description of the characteristics of the PC-CT prototype PIXSCAN based on both the XPAD3/Si and the XPAD3/CdTe hybrid pixel detectors.

B. Data preprocessing

Raw data are pre-processed in a pipeline that involves three main steps.

- Bad pixels are first identified thanks to the dark fields. Any pixel counting more than 10 photons in one of the dark fields is reported as bad.
- Dead and faulty pixels are then identified thanks to the flat fields. Any flat field is first normalized by the average flat field computed from all of them to attenuate the Poisson

[1] Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France.
[2] Institut de Mathématiques de Marseille, Aix Marseille Univ, CNRS, Centrale Marseille, Marseille, France.
This work was partly funded by the French program "Investissement d'Avenir" run by the 'Agence Nationale pour la Recherche' ; the grant reference is 'Infrastructure d'avenir en Biologie Santé - ANR-11-INBS-0006'

noise and cancel the inhomogeneous source profile of the cone beam. A cut is set to three times the standard-deviation of the histogram of normalized counts over all the pixels. Any pixel outside this cut is reported as bad.

- Planar raw images are eventually processed by 2D linear interpolation in order to i) estimate missing information due to bad pixels and ii) correct the complex internal geometry of the XPAD3/Si (tiled geometry) or XPAD3/CdTe (planar geometry with small gaps between readout chips). The internal geometry of each detector is known via metrology measurements.

Geometrical alignment of the micro-CT is obtained from the scan of steal-balls glued inside a hollow cylinder. Geometrical parameters are derived from the trajectories of the shadows of the balls on the detector, more details can be found in [6].

The energy threshold of each pixel is tuned independently before each acquisition. The energy threshold can be set over the noise or at a K-edge energy after precalibration using a filter made of the K-edge material, more details on this calibration process can be found in [2].

C. Performance with XPAD3/Si and XPAD3/CdTe pixels

In the first experiment, three series of acquisitions of the same phantom were acquired, one with the XPAD3/Si detector and two with XPAD3/CdTe detector. This experiment allows us to compare the performance of the proposed algorithm with the two pixel types at equivalent dose (statistics 11 times higher with the XPAD3/CdTe pixels and hence a lower level of noise) and at equivalent photon statistics (same level of noise with the two cameras, dose 11 times lower with the XPAD3/CdTe pixels). Each series consists in seven scans of 360 images for which only the energy threshold of the XPAD3 is varying through the K-edge of iodine (33.2 keV). We then processed with ProMeSCT each set of data constructed from three scans with thresholds at $\{33.2-\Delta, 33.2, 33.2+\Delta\}$ keV with $\Delta \in \{1.5, 3.0, 4.5\}$ keV. All the scans were performed under a 60 kVp tungsten anode X-ray spectrum filtered with 0.5 mm of Al with an almost parallel beam geometry.

The cylindrical phantom has 1 cm diameter with four cylindrical inserts of 3 mm diameter. The inserts were filled with water, two solutions of iodine (80 mg/mL and 40 mg/mL) and one solution of $CuSO_4$ at 1 M in order to compare with a material having an equivalent density, but without K-edge in the considered energy range. The model of the spectral response of the imaging systems provided in ProMeSCT incorporates the tungsten anode X-ray spectrum as provided by the manufacturer, the filtration by 0.5 mm of aluminum and the efficiency of the XPAD3/Si or XPAD3/CdTe pixels, given their respective thicknesses. Note that the theoretical model of the efficiency of the XPAD3/CdTe pixels does not take into account the charge sharing between the pixels, nor the fluorescence escape from the K_α rays of Cd and Te. For a fair comparison, we processed the XPAD3/Si data twice with ProMeSCT, one with the theoretical model and one with the efficiency curve simulated with GATE [4] including the charge sharing effect.

The iodine component maps estimated with ProMeSCT on the dataset with $\Delta = 4.5$ keV in these different cases are presented on Fig. 2 with a systematic comparison with K-edge subtraction [3]. The maps obtained with K-edge subtraction are noisier. The ProMeSCT maps exhibit circular patterns due to an inaccurate calibration of some pixels. Note that these patterns are not present outside of the rods on XPAD3/CdTe data at equivalent dose thanks to the high statistics (Fig. 2, bottom right). Quantitative results, *i.e.* estimation of iodine concentrations, are provided in Fig. 3. The estimation of concentration with ProMeSCT is quite satisfactory whatever the value of Δ.

D. In vivo imaging with the XPAD3/CdTe pixels

In the second experiment, a mouse injected with a contrast agent containing barium nanoparticles[1] to label Kupffer cells of the liver was imaged simultaneously with three different thresholds through the K-edge of barium: 33.9, 37.4 and 40.9 keV ([5]). Simultaneity of acquisitions is necessary to avoid motion artifacts that occur despite anesthesia, and is made possible by grouping pixels 3×3 with random pattern of thresholds evenly distributed throughout these three values in order to reduce aliasing effects. In this case, two thirds of the pixels do not provide any measurement in each energy channel. This missing information is estimated in preprocessing by 2D interpolation. The optimized scan protocol acquires 720 images followed by 50 flat fields and 50 dark fields under a 50 kVp tungsten anode X-ray spectrum with 0.6 mm of Al filtration. The dose for the scan amounts 330 mGy and a short acquisition time of about 12 minutes consistant with isoflurine gas anesthesia.

Fig. 4 shows the results obtained by K-edge subtraction and after 1000 iterations of the ProMeSCT algorithm. For a reference image, we also performed a standard acquisition with thresholds tuned over the noise followed by a standard FDK reconstruction. It is clear that the ProMeSCT approach allows a huge gain of quality on this example with respect to K-edge subtraction. The liver structure in barium maps are more accurate and better delineated with ProMeSCT with a significant reduction of noise. Note that the concentration of barium is higher in the estimation of ProMeSCT than with K-edge, which reflects the same trend previsouly observed. No ground truth is unfortunately available on this dataset. A residual signal of the gallow used to maintain the mouse, which is made of carbon, is visible at the level of noise. This low signal is the only crosstalk problem observed in barium maps. The water maps obtained with ProMeSCT (equivalent to Virtual Non Contrast images) exhibits clearly the bones that are only visible in this channel as well as all the soft tissues in this anatomic region. Few pixels randomly distributed in the liver have artificially high values in the water map, due to a bad identification of barium probably induced by noise in raw data.

E. Performance with the XPAD3/GaAs pixels at high energies

In the third experiment, we assessed the XPAD3/GaAs pixels performances for discriminating in the one hand contrast agents with closed K-edges (iodine and barium), and on the other hand for separating contrast agents at very high K-edge (gold was added to iodine and barium).

Firstly, the phantom was filled with a 40 mg/mL iodine solution, an 80 mg/mL iodine solution and a barium nanoparticle sample. The source was set at 60 kVp and 500 μA with a 0.6 mm aluminium filter. Six datasets of 360 images each were acquired around the K-edge of iodine (29.9, 33.2 and 36.5 keV) and around the K-edge of barium (34.1, 37.4 and 40.5 keV),

[1] Exitron™nano 1200, Miltenyi Biotec, Bergisch Gladbach, Germany

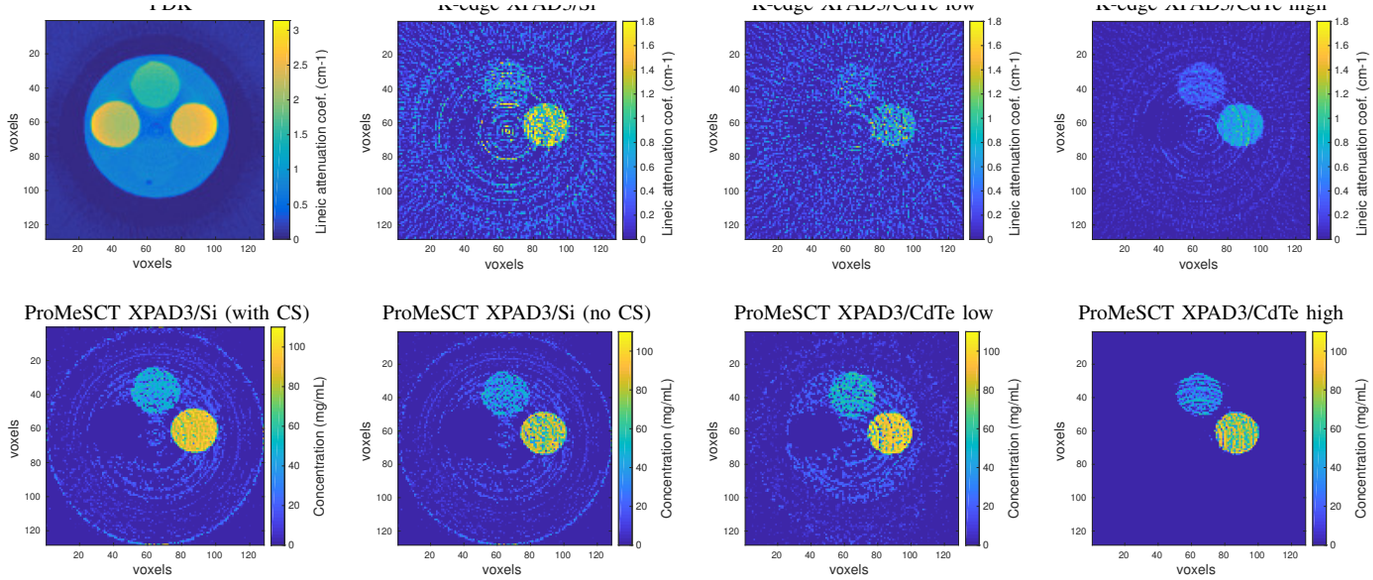


Fig. 2. Top, left to right: FDK reconstruction over the noise, iodine concentration maps obtained by K-edge subtraction for XPAD3/Si, XPAD3/CdTe at equivalent statistics (low) and XPAD3/CdTe at equivalent dose (high). Bottom, left to right: iodine concentration maps obtained at 400 iterations of ProMeSCT for XPAD3/Si with charge sharing modeling, XPAD3/Si without charge sharing modeling, XPAD3/CdTe at equivalent statistics (low) and XPAD3/CdTe at equivalent dose (high).

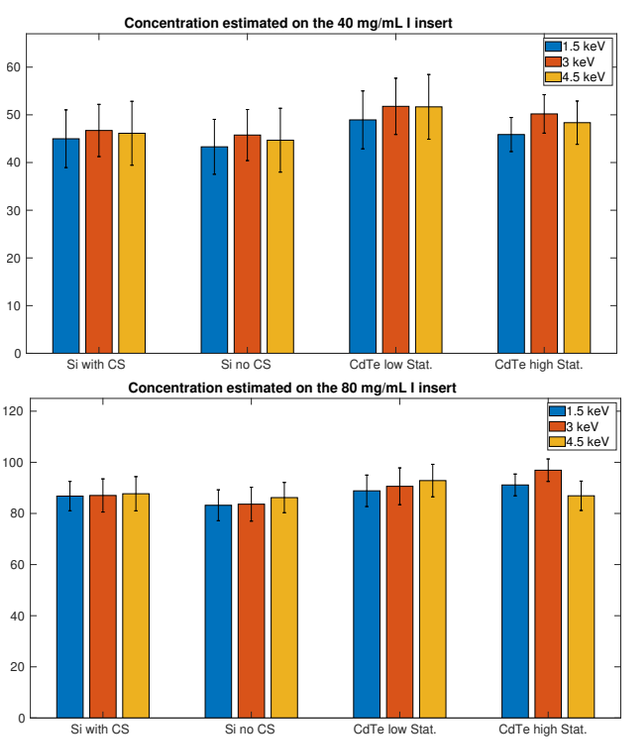


Fig. 3. Barplots display the concentrations of iodine inserts with 40 mg/mL (top) and 80 mg/mL (bottom) reconstructed with ProMeSCT at 400 iterations. The bars are centered on mean values and their error is the standard deviation.

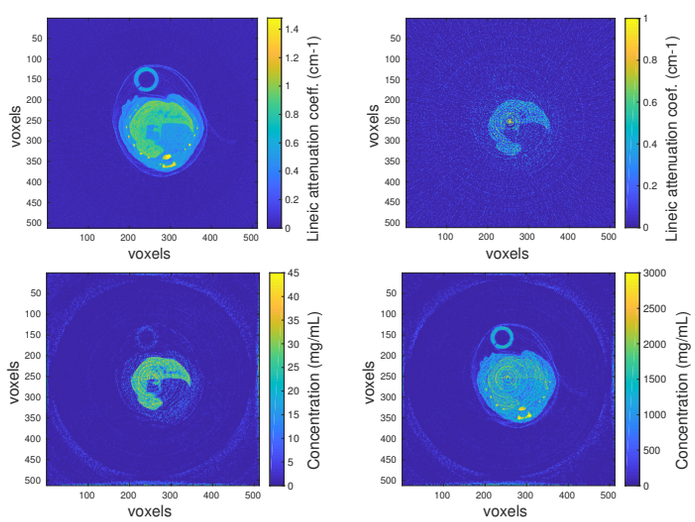


Fig. 4. In vivo mouse scan results reconstructed through a transverse slice. From top to bottom and left to right: FDK reconstruction for a standard attenuation scan over-the-noise, barium maps obtained with K-edge subtraction and with ProMeSCT, and water maps obtained with ProMeSCT.

preprocessed, then reconstructed with 1500 iterations of the ProMeSCT algorirhm. The Fig. 5 displays the results obtained with the ProMeSCT algorithm and shows the good separation of the two materials despite their close K-edge. Images obtained are slightly less noisy than with a post-reconstruction image subtraction approach because the algorithm uses single-screened scans and thus avoids the artificial increase in variance when subtracting scans in energy windows.

The second part of this experiment includes gold, a material with a very high K-edge energy (80.7 keV). The phantom is made up of a 350 mg/mL iodine solution, a solution of barium nanoparticles and a 10 μm gold foil wrapped around itself. In this case, nine datasets of 360 images each were acquired around the K-edge of iodine (29.9, 33.2 and 36.5 keV), around the K-edge of barium (34.1, 37.4 and 40.5 keV) and the K-edge of gold (74.7, 80.7 and 86.7 keV), then processed as previously described in the first part. We can observe a good separation of the different components: iodine, barium and gold have positive values in their respective channels. However, in the iodine channel, a partial signal of barium is always present, that can be explained by some calibration instabilities in energy thresholds for the XPAD3/GaAs prototype. On the other hand, for gold we can see that separation is possible and the ProMeSCT algorithm allows it to be placed in the right channel. The very weak signal from iodine and barium in the gold channel can be likened to noise, and a higher number of iterations of ProMeSCT could

make it possible to suppress this signal.

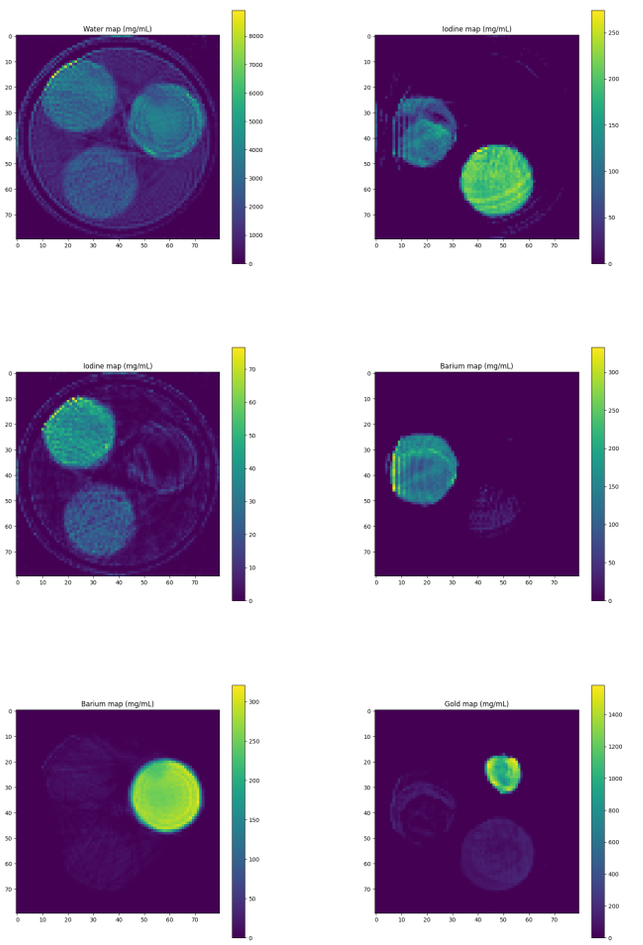


Fig. 5. Left: Results of spectral tomographies obtained with the ProMeSCT algorithm on the Iodine/Barium phantom: water-like map (top), iodine map (center) and barium map (bottom). Right: Results of spectral tomographies obtained with the ProMeSCT algorithm on the Iodine/Barium/Gold phantom: iodine map (top), barium map (center) and gold map (bottom). All the images represent an average of several tomographic slices for better visualisation of the results.

The various proofs of concept carried out with the XPAD3/GaAs detector have enabled us to confirm the feasibility of spectral tomography at high energies. At low energies, for iodine and barium, the separation is correct and the ProMeSCT algorithm enables us to obtain quantifications of the concentration of these different materials. A limit to this quantification remains the still imperfect knowledge of the response of the XPAD3/GaAs pixels in terms of charge sharing, which generates systematic uncertainties in the concentration estimates. We have also shown that we can use the XPAD3/GaAs detector at higher energies, in particular at the K-edge of gold, a material that is increasingly used in nanoparticle form.

Acknowledgement

This work was partly funded by the French program "Investissement d'Avenir" run by the 'Agence Nationale pour la Recherche' ; the grant reference is 'Infrastructure d'avenir en Biologie Santé - ANR-11-INBS-0006'. This work was also supported by the 80—Prime project DePIcT of the Mission pour les Initiatives Transverses et Interdisciplinaires (MITI) of CNRS.

References

[1] S. Tairi et al., ProMeSCT: a proximal metric algorithm for spectral CT, *IEEE Trans. Radiat. Plasma Med. Sci.* 5 (2021) 548-558.
[2] F. Cassol et al., Characterization of the imaging performance of a micro-CT system based on the photon counting XPAD3/Si hybrid pixel detectors, *Biomed. Phys. Eng. Express* 2 (2016) 025003.
[3] F. Cassol et al., K-edge imaging with the XPAD3 hybrid pixel detector, direct comparison of CdTe and Si sensors, *Phys. Med. Biol.* 60 (2015) 5497-5511.
[4] F. Cassol et al., Study of the charge sharing effect in the photon- counting pixel detector XPAD3-S, *Nucl. Instrum. Methods Phys. Res.* A, vol. 633, 05 2011.
[5] F. Cassol et al., Tracking dynamics of spontaneous tumours in mice using Photon Counting Computed Tomography, *Science* 21 (2019) 68-83.
[6] R. Khoury et al., A geometrical calibration method for the PIXSCAN micro-CT scanner, *JINST* 14-18

Accelerated iterative reconstruction for cone beam CT using measurement-domain filtering

Ville-Veikko Wettenhovi, Ari Hietanen, Ville Kolehmainen, Kati Niinimäki, and Marko Vauhkonen

***Abstract*—In this work we present an accelerated iterative reconstruction technique for cone beam CT (CBCT) image reconstruction. One of the major issues in applying iterative reconstruction for CBCT image reconstruction are the long computational times. While this can be alleviated by the use of subsets or, in some algorithms, custom computation of the step-sizes, the reconstruction speeds can still be quite long. This is even more highlighted when using regularization, especially regularization that is also computationally intensive. Thus it is important to find methods that allow iterative reconstructions in just 10-20 iterations, or even less. In this work, we present a measurement-domain based preconditioner for the primal-dual hybrid gradient algorithm. With this filtering-based measurement-domain preconditioner, combined with subsets, we can achieve sufficiently converged images in 10 iterations.**

***Index Terms*—cone beam, iterative, image reconstruction**

I. Introduction

ITERATIVE reconstruction is commonplace in regular computed tomography (CT) and in other similar imaging modalities such as positron emission tomography (PET) or single photon computed tomohpraphy (SPECT). However, for cone beam CT (CBCT), iterative image reconstruction is still rarely used. This is due to many of the issues present with the iterative reconstruction framework such as dealing with out-of-field-of-view data. One of the major issues is also the large computational costs, especially when reconstructing high-resolution, and thus high-dimensional, volumes. While some iterative algorithms can compute images with less than hundred iterations, these algorithms tend to lack the flexibility and robustness of some slower algorithms.

To accelerate computations in CBCT, a graphics processing unit (GPU) based reconstruction is very important. The core component of both the forward projection operator as well as the backprojection operator, ray-tracing, can be effectively parallized. This allows for significant increase of computational speeds. For effective iterative reconstruction, however, the forward projector should be ray-based while the backprojection should be voxel-based. Inherently these are not adjoint, which can cause issues with certain algorithms. However, it is possible to modify the backprojection operator to be nearly adjoint with the forward projection operator [1]. This modification allows the usage of ray-based forward projector and voxel-based backprojector with algorithms that require adjoint projectors.

V-V. Wettenhovi is with the Department of Technical Physics, University of Eastern Finland, Kuopio, Finland and Planmeca Group, Helsinki, Finland, e-mail: villewe@uef.fi.

A. Hietanen and K. Niinimäki are with Planmeca Group, Helsinki, Finland.

V. Kolehmainen and M. Vauhkonen are with the Department of Technical Physics, University of Eastern Finland, Kuopio, Finland.

Many different algorithms have been developed for iterative CT reconstruction. While most of the methods linearize the data, the original Poisson data could be used as well. An example of an algorithm using linearized data is the conjugate gradient least-squares (CGLS) [2] and for Poisson data the maximum-likelihood expectation maximization (MLEM) [3]. One common way to accelerate convergence, especially with Poisson-based algorithms, is to divide the measurement data into subsets that contain, for example, only a subset of the X-ray projections. However, even with the use of subsets, the convergence rate can still be subpar in order to obtain less than a minute reconstruction times.

Another method to accelerate convergence is the use of preconditioners. Preconditioners, usually diagonal matrices, act as specific weights in the reconstruction. Usually the preconditioners are applied to the image-domain, but measurement-domain is also applicable. Measurement-domain preconditioning, however, often does not have solid mathematical foundation and thus can be reliable used with only some algorithms.

In the last few decades, there has been an interest in primal-dual optimization methods. These algorithms do not simply solve the *primal* problem, which in the least-squares sense is the computation of the norm of the residual, but also solve the *dual* problem. Due to this, the primal-dual algorithms are more robust and flexible and, for example, allow mathematically solid measurement-domain preconditioning. One such an algorithm is the the primal-dual hybrid gradient (PDHG) algorithm [4], [5]. The PDHG algorithm also allows acceleration with the use of subsets [6] making it a potential algorithm. However, the PDHG algorithm also has its drawbacks. These include slow convergence, even with the use of subsets, and the requirement for matched forward and backward projection operators. The latter, however, is easy to overcome. There also exists methods to dynamically adjust the primal and dual step-sizes in PDHG in every iteration [7]. These methods accelerate the convergence at the expense of additional computations and increase in memory requirements.

In this work we present a new preconditioner to accelerate reconstruction in CBCT using GPU-based computing. While this preconditioner is implemented for the subset-based PDHG, it is viable in other algorithms as well, for example FISTA [8]. The preconditioner works in the measurement-domain by filtering the residual in the computation of the dual-parameter. This is a very similar procedure to that of the FDK algorithm, where the projections are also filtered before backprojection. This is also computationally very efficient as only 1D FFTs are required. For PDHG, the filtering-based preconditioner can also be shown to be mathematically viable.

A similar type of preconditioner has been implemented for non-Cartesian MRI reconstruction in [9]. Example reconstructions are presented with a knee phantom measured with the Planmeca Viso G7 CBCT scanner.

II. Mathematical background

The PDHG algorithm uses the linearized version of the Lambert-Beer law

$$\hat{\boldsymbol{n}} = \ln\left(\frac{n_0}{\boldsymbol{n}}\right) = \boldsymbol{A}\boldsymbol{f}, \tag{1}$$

where $\boldsymbol{n}$ are the original projection measurements, n_0 the flat value, $\boldsymbol{A}$ the system matrix and $\boldsymbol{f}$ the image.

The PDHG is a primal-dual algorithm meaning it includes the primal minimization of

$$\min_x \left(F(\hat{\boldsymbol{A}}\boldsymbol{x}) + G(\boldsymbol{x})\right), \tag{2}$$

and the dual maximization of

$$\max_y \left(-F^*(\boldsymbol{y}) - G^*(-\hat{\boldsymbol{A}}^T\boldsymbol{y})\right), \tag{3}$$

where * denotes the convex conjugate. Now

$$F\begin{pmatrix} y \\ z \end{pmatrix} = F_1(y) + F_2(z) \tag{4}$$

$$F_1(y) = \frac{1}{2}\left\|\boldsymbol{y} - \hat{\boldsymbol{n}}\right\|_2^2 \tag{5}$$

$$F_2(z) = \beta R(\boldsymbol{z}) \tag{6}$$

$$G(x) = 0 \tag{7}$$

$$x = f \tag{8}$$

$$y = Af \tag{9}$$

$$\hat{A} = A. \tag{10}$$

The definition of z depends on the prior function R and $\hat{A}$ should include the necessary operator to achieve $F(\hat{A}x) = F\begin{pmatrix} y \\ z \end{pmatrix}$, i.e. it is the linear map from one vector space to another. Prior functions that are, for example, non-linear do not fulfill this requirement and thus they cannot be used with the PDHG. In this case the prior function is from now on ignored as regularization is not considered in this paper.

The general form of (primal) PDHG is defined in algorithm 1. Note that the PDHG can also be used as a "dual" version, where the update equation, with θ, is applied to the dual parameter of $\hat{p}$. The subset version later utilizes this dual method. The primal and dual step-sizes, τ and σ respectively, require the computation of the largest singular value, or eigenvalue, L_{singular} of the matrix/operator $\boldsymbol{A}$. This can be efficiently computed with the power method.

The primal and dual step-sizes can be modified, or completely replaced, by preconditioner matrices. In this case, the filtering preconditioner is includes with σ as the matrix $\boldsymbol{C}\mathcal{F}^{-1}\boldsymbol{M}\mathcal{F}$, where $\mathcal{F}$ is the Fourier transform operator and $\boldsymbol{M}$ is a diagonal matrix containing the filtering coefficients. The Fourier transform operations are performed in 1D for each row in the projection. The dual PDHG with the filtering preconditioner is presented in algorithm 2.

The final preconditioned version of PDHG is presented in algorithm 3 with S number of subsets.

Algorithm 1 General (primal) PDHG algorithm

$L_{\text{singular}} \leftarrow \|A\|_2$, $\sigma\tau \leftarrow 1/L_{\text{singular}}^2$, $\theta \leftarrow 1$, $k \leftarrow 0$
$\hat{f}_0 \leftarrow f_0$
for $k = 1, \ldots, K$ **do**
  $\hat{p}_k \leftarrow prox_\sigma\left[F^*\right]\left(\hat{p}_{k-1} + \sigma A\hat{f}_{k-1}\right)$
  $f_k \leftarrow prox_\tau\left[G\right]\left(f_{k-1} - \tau A^T\hat{p}_k\right)$
  $\hat{f}_k \leftarrow f_k + \theta\left(f_k - f_{k-1}\right)$
end for

Algorithm 2 L_2 norm (dual) PDHG algorithm with a positivity constraint

$L_{\text{singular}} \leftarrow \|A\|_2$, $\sigma\tau \leftarrow 1/L_{\text{singular}}^2$, $\theta \leftarrow 1$, $k \leftarrow 0$
$\hat{f}_0 \leftarrow f_0$
for $k = 1, \ldots, K$ **do**
  $\hat{p}_k \leftarrow prox_\sigma\left[F^*\right]\left(\hat{p}_{k-1} + \sigma CAf_{k-1}\right)$
  $\bar{p}_k \leftarrow \hat{p}_k + \theta\left(\hat{p}_k - \hat{p}_{k-1}\right)$
  $f_k \leftarrow prox_\tau\left[G\right]\left(f_{k-1} - \tau A^T\bar{p}_k\right)$
end for

The inverse in algorithm 3 can be computed efficiently since $\boldsymbol{C}$ and $\boldsymbol{I}$ are circulant matrices [10], [11], and thus their summation is also circulant,

$$\begin{aligned} \boldsymbol{y} &= \boldsymbol{C}\boldsymbol{x} \rightarrow \\ \boldsymbol{y} &= \boldsymbol{c} * \boldsymbol{x} \rightarrow \\ \mathcal{F}(\boldsymbol{y}) &= \mathcal{F}(\boldsymbol{c})\mathcal{F}(\boldsymbol{x}) \rightarrow \\ \boldsymbol{x} &= \mathcal{F}^{-1}\left(\frac{\mathcal{F}(\boldsymbol{y})}{\mathcal{F}(\boldsymbol{c})}\right), \end{aligned} \tag{11}$$

where $*$ denotes convolution and $\boldsymbol{c}$ is the first column/row of $\boldsymbol{C}$. The division is done element-wise.

III. Methods

We use knee phantom data measured with the Planmeca Viso G7 CBCT scanner. The total number of projections used is 500 with projections of size 738×616 and the detector size of 0.278 mm. The transaxial field-of-view (FOV) is 120 mm while the axial is 100 mm. The phantom is 29 mm offset from

Algorithm 3 L_2 norm PDHG algorithm with subsets, a positivity constraint and preconditioners

$L_{\text{singular}} \leftarrow \|\boldsymbol{A}\|_2$, $\sigma\tau \leftarrow 1/L_{\text{singular}}^2$, $\theta \leftarrow 1$, $k \leftarrow 0$
$\hat{\boldsymbol{f}}_0 \leftarrow f_0$
for $k = 1, \ldots, K$ **do**
  for $l = 1, \ldots, S$ **do**
    $\bar{\boldsymbol{p}}_l \leftarrow \left(\hat{\boldsymbol{p}}_{k-1,l} + \sigma\boldsymbol{C}_l\left(\boldsymbol{A}_l\boldsymbol{f}_{k,l-1} - \hat{\boldsymbol{n}}\right)\right)\left(\boldsymbol{I} + \sigma\boldsymbol{C}_l\right)^{-1}$
    $\Delta\boldsymbol{g} \leftarrow \boldsymbol{A}_l^T\left(\bar{\boldsymbol{p}}_l - \hat{\boldsymbol{p}}_{k-1,l}\right)$
    $\boldsymbol{g} \leftarrow \boldsymbol{g} + \Delta\boldsymbol{g}$
    $\hat{\boldsymbol{p}}_{k,l} \leftarrow \bar{\boldsymbol{p}}_l$
    $\bar{\boldsymbol{g}} \leftarrow \boldsymbol{g} + \frac{\theta}{P_l}\Delta\boldsymbol{g}$
    $\boldsymbol{f}_{k,l} \leftarrow \text{pos}\left(\boldsymbol{f}_{k,l-1} - \tau\bar{\boldsymbol{g}}\right)$
  end for
  $\boldsymbol{f}_{k+1,0} \leftarrow \boldsymbol{f}_{k,S}$
end for

the origin. The final reconstructed image matrix size is $801 \times 801 \times 668$. However, to avoid out-of-FOV artifacts we utilized extended FOV with multi-resolution reconstruction [12] in the axial direction. The matrix size with multi-resolution is $801 \times 801 \times 1202$. In multi-resolution imaging, the extended FOV region has lower resolution than the actual FOV. In this case, the extended FOV size has a voxel size of $4\times$ of the original size. The original voxel size is 0.15 mm. All reconstructions used 10 full iterations (epochs) and 20 subsets.

To further alleviate the out-of-FOV artifacts, we also extrapolated the projections in the column-dimension. The size of the extrapolated projections is 738×860. The extrapolation was a simply next/previous interpolation with sine-based weighting, starting from 1 and going to 0.

The projection data was precorrected for bad pixels and flat-field, no other corrections were performed. All computations were performed by using our own open-source software OMEGA [13] and using AMD Radeon 7900 XT GPU. All reconstruction related computations were performed on the GPU. For the forward and backward projections, we used a ray-based approach for forward projection and a voxel-based one for backprojection, weighted with the weights presented in [1]. All the GPU code was implemented in OpenCL [14], while the filtering was computed by using the open-source software ArrayFire [15] using the OpenCL backend.

IV. Results

Figure 1 shows the primal-dual gap value for 20 iterations and 20 subsets, leading to a total of 400 subiterations, for five different cases. Figure 2 we show comparison images for the five different cases, in addition with a "ground truth" image which is the converged (noisy) image. In Figure 2 with all cases, 20 subsets are used and, aside from the ground truth, 10 full iterations (epochs). Most of the cases use air image as the initial value, i.e. the initial value corresponds to an image with attenuation coefficient close to that of air while some use the FDK reconstruction obtained by the Planmeca software as the initial value. Some of the cases use the adaptive step size used in [7]. The top left subfigure in figure 2 is PDHG with no filtering, no adaptive step sizes and the air image as the initial value. The top right one is the same except that it uses adaptive step-sizes. The center left uses the FDK image as the initial value in addition to the adaptive step sizes. Center right uses FDK initial value, but no adaptive step-sizes. Bottom right is the only subfigure with filtering with the first 5 full iterations (epochs) with filtering and the later 5 without. It also does not have adaptive step-sizes and uses air image as the initial value. Bottom right serves as the (noisy) ground truth, and has 500 full iterations (epochs) with 20 subsets. No adaptive step-sizes or filtering was used with the ground truth and air image was used as the initial value.

Comparing the figures in Figure 2, it is clear to see that the filtering causes much better convergence than the others. While the adaptive one is better than the one without it or the one with the FDK initial value, it is still not quite close to that of filtering. The FDK initial value helps, but does not provide significant benefit with adaptive step sizes. Curiously, with the primal-dual gap value the filtering exhibits noticeable increase after the initial decrease, though decreasing again after this. In general, the primal-dual gap does not correlate well with the visual assessment of convergence. Similar behavior is shown in [9].

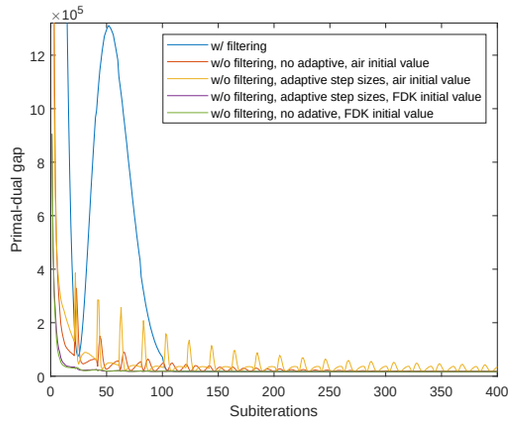


Fig. 1. Primal-dual gap value for each subiteration for all five different tested methods.

We also used filtering with the FDK initial value and adaptive step-sizes, but neither provided any benefit. The filtering is also only about 8-15% slower than without filtering, while about 20% faster than with adaptive-step sizes.

V. Conclusion

In this paper, we presented a new preconditioning method to accelerate convergence in CBCT image reconstruction. While the preconditioner is usable with other algorithms too, it is especially well-suited for use with the PDHG algorithm. As shown, it is very fast to compute as it only required 1D FFT operations, yet it provides very fast convergence. It can also be combined with regularization methods, allowing quickly converged regularized reconstruction and thus reducing the need for many regularization steps. The downside is that since the preconditioner works in the measurement-domain, it is not naturally applicable to all algorithms.

References

[1] X. Jia, B. Dong, Y. Lou, and S. B. Jiang, "GPU-based iterative cone-beam CT reconstruction using tight frame regularization," *Physics in Medicine and Biology*, vol. 56, pp. 3787–3807, may 2011.

[2] M. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, vol. 49, p. 409, dec 1952.

[3] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. B.*, vol. 39, no. 1, pp. 1–38, 1977.

[4] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[5] 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 and Biology*, vol. 57, pp. 3065–3091, apr 2012.

[6] M. J. Ehrhardt, P. Markiewicz, and C.-B. Schönlieb, "Faster PET reconstruction with non-smooth priors by randomization and preconditioning," *Phys. Med. Biol.*, vol. 64, p. 225019, nov 2019.

[7] A. Chambolle, C. Delplancke, M. J. Ehrhardt, C.-B. Schönlieb, and J. Tang, "Stochastic primal dual hybrid gradient algorithm with adaptive step-sizes," Jan. 2023.

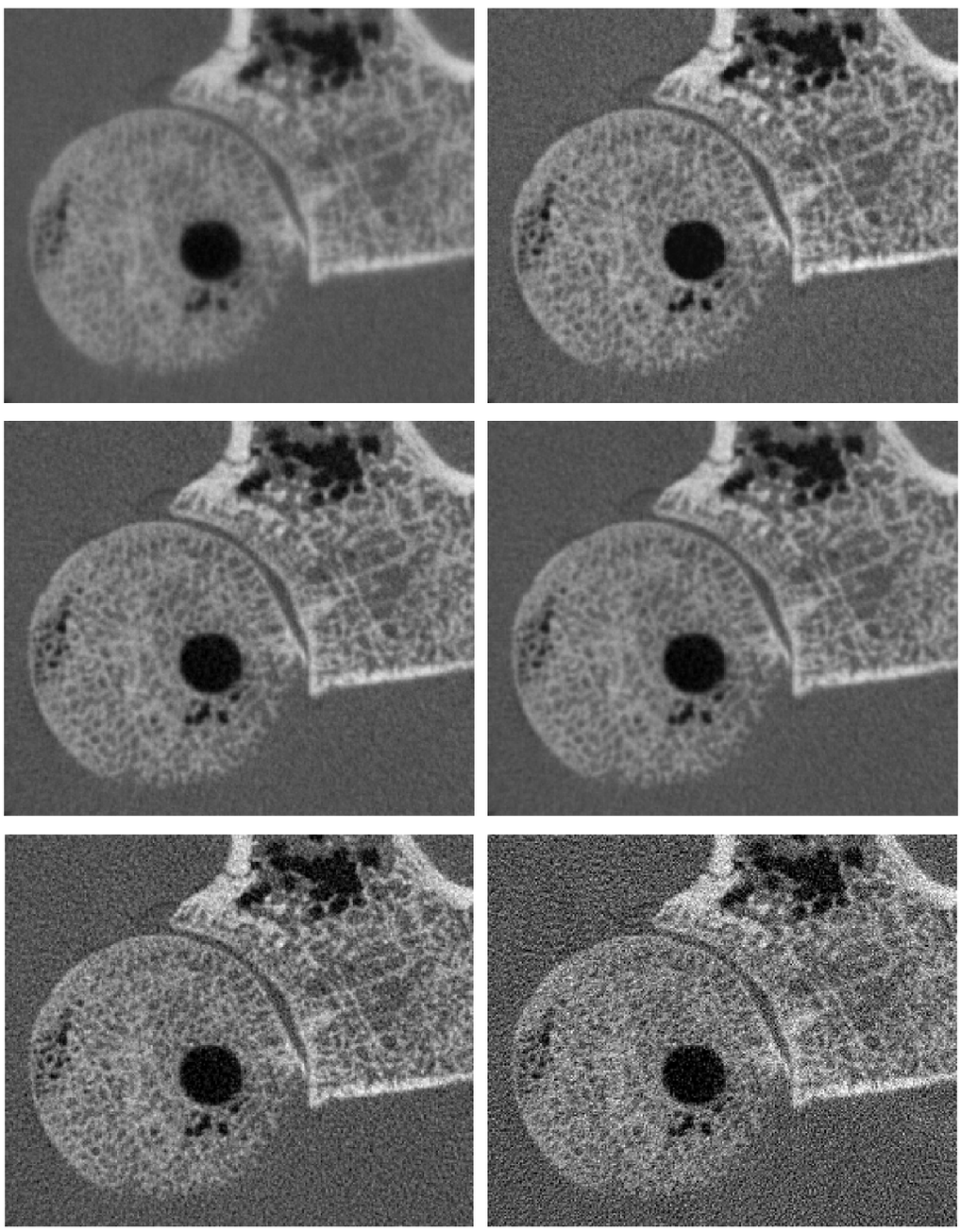


Fig. 2. Close-up reconstructions of the knee phantom with six different case. Reconstruction times for each are presented in parenthesis. Top left: No filtering, no adaptive step-sizes, air image as the initial value (59.8 s). Top right: No filtering, adaptive step-sizes, air image as the initial value (80.2 s). Center left: No filtering, adaptive step-sizes, FDK image as the initial value (79.0 s). Center right: No filtering, no adaptive step-sizes, FDK image as the initial value (52.6 s). Bottom left: 5 filtering epochs, no adaptive step-sizes, air image as the initial value (64.0 s). Bottom right: Ground truth with 500 epochs, no filtering, no adaptive step-sizes, air image as the initial value (2362.5 s).

[8] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, pp. 183–202, jan 2009.

[9] F. Ong, M. Uecker, and M. Lustig, "Accelerating non-Cartesian MRI reconstruction convergence using k-space preconditioning," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 1646–1654, May 2020.

[10] R. Aldrovandi, *Special Matrices of Mathematical Physics*. World Scientific Publishing Company, 2001.

[11] P. J. Davis, *Circulant matrices*. Wiley, 1979.

[12] H. Dang, J. W. Stayman, A. Sisniega, W. Zbijewski, J. Xu, X. Wang, D. H. Foos, N. Aygun, V. E. Koliatsos, and J. H. Siewerdsen, "Multi-resolution statistical image reconstruction for mitigation of truncation effects: application to cone-beam CT of the head," *Physics in Medicine and Biology*, vol. 62, pp. 539–559, dec 2016.

[13] V.-V. Wettenhovi, M. Vauhkonen, and V. Kolehmainen, "OMEGA—open-source emission tomography software," *Phys. Med. Biol.*, vol. 66, p. 065010, mar 2021.

[14] J. E. Stone, D. Gohara, and G. Shi, "OpenCL: A parallel programming standard for heterogeneous computing systems," *Comput. Sci. Eng.*, vol. 12, pp. 66–73, may 2010.

[15] P. Yalamanchili, U. Arshad, Z. Mohammed, P. Garigipati, P. Entschev, B. Kloppenborg, J. Malcolm, and J. Melonakos, "ArrayFire - A high performance software library for parallel computing with an easy-to-use API," 2015.

A Learnt Half-Quadratic Splitting-Based Algorithm for Fast and High-Quality Industrial Cone-beam CT Reconstruction

Aniket Pramanik, Singanallur V. Venkatakrishnan, Obaidullah Rahman, Amirkoushyar Ziabari

***Abstract*—Industrial X-ray cone-beam CT (XCT) scanners are widely used for scientific imaging and non-destructive characterization. Industrial CBCT scanners use large detectors containing millions of pixels and the subsequent 3D reconstructions can be of the order of billions of voxels. In order to obtain high-quality reconstruction when using typical analytic algorithms, the scan involves collecting a large number of projections/views which results in large measurement times - limiting the utility of the technique. Model-based iterative reconstruction (MBIR) algorithms can produce high-quality reconstructions from fast sparse-view CT scans, but are computationally expensive and hence are avoided in practice. Single-step deep-learning (DL) based methods have demonstrated that it is possible to obtain fast and high-quality reconstructions from sparse-view data but they do not generalize well to out-of-distribution scenarios. In this work, we propose a half-quadratic splitting-based algorithm that uses convolutional neural networks (CNN) in order to obtain high-quality reconstructions from large sparse-view cone-beam CT (CBCT) measurements while overcoming the challenges with typical approaches. The algorithm alternates between the application of a CNN and a conjugate gradient (CG) step enforcing data-consistency (DC). The proposed method outperforms other methods on the publicly available Walnuts data-set.**

***Index Terms*—Convolutional neural network, Cone-beam CT, hybrid, regularization, deep-learning, iterative algorithm.**

I. Introduction

Industrial X-ray cone-beam CT (XCT) scanners are widely used for high-resolution scientific imaging and non-destructive characterization. In order to conduct a CT scan, an object is rotated about a single axis and multiple projections are acquired (see Fig. 1), that are typically processed by an analytic algorithm such as the Feldkamp-Davis-Kress (FDK) [1] after appropriate pre-processing to obtain a 3D reconstruction. Unlike in medical CT, industrial XCT scanners use large detectors containing millions of pixels and the subsequent 3D reconstructions can be of the order of billions of voxels. A heuristic that is often used to obtain high-quality reconstruction is that the number of views needed is of the order of the number of columns of the detector, which can corresponds to total scan times of the order of several tens of minutes or even hours depending on the the material, desired reconstruction grid size and energy of the source.

One way to accelerate the XCT scans is to acquire fewer projections, but naively processing them using the FDK algorithm results in reconstructions with significant artifacts. Model-based iterative reconstruction (MBIR) algorithms [2], [3], that exploit the underlying imaging physics and edge-preserving models to regularize the image recovery, have been developed and shown superior performance over analytical techniques for sparse-view cone-beam medical XCT. However, these techniques are computationally very expensive when dealing with large volumes as is typical for industrial cone-beam XCT which has made them impractical for routine use.

Single-stage deep learning (DL) approaches that utilize a convolutional neural network (CNN) to directly infer the final reconstruction from an initial analytic reconstruction [4] have demonstrated that it is possible to obtain high-quality reconstructions while maintaining a lower computational complexity from sparse-view industrial XCT data [5], [6]. However, such methods require appropriate types of training data and do not generalize well to out of distribution scenarios such as varying measurement conditions. Furthermore, such methods lack any knobs to control image quality at test time in case there is any variation in imaging conditions. Another class of algorithms developed recently are Deep Unrolled (DU) networks which combine the physics-based model with a data-driven DL model. These methods avoid overfitting and have shown reduced training data demands while improving performance [7], [8]. However, these methods can require a large amount of memory that scales with the number of iterations - making them challenging for large scale industrial XCT. Anish et. al [9] have proposed a memory-efficient DU approach consisting of a multi-iteration network alternating between a de-streaking CNN and the physics-based model and demonstrated its value for small data-sets.

In this paper, we propose a learnt half-quadratic splitting algorithm that uses CNN to replace the proximal operation. The algorithm alternates for a fixed number of outer iterations between a CNN and the data-consistency (DC) block as shown in Fig. 2. A regularization parameter controls the effect of both the blocks on reconstruction. The CNN essentially performs a denoising operation which is equivalent to solving for proximal operation for the regularizer. Unlike DU, it has a memory efficient implementation where CNN in each iteration is trained separately. We utilize a 2D CNN for training it with 2D patches obtained from the 3D volume images. We validate our method on the publicly available Walnuts[10] cone-beam CT data-set - and demonstrate superior performance compared

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). Research sponsored by the US Department of Energy, Office of Energy Efficiency and Renewable Energy, Advanced Manufacturing Office and Technology Commercialization Fund (TCF-21-24881), under contract DE-AC05-00OR22725 with UT-Battelle, LLC. The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

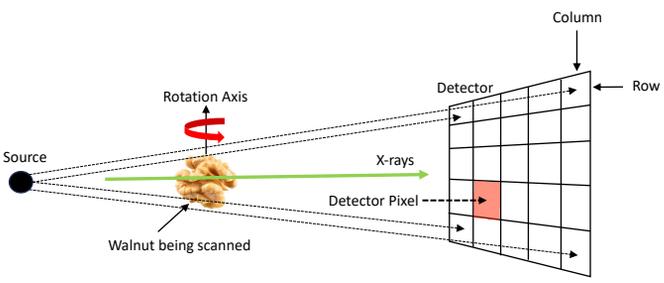


Fig. 1. The Cone-beam CT geometry. A cone-beam source of X-rays is incident on an object and the resulting transmitted signal is captured by a detector. The object is rotated and scanned at different angles (views).

to the traditional MBIR and a single-stage U-Net approach. Overall, it reduces the time complexity significantly compared to MBIR. A key benefit of this memory efficient framework is its ability to be extended to large-scale image reconstruction problems without compromising on the performance.

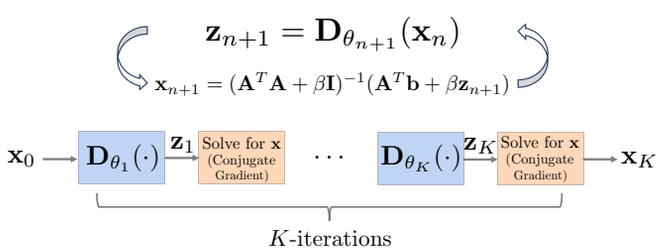


Fig. 2. Iterative rule of the proposed algorithm and its network architecture. **D** represents a artifact suppression CNN and the block in orange represent a data-consistency block that ensures a good fit to the measurements.

II. Theory

MBIR methods have been widely developed for X-ray CT and obtain the reconstruction by minimizing a cost function

$$\arg\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \mathbf{Q}(\mathbf{x}) \tag{1}$$

where $\mathbf{b} \in \mathbb{R}^M$ is the log-normalized measurement vector, $\mathbf{x} \in \mathbb{R}^N$ is the 3D image to be recovered, $\mathbf{A} \in \mathbb{R}^{M \times N}$ is the cone-beam forward projector, $\mathbf{Q}(\cdot)$ is a regularization function that enforces certain desirable properties in the solution, and λ balances the effect of regularizer and the data-consistency (DC) term. Several algorithms such as optimized gradient methods [11] and iterative coordinate descent [2], [3] have been developed to find a solution to (1) when using differentiable regularizers, but these are computationally very expensive when dealing with large data-sets in industrial XCT.

Instead, we propose to use a different approach that uses a fixed number of outer iterations of the half-quadratic splitting (HQS) method and replace certain blocks with learnable functions. The HQS for solving (1) results in a algorithm that alternates between the following two steps,

$$\mathbf{z}_{n+1} = \arg\min_{\mathbf{z}} \lambda \mathbf{Q}(\mathbf{z}) + \frac{\beta}{2}\|\mathbf{x}_n - \mathbf{z}\|_2^2 \tag{2}$$

$$\mathbf{x}_{n+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \frac{\beta}{2}\|\mathbf{x} - \mathbf{z}_{n+1}\|_2^2. \tag{3}$$

where β is a penalty parameter. The solution to the sub-problems in (2), (3) alternates between a DC step and a proximal operation,

$$\mathbf{x}_{n+1} = (\mathbf{A}^T\mathbf{A} + \beta\mathbf{I})^{-1}(\mathbf{A}^T\mathbf{b} + \beta\mathbf{z}_{n+1}) \tag{4}$$

$$\mathbf{z}_{n+1} = \text{prox}_{\frac{\lambda}{\beta}\mathbf{Q}}(\mathbf{x}_n). \tag{5}$$

The step in (5) is a proximal operation for the regularizer $\mathbf{Q}(\cdot)$. It is equivalent to solving a Gaussian denoising problem whose maximum a posteriori (MAP) estimate is equivalent to solving for (2).

In the spirit of methods like plug-and-play priors (PnP) [12] we replace the denoising step in (5) with a CNN $\mathbf{D}_\theta(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^N$ with weights θ. Therefore, the sub-problems (2) and (3) are solved approximately using the following equations,

$$\mathbf{z}_{n+1} = \mathbf{D}_{\theta_{n+1}}(\mathbf{x}_n) \tag{6}$$

$$\mathbf{x}_{n+1} = (\mathbf{A}^H\mathbf{A} + \beta\mathbf{I})^{-1}(\mathbf{A}^H\mathbf{b} + \beta\mathbf{z}_{n+1}). \tag{7}$$

The network architecture is shown in Fig. 2. The CNN $\mathbf{D}_\theta$ can have either shared or unshared weights across iterations i.e we can train a collection of CNNs - one for each outer iteration. In practice, the DC step in (7) is solved approximately using a fixed number of conjugate gradient (CG) algorithm iterations. The optimization steps in (6), (7) are essentially solving for,

$$\mathbf{x}_{n+1} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \frac{\beta}{2}\|\mathbf{x} - \mathbf{D}_{\theta_{n+1}}(\mathbf{x}_n)\|_2^2. \tag{8}$$

The regularization parameter β in (8) is chosen empirically and kept fixed across the iterations during training. β balances between the denoising effect of the CNN and data-consistency. β acts like a knob that can be varied during inference for unseen imaging conditions.

A. Memory Efficient Training and Implementation

We propose to train the CNNs in each iteration separately. We consider CNNs with 2D convolution layers and therefore split the 3D CBCT data into several 2D patches for training. This approach reduces GPU memory consumption. During inference, the 3D volume image is obtained through the CNN in 2D slice-by-slice fashion. The CNNs from subsequent iterations are trained in a similar fashion with the data reconstructed from the previous iteration. The training is parallelized across multiple GPUs along the batch dimension of the training data. The DC step is implemented using the CBCT forward operator from the ASTRA Toolbox [13] which can also be distributed over multiple GPUs. The training loss is the mean squared error (MSE) with expectation over the training samples,

$$\theta_{n+1} = \arg\min_{\theta} \mathbb{E}\left[\|\mathbf{D}_\theta(\mathcal{P}(\mathbf{x}_n)) - \mathcal{P}(\mathbf{x}_{\text{MBIR/FDK}})\|_2^2\right] \tag{9}$$

where ground truth is either MBIR ($\mathbf{x}_{\text{MBIR}}$) or FDK ($\mathbf{x}_{\text{FDK}}$) reconstruction on the corresponding dense-view data, and $\mathcal{P}(\cdot)$ is a 2D patch extraction operator. The pseudo-code for the algorithm is shown in Algorithm 1. The variable ep denotes training epoch, n is for iteration number and i ranges over N_t training samples.

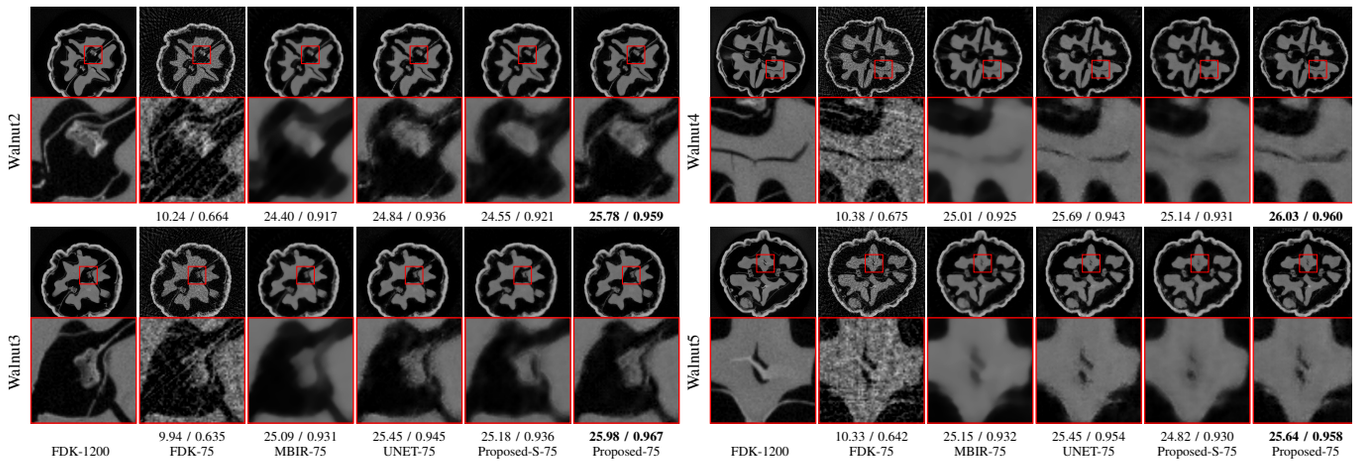


Fig. 3. Performance comparisons on Walnut Data. Walnut1 is the training data and the rest are testing data. Proposed-S-75 is the shared weights version of the algorithm applied on 75 views - a subsampling factor of 16 compared to the original measurements. PSNR in dB and SSIM values are reported for the zoomed patches. The proposed method with unshared weights preserves edges and details better compared to rest of the methods.

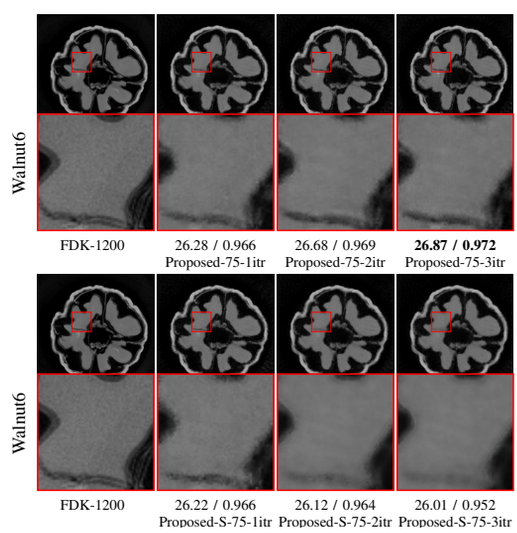


Fig. 4. Performance comparison of the two versions of the proposed method, shared vs unshared CNN weights. Each row displays results from three iterations. Proposed-S-75-Kitr stands for the shared weights version on 75 views at K^{th} iteration while Proposed is the unshared one. Proposed-S is missing details due to excessive smoothing while the edges get sharper with more iterations for unshared weights. PSNR in dB and SSIM values are reported for the zoomed patches.

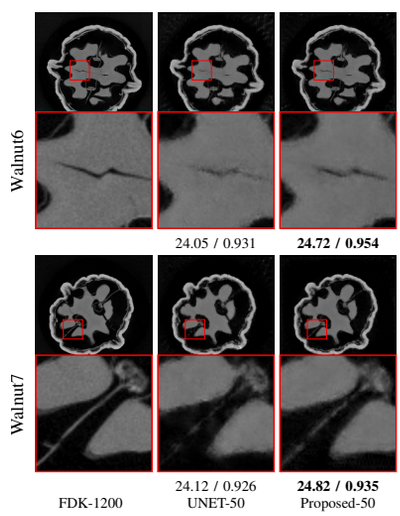


Fig. 5. Performance comparison of U-Net and the proposed method with unshared weights on out-of-distribution data as characterized by a different sparse-view measurement scenario. PSNR in dB and SSIM are reported for the zoomed patches. The results are shown on data from different sub-sampling scenario (50 views) which the networks have not been trained with. The proposed method shows significantly improved performance both visually and also in terms of performance metrics.

Considering K as number of iterations, the proposed method and MBIR have a time and space complexity of the order of $\mathcal{O}(K)$ and $\mathcal{O}(1)$ respectively. DU methods instead perform end-to-end training, and therefore, have a complexity of $\mathcal{O}(K)$ for both time and space. DL has the best possible complexity of $\mathcal{O}(1)$ for both. Therefore, our proposed approach reduces space complexity by a factor of K over DU.

III. Experiments and Results

A. Data-set, CNN Architecture and Methods for Comparisons

Experiments have been conducted on the publicly available CBCT Walnut data [10]. The detector size is 972 $\times$ 768 pixels and subsequently each scan consists of projection data from 1200 views (angles). In order to evaluate our algorithm on sparse-view data, we sub-sampled the views by factor of 16 to obtain sparse-view data (75 views) - simulating more than an order of magnitude faster scan time than the typical rate. The ground truth for training is an FDK reconstruction with 1200 views. The reconstructed volume has dimensions 450 $\times$ 380 $\times$ 380 voxels. The training data only includes a single scan (Walnut1) and all the methods are tested on six different scans (Walnut2-7).

We consider $K = 3$ iterations for the proposed method. A full-size U-Net with four pooling and unpooling layers is chosen as the CNN for each iteration. It consists of 3 x 3 filters with the initial number of output channels being 64. The CNNs are trained with 2D patches of size 256 x 256 and the batch-size is set as 64. The training is parallelized across four 16GB Nvidia P100 GPUs. The CNN parameters are optimized using

Algorithm 1 : Training: input=training data $\mathbf{x}_i; i = 1, .., N_t$

```
for ep = 1, 2, ... do
    for n = 1, 2, ..., K do
        for i = 1, 2, ..., N_t do
            Extract 2D patches from x_{i,n} using P
            Train D_{θ_{n+1}}(·) using the loss function in (9)
            Compute z_{i,n+1} = D_{θ_{n+1}}(x_{i,n}).
            Compute x_{i,n+1} from z_{i,n+1} using (7).
        end for
    end for
end for
```

the Adam Optimizer for 500 epochs at a learning rate of 10^{-4}. The inference was performed on each 2D slice sequentially to obtain the 3D volume image. The DC block consists of the CG algorithm operating on the 3D image generated by the CNN. The CG algorithm is run for ten iterations. The ASTRA forward operator is distributed over the four GPUs. We empirically choose the regularization parameter $\beta = 5 \times 10^{-2}$ and keep it fixed across all the iterations during training.

We compare the proposed method with unshared weights against FDK [1], MBIR [3], [14], U-Net [15] and its shared weights version. Both U-Net and the proposed method are fed with FDK reconstruction on the sparse-view data. We also compare the iteration-by-iteration reconstruction quality of both the versions of the proposed method. MBIR has been implemented using pyMBIR [14] with ASTRA [16].

B. Performance and Run-time Comparisons

The reconstruction performances of various methods are shown in Fig. 3. The PSNR and SSIM values are reported for the zoomed patches. The proposed method with unshared weights outperforms all the other methods. The one with shared weights (Proposed-S) loses key details due to excessive smoothing. The single-stage U-Net reconstructions are slightly inferior to the proposed due to overfitting errors and lack of measurement information through DC. MBIR gets rid of streak artifacts but loses some details due to blurring.

We compare the reconstruction quality at various iterations of the proposed method in Fig. 4. Proposed-S is the method with shared weights and its performance drops with iterations due to aggressive denoising (over-smoothing) at subsequent iterations. The one with unshared weights show sharper edges with iterations due to CNNs weights being optimal for other iterations. The PSNR and SSIM metrics drop for Proposed-S with iterations while they improve for the unshared weights.

Fig. 5 displays the performance of U-Net and the proposed method with unshared weights on out-of-distribution data. We performed reconstructions on 50 views instead of 75 and found the proposed method to be outperforming U-Net visually as well as in terms of PSNR and SSIM metrics. The results validate the robustness of the proposed method on imaging conditions unseen by the network during training.

We report the average run-time of each method in Table I. MBIR requires approximately two hundred iterations and therefore it is the slowest. The proposed method requires only three outer iterations and significantly reduces run-time. U-Net is much faster compared to the other two due to the absence of CG step.

MBIR	U-Net	Proposed	FDK
1440	5	30	4

TABLE I
AVERAGE RUN-TIME REPORTED IN SECONDS FOR A VOLUME OF SIZE $450 \times 380 \times 380$ VOXELS FROM MEASUREMENTS CONTAINING 75 VIEWS EACH OF SIZE 962×768 PIXELS. THE PROPOSED METHOD IS SLOWER THAN THE SINGLE-STAGE DEEP-LEARNING METHOD (U-NET) BUT IS MORE THAN AN ORDER OF MAGNITUDE FASTER THAN MBIR.

IV. CONCLUSION

We introduced a learnt half-quadratic splitting-based algorithm for CBCT reconstruction from sparse measurements. The proposed algorithm is a hybrid approach that uses CNN as a regularizer and a physics-based data consistency term; offering a memory-efficient implementation for large scale problems. It is significantly faster than the traditional MBIR approach, requiring very few iterations and also reduces overfitting issues often dealt by DL methods such as U-Net. Experiments show that the proposed algorithm outperforms the state-of-the-art methods when validated on a publicly available CBCT data.

REFERENCES

[1] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Josa a*, vol. 1, no. 6, pp. 612–619, 1984.

[2] Z. Yu, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "Fast model-based x-ray ct reconstruction using spatially nonhomogeneous icd optimization," *IEEE Transactions on image processing*, vol. 20, no. 1, pp. 161–175, 2010.

[3] J.-B. Thibault, K. D. Sauer, C. A. Bouman, and J. Hsieh, "A three-dimensional statistical approach to improved image quality for multislice helical ct," *Medical physics*, vol. 34, no. 11, pp. 4526–4544, 2007.

[4] 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.

[5] A. Ziabari *et al.*, "Enabling rapid x-ray ct characterisation for additive manufacturing using cad models and deep learning-based reconstruction," *npj Computational Materials*, vol. 9, no. 1, p. 91, 2023.

[6] O. Rahman *et al.*, "Deep learning based workflow for accelerated industrial x-ray computed tomography," in *2023 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2023, pp. 2990–2994.

[7] J. Xiang, Y. Dong, and Y. Yang, "Fista-net: Learning a fast iterative shrinkage thresholding network for inverse problems in imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 5, pp. 1329–1339, 2021.

[8] T. Su *et al.*, "Generalized deep iterative reconstruction for sparse-view ct imaging," *Physics in Medicine & Biology*, vol. 67, no. 2, p. 025005, 2022.

[9] A. Lahiri, G. Maliakal, M. L. Klasky, J. A. Fessler, and S. Ravishankar, "Sparse-view cone beam ct reconstruction using data-consistent supervised and adversarial learning from scarce training data," *IEEE Transactions on Computational Imaging*, vol. 9, pp. 13–28, 2023.

[10] H. Der Sarkissian *et al.*, "A cone-beam x-ray computed tomography data collection designed for machine learning," *Scientific data*, vol. 6, no. 1, p. 215, 2019.

[11] D. Kim and J. A. Fessler, "Generalizing the optimized gradient method for smooth convex minimization," *SIAM Journal on Optimization*, vol. 28, no. 2, pp. 1920–1950, 2018. [Online]. Available: https://doi.org/10.1137/17M112124X

[12] 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*, 2013, pp. 945–948.

[13] W. Van Aarle *et al.*, "The astra toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.

[14] S. V. Venkatakrishnan, "pyMBIR," 6 2019. [Online]. Available: https://www.osti.gov//servlets/purl/1550791

[15] 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*. Springer, 2015, pp. 234–241.

[16] W. Van Aarle *et al.*, "Fast and flexible x-ray tomography using the astra toolbox," *Optics express*, vol. 24, no. 22, pp. 25 129–25 147, 2016.

Joining photogrammetry and X-ray computed tomography

Nico Heinz, Andreas Maier, Christopher Syben and Richard Schielein

***Abstract*—Photogrammetry and X-ray computed tomography are two methods to obtain a three-dimensional model of a scanned object. While photogrammetry only yields a textured description of the object's surface, X-ray computed tomography can provide insights into the internal structure. We propose to combine both modalities during reconstruction in order to improve the reconstruction quality of both modalities. This is achieved by incorporating a consistency assumption which describes the relation between both reconstructions and integrate the consistency assumption into a joint optimization objective within a Bayesian framework. On synthetic datasets we successfully demonstrate experimentally that the combination of photogrammetry and X-ray computed tomography can to enhance overall reconstruction quality.**

***Index Terms*—CBCT, Multi-modal reconstruction, Photogrammetry**

I. Introduction

WHILE X-ray CT provides information about the internal structure of an object, photogrammetry can be used to obtain a detailed and textured description of the object's surface. The combination of the information from both modalities consequently provides a more complete description of the scanned object. As an exemplary previous work, [1] combines both modalities to capture antique glasswork for the preservation of cultural heritage. This motive is also found in [8], who additionally add endoscopy to preserve ancient gyroscopes. Another combination of X-ray CT and photogrammetry is the scanning of natural objects, such as insects and plants [2]. These previous works combine both modalities in order to obtain a more complete description of the scanned object by exploiting each modality's advantages. Consequently, the merge of the information from both modalities is done as a post-processing step after reconstructing the object's geometry, most commonly from X-ray cone-beam CT (CBCT).

In contrast, the focus of our work lies within improving the reconstruction quality of each modality by incorporating information from the other modality within the reconstruction algorithm. To our best knowledge, this is the first work to propose a reconstruction framework for this task.

A. Photogrammetry and Neural Radiance Fields

Reconstructing the three-dimensional geometry and texture of an object's surface from registered photographs is known as the multi-view stereo (MVS) problem. Recently, many researchers focused on Image Based Rendering (IBR) to solve the MVS problem, which employs an iterative optimization scheme to generate simulated perspectives from a virtual scene, parameterized by its parameters θ^P and alters the parameters θ^P to minimize the disparity J^{RGB}, commonly expressed as the mean squared error, between the synthesized rendered pixel colors $\hat{C}$ and the corresponding observed pixel colors C in the photographs.

$$\hat{\theta}^P = \mathrm{argmin}_{\theta^P} J^{\mathrm{RGB}}(\hat{C}(\theta^P), C) \tag{1}$$

Notably, Neural Radiance Fields (NeRFs) by Mildenhall et al. have emerged as a prominent IBR method, offering differentiable rendering capable of handling complex lighting effects and occlusions [3]. NeRFs describe a scene using a semi-transparent radiance field f^R which describes the local color $\mathbf{c}$ and volume density σ for every spatial position $\mathbf{p}$ and viewing direction $\mathbf{d}$ in the scene. The first step in rendering is executed by tracing rays $\mathbf{p}(t)$ from the camera's focal spot along the direction $\mathbf{d}$ through each pixel of the to-be-synthesized image and sampling spatial locations $\mathbf{p}_i$ along each ray. The second step of rendering consists of querying the model for the density σ and color $\mathbf{c}$ at the spatial positions $\mathbf{p}_i$. With these estimates, the rendering equation

$$\hat{\mathbf{C}} = \int_{t_1}^{t_2} T(t) \cdot \sigma(\mathbf{p}(t)) \cdot \mathbf{c}(\mathbf{p}(t), \mathbf{d}) dt \tag{2}$$

can be integrated numerically, where t_1 and t_2 denote the nearest and furthest distance evaluated from the ray origin. $T(t)$ is the accumulated transmittance, which describes the probability that the point $\mathbf{p}(t)$ is visible from the ray origin:

$$T(t) = \exp\left(-\int_{t_1}^{t} \sigma(\mathbf{p}(u)) du\right) \tag{3}$$

An extension of the NeRF model is the VolSDF model [5], which explicitly distinct in between the object's interior and exterior and induces a bias towards learning smooth geometries. A main drawback of the NeRF model are the long training times, which can last for several days. Multiresolution hash-grids reduce the training time to only seconds while achieving similar quality [4].

B. X-ray CT and Neural attenuation fields

X-ray computed tomography employs X-rays to generate detailed three-dimensional reconstructions of the internal structure of the scanned object. According to the Bee-Lambert

N. Heinz was with the Pattern Recognition Lab (LME) at Friedrich Alexander Universität Erlangen-Nürnberg (FAU) and Fraunhofer IIS. heinz.nico@mailbox.org

A. Maier is with the LME at FAU and Fraunhofer IIS

C. Syben and R. Schielein were with Fraunhofer IIS

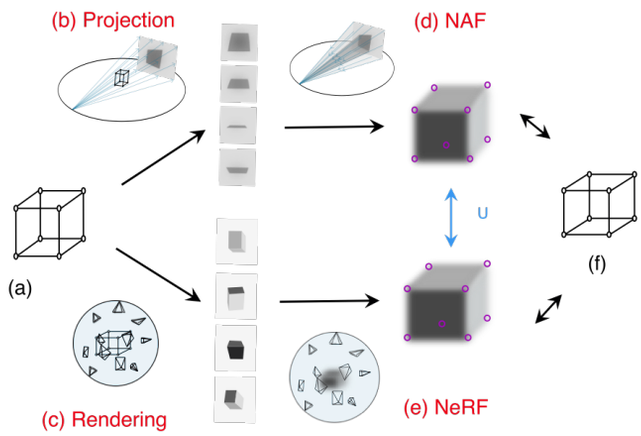


Fig. 1: The processing pipeline: A mesh (a) defines the synthetic phantom, which is used to (b) simulate X-ray projection images and (c) to render photographs of the phantom. During the tomographic reconstruction of the phantom from the projection images using Neural attenuation fields (NAFs) (d) and during the reconstruction of the phantom from photographs using Neural Radiance Fields (NeRFs) (e), both models exchange information via a regularization term U to induce bimodal consistency. A quantitative quality assessment (f) using various image quality metrics experimentally indicates the positive effect of the induced bimodal consistency assumptions on the final reconstruction quality of both modalities.

law, X-ray beams become attenuated to varying degrees depending upon the spatially varying attenuation coefficient $\mu(\mathbf{p})$ of the materials they encounter when they pass through an object. The negative of the logarithm of the ratio of the intensity I of the X-ray beam after its traversal through the object along the beam path S and the initial beam intensity I_0 is a linear projection of the attenuation coefficient along the ray path:

$$-\log(I/I_0) = \int_0^{t_D} \mu(\mathbf{p}(t))dt \tag{4}$$

with the source-detector distance t_D. Consequently, the NeRF framework can be easily adapted for X-ray CT reconstruction by replacing the NeRF rendering equation (2) with equation (4), with the model f^X describing the estimated local attenuation coefficient $\hat{\mu}(\mathbf{p}) = f^X(\mathbf{p}; \theta^X)$, so the framework is called Neural attenuation fields (NAFs). [7]

II. Methods

Our primary contribution involves developing and evaluating a bimodal reconstruction framework that enhances overall reconstruction results by concurrently reconstructing an object using X-ray CBCT using Neural Attenuation Fields and photogrammetry using Neural Radiance Fields while incorporating bimodal consistency assumptions through a regularization term U. The proposed algorithm is assessed on synthetic data, comparing bimodal reconstruction results with those achieved by standalone reconstruction using common quality assessment metrics. The entire processing pipeline from the creation of the phantom to the final evaluation is illustrated in Figure 1. We utilize the SDFStudio Python package [6] for implementing the bimodal reconstruction algorithm and evaluating results.

A. Bi-Modal consistency

We propose to induce a bimodal consistency assumption via a prior regularization term U_{bimodal} in a joint training objective. Then the simultaneous optimization of both the photogrammetry model and the X-ray model with respect to a joint objective J^{Joint} link the attenuation field $\hat{\mu}$ and the density field σ to each other:

$$J^{\text{Joint}}(\theta_P, \theta_X) = J^{\text{RGB}}(\theta_P) + \gamma_X J^X(\theta_X) + \gamma_U U(\theta_P, \theta_X) \tag{5}$$

If X-rays become attenuated inside an object, then both the tomographic model and the photogrammetric model agree whether some position belongs to the object or is empty, leading to the following assumption:

The classification of some position in space with respect to holding an object or being empty must be consistent across modalities.

In order to apply this assumption, both the optical density σ and the attenuation coefficients μ need a classification rule $h : \mathbb{R}_+ \to [0, 1], h(\mathbf{x}) = p_F(\mathbf{x})$ that assigns values of σ or μ to a corresponding probability $p_F(\cdot)$ of being object. $h(\cdot)$ should be monotone in its argument. The bimodal consistency assumption is incorporated into the reconstruction process by adjusting σ and $\hat{\mu}$ to satisfy $h_\sigma(\sigma) \approx h_\mu(\hat{\mu})$. To achieve this, the regularization term U in the optimization objective (5) penalizes differences between $h_\sigma(\sigma)$ and $h_\mu(\hat{\mu})$, promoting similarity or equivalence between the two classifications. A suitable regularization term, such as binary cross-entropy loss, can be employed for this purpose.

$$J^{\text{BCE}}(h(\mathbf{x}), h(\mathbf{y})) = -(h(\mathbf{y}) \log(h(\mathbf{x})) + (1 - h(\mathbf{y})) \log(1 - h(\mathbf{x}))) \tag{6}$$

We propose deriving a classification rule h by treating h as the prediction of a Gaussian Mixture Model (GMM). A GMM assumes that some probability distribution $p^{GMM}(\mathbf{x}|\theta^{\text{GMM}})$ is a mixture of K Gaussian components p_k^{GMM}, each being parameterized by its parameters θ_k consisting of the mean



(a) CiC photograph (b) CiC projection (c) Ficus photograph (d) Ficus projection

Fig. 2: Our datasets, containing photographs and projection images of a digital Cube in Cube (CiC) and ficus phantom.

vector $\mu_k \in \mathbb{R}^l$, covariance matrix $\Sigma_k \in \mathbb{R}^{l\times l}$ and mixing probability $\pi_k \in [0,1]$:

$$p^{\mathrm{GMM}}(\boldsymbol{x}|\theta_{\mathrm{GMM}}) = \sum_{k=1}^{K} p_k^{\mathrm{GMM}}(\boldsymbol{x}|\theta_k) \tag{7}$$

with

$$p_k^{\mathrm{GMM}}(\boldsymbol{x}|\theta_k) = \frac{\pi_k}{\sqrt{2\pi|\Sigma_k|}} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu}_k)^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_k)}{2}\right) \tag{8}$$

A maximum likelihood estimate of the parameters θ^{GMM} can be calculated using the expectation maximization algorithm.

The inherent ability to model two-dimensional inputs enhances information content, improving the discrimination between foreground and background classes. Consequently, the predictions of the two-dimensional $\mathrm{GMM}_{\sigma,\hat{\mu}}$ can be treated as a targets for the predictions of the one dimensional Gaussian Mixture Models GMM_{σ} and $\mathrm{GMM}_{\hat{\mu}}$.

Accordingly, a proper regularization term U_{GMM} is as follows:

$$\begin{aligned} U_{\mathrm{GMM}} = {} & J_{\mathrm{BCE}}(p_0^{\mathrm{GMM}}(\sigma|\theta_{GMM_\sigma}), p_0^{\mathrm{GMM}}(\sigma,\hat{\mu})|\theta_{GMM_{\sigma,\hat{\mu}}})) \\ & + J_{\mathrm{BCE}}(p_0^{\mathrm{GMM}}(\hat{\mu}|\theta_{\mathrm{GMM}_{\hat{\mu}}}), p_0^{\mathrm{GMM}}((\sigma,\hat{\mu})|\theta_{GMM_{\sigma,\hat{\mu}}})) \end{aligned} \tag{9}$$

We use the mean squared error for J^{RGB} and J^{X}.

B. Data generation and model structure

We create two different datasets for our experiments by simulating photos with Blender, an open-source 3D creation suite, and X-ray projection images using our proprietary software based on ray casting, considering frequency-dependent phenomena. The first dataset, Cube in Cube (CiC), comprises two cubes with distinct densities and orientations, presenting a challenge for reconstruction due to the visibility of the outer cube in photographs. The X-ray CT system acquires 16 cone-beam projection images along a circular trajectory, accompanied by 16 photographs with a high noise level. The second dataset represents a manipulated ficus plant for X-ray projection simulation, featuring 128 projection images and 32 photographs. Figure 2 provides exemplary views of the datasets. Note that the ficus phantom is the same as in previous works, e.g. by [3], but the set of photographs is different to increase sparsity.

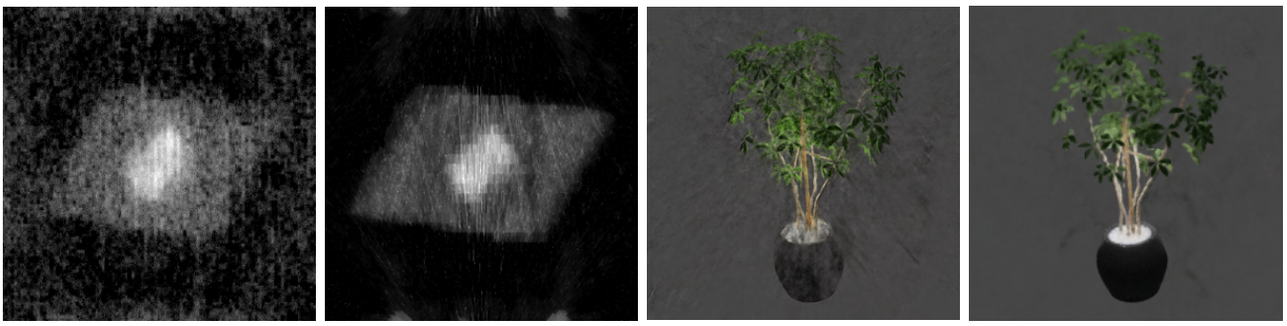
For the CiC dataset, we opt for the VolSDF model for photogrammetry, due to the simple geometry of the object. In contrast, the ficus phantom has a very complex geometry, which is why we use the plain NeRF model for photogrammetry. For both the NeRF models and the NAF models, we use multiresolution hash-encoding in order to accelerate training times.

The quantification of the effectiveness of the bimodal reconstruction approach in enhancing the quality of X-ray CT and photogrammetric reconstruction involves the calculation of image quality metrics for comparison with standalone reconstruction results. For photogrammetry assessment, this assessment covers the quality of rendering novel viewpoints not part of the training dataset by employing the LPIPS, PSNR and SSIM image metrics, following established procedures in the NeRF research community. In the context of X-ray CT, quality evaluation involves estimating the fidelity of the reconstructed volume in terms of the three-dimensional PSNR and SSIM, along with LPIPS for a representative cross-sectional image.

III. Results

For the X-ray reconstruction, figure 3 and table I reveal that U_{GMM} regularization significantly reduces the noise artifacts outside of the object for the CiC dataset. As a consequence, the LPIPS score improves from 0.66 to 0.57, the PSNR increases significantly from 24.7 to 29.7 and the SSIM confirms the improved model quality with an increase from 0.42 to 0.57. To our surprise, not only does the bimodal approach decrease the noise in the object's exterior, but also increases sharpness and contrast in the object's interior significantly, even though the photographs do not provide any information about the interior.

A sparse CBCT reconstruction is characterized by its underdetermined nature, resulting in the existence of multiple viable solutions. Within this framework, along a given ray path, the distribution of attenuation at certain spatial coordinates external to the imaged object may be offset by adjusting attenuation levels within the object itself, driven by the inherent underdeterminedness of the linear system. Conversely, the reconstruction of a different ray traversing through regions of reduced internal attenuation may counterbalance this by increasing attenuation somewhere



(a) Standalone (b) Bimodal (c) Standalone (d) Bimodal

Fig. 3: X-ray CT reconstruction for the CiC dataset and rendering of a novel view on the ficus dataset for standalone and bimodal models.

TABLE I: Evaluation scores of tomography (Tomo) and photogrammetry (photo) for standalone and bimodal reconstruction

Type	Dataset		LPIPS	PSNR	SSIM
Tomo	CiC	Standalone	0.66	24.7	0.42
		Bimodal	0.57	29.7	0.57
	Ficus	Standalone	0.22	25.6	0.95
		Bimodal	0.06	27.7	0.99
Photo	CiC	Standalone	0.02	33.1	0.95
		Bimodal	0.03	33.3	0.95
	Ficus	Standalone	0.11	28.0	0.87
		Bimodal	0.06	29.2	0.92

else along its path, potentially within the interior regions, not encountered along the former ray path. Consequently, the introduction of noisy or erroneously allocated attenuation outside the object can engender reconstruction artifacts within the object's interior due to this interplay. By enforcing the removal of reconstructed attenuation originating outside the object, the resultant reduction in noise levels within the object's interior is a direct consequence of this corrective action.

Bimodal regularization demonstrates substantial enhancement in photogrammetry as well. For the ficus dataset, LPIPS improves from 0.11 to 0.06. The mean PSNR across all evaluation images increases from 28.0 to 29.2, and SSIM increases from 0.87 to 0.92. The primary impact of the bimodal approach is the reduction of background artifacts originating from the NeRF model, addressing the optical density applied in empty space, a challenge known as shape-appearance ambiguity. However, a closer look at the image reveals that U_{GMM} regularization cannot remove the noise artifacts entirely. The underlying cause of this phenomenon can be attributed to the substantial magnitude of noise in the reconstruction, which leads to a misclassification of regions containing noise as objects, thereby outweighing the predictions made by the photogrammetric model.

IV. Summary

This study presents a novel bimodal reconstruction algorithm that combines photogrammetry and X-ray CT reconstruction by incorporating the consistency assumption that *The classification of some position in space with respect to holding an object or being empty must be consistent across modalities.*. We implement this assumption using Gaussian Mixture Models and a Bayesian reconstruction framework. Our findings demonstrate notable noise reduction and enhanced reconstruction quality in both X-ray CT and photogrammetry. These advancements hold promise for potentially lowering CT radiation dosage levels in future applications.

References

[1] Peter Fried et al. "3D scanning of antique glass by combining photography and computed tomography". In: *Digital Applications in Archaeology and Cultural Heritage* 18 (2020), e00147.

[2] Takashi Ijiri et al. "Digitization of natural objects with micro CT and photographs". In: *PLoS One* 13.4 (2018), e0195852.

[3] Ben Mildenhall et al. "Nerf: Representing scenes as neural radiance fields for view synthesis". In: *Communications of the ACM* 65.1 (2021), pp. 99–106.

[4] Thomas Müller et al. "Instant neural graphics primitives with a multiresolution hash encoding". In: *ACM Transactions on Graphics (ToG)* 41.4 (2022), pp. 1–15.

[5] Lior Yariv et al. "Volume rendering of neural implicit surfaces". In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 4805–4815.

[6] Zehao Yu et al. *SDFStudio: A Unified Framework for Surface Reconstruction*. 2022. URL: https://github.com/autonomousvision/sdfstudio.

[7] 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.

[8] K Zhan, D Fritsch, and JF Wagner. "Integration of photogrammetry, computed tomography and endoscopy for gyroscope 3D digitization". In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 46 (2021), pp. 925–931.

A 2D Sinogram-Based Approach to Defect Localization in Computed Tomography

Yuzhong Zhou, Linda-Sophie Schneider, Fuxin Fan, Andreas Maier

***Abstract*—The rise of deep learning has introduced a transformative era in the field of image processing, particularly in the context of computed tomography. Deep learning has made a significant contribution to the field of industrial Computed Tomography. However, many defect detection algorithms are applied directly to the reconstructed domain, often disregarding the raw sensor data. This paper shifts the focus to the use of sinograms. Within this framework, we present a comprehensive three-step deep learning algorithm, designed to identify and analyze defects within objects without resorting to image reconstruction. These three steps are defect segmentation, mask isolation, and defect analysis. We use a U-Net-based architecture for defect segmentation. Our method achieves the Intersection over Union of 93.92% on our simulated data, with an average position error of 1.3 pixels for defect detection on a 512-pixel-wide detector.**

***Index Terms*—Computed Tomography, image segmentation, defect localization, sinogram**

I. Introduction

COMPUTED Tomography (CT) is a remarkable technology that produces highly detailed cross-sectional images that provide valuable insights into the internal structures of objects, such as industrial components or the human body in the medical field. These images are essential for detecting abnormalities, which can range from medical conditions such as tumors and fractures in the human body to imperfections or defects in industrial products.

Detecting defects in CT images is of key importance in a variety of applications. In the medical field, early detection of medical conditions can be a determining factor for successful treatment and patient outcomes. Deep learning, a subset of machine learning, has emerged as a transformative force in CT image analysis. Much research has shown that convolutional neural networks can significantly improve the performance of defect detection [1], [2], especially semantic segmentation networks such as fully convolutional neural networks and U-Net.

Most research about defect detection focuses on image processing operations performed on reconstructed domains. Therefore, regardless of how efficient these algorithms are, they must wait for the completion of the reconstruction before they can start. In particular, prominent reconstruction algorithms such as iterative reconstruction (IR) introduce significant computational overhead. As shown in previous work [6], the IR algorithm, called real space iterative reconstruction, requires approximately 30 seconds to reconstruct a 3D volume measuring $243 \times 243 \times 243$. Consequently, the time required for defect detection in this pipeline is at least 30 seconds.

These studies based on the reconstructed domain overlook the valuable information in the raw sensor data, represented as sinograms—X-ray attenuation measurements taken from different angles. Traditional image reconstruction from sinograms, while standard, has drawbacks such as loss of detail, the introduction of artifacts, and high computational demands. Instead of using the reconstructed images, analyzing CT data directly in the sinogram domain should provide more accurate and consistent information by avoiding the information loss [8].

This research focuses on defects that have high attenuation. We exploits the potential of sinograms in defect detection by proposing direct analysis in sinogram space. This could allow simultaneous defect detection and reconstruction, or even skip reconstruction when unnecessary. This work aims to develop a methodology for defect detection and localization in parallel beam computed tomography using a U-Net-based architecture tailored for the analysis of CT sinograms. Our approach, which is divided into sinogram segmentation, instance segmentation, and defect analysis, can detect defects directly within sinograms, bypassing the need for reconstruction. This method detects defects efficiently and can be integrated with reconstruction processes in parallel beam CT. With this approach, we have not only improved defect detection in computed tomography, but also significantly reduced the computational resources required, paving the way for more efficient and accurate defect detection.

II. Methods

A. Sinogram Segmentation

Sinogram segmentation, which is a semantic segmentation task, is initially addressed using a U-Net model. This model takes the sinogram of the object (Fig. 1(a)) and processes it through 3 encoder layers for downsampling and feature extraction, and 3 decoder layers for upsampling and classification of pixels. Starting with 32 feature channels, each layer of the U-Net doubles the channels of the previous layer. This U-Net performs a binary classification of the sinogram pixels, identifying abnormal pixels as foreground and others as speciments. The goal is to isolate defect information in the sinogram mask (Fig. 1(b)). Our architecture is rigorously trained, with continuous validation using a separate dataset to monitor metrics such as Intersection over Union (IoU), precision, recall, and F1-score.

All authors are with the Pattern Recognition Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

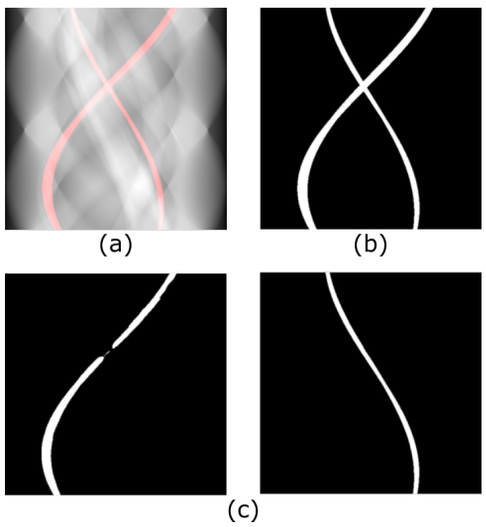


Fig. 1. (a) Input of the U-Net, pixels containing information of defects are painted red; (b) Result of sinogram segmentation is a binary mask that contains information about all found defects; (c) Result of instance segmentation, the information in (b) is separated based on which defect it belongs to.

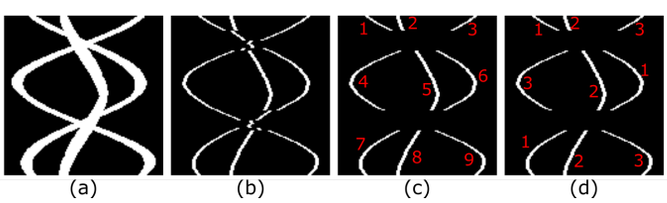


Fig. 2. (a) Result of semantic segmentation, which is also the input of the instance segmentation algorithm; (b) Raw result of skeletonization. After erasing the intersection part we get the final skeletonization result (c). (d) is the result of reclassification. For ease of presentation, the results of the skeletonization in (b), (c), and (d) have been widened.

B. Instance Segmentation

Sinogram segmentation produces a binary mask that contains all defects. To analyze these defects individually, we implement instance segmentation, which, unlike semantic segmentation, distinguishes separate defects within the same mask (Fig. 1(c)). In Fig. 1(b), the distinct positioning of two defects in the volume leads to separate projections in the sinogram. Using the Radon transform, identifying the central sinusoid of each defect allows them to be separated. This is achieved through skeletonization and reclassification techniques.

1) Skeletonization: we address the clustering of defect projections by adapting the Zhang-Suen algorithm [9] to the characteristics of the Radon transform. In each row of the sinogram, we reduce every non-zero region to its central point, producing an initial skeletonization result (Fig. 2(b)). Next, intersections are removed, resulting in distinct paths (Fig. 2(c)), each path containing information about a specific defect.

2) Reclassification: In Fig. 1(c), the original mask is thinned out, with the remaining pixels representing the centers of each defect's projection in a given direction. These pixels are crucial for computing the sinusoids needed to differentiate the defects. Given our projection's known frequencies and offsets, we only need to determine amplitude and phase. We choose not to use the Hough transform directly for parameter extraction; instead, we first categorize these pixels based on Radon transform characteristics. Pixels in the same group contribute to the calculation of an amplitude-phase pair, which is then fed into the Hough transform to derive the necessary parameters. This process is detailed in Algorithm 1.

Algorithm 1 Reclassification

Require: Result of skeletonization M_s
Ensure: Parameters for separation
  Initialize M_l as a zero matrix
  Initialize working row r to zero
  Initialize start of label s to 1
  Initialize end of label e to 1
  while r is less than the height of M_s **do**
    idx = indices of nonzero elements in $M_s[r]$
    if idx is empty **then**
      $s = e$
    else
      $e = s$ + size of idx
      $M_l[r, idx]$ = range from s to e
    end if
  end while
  Apply Hough Transform for sinusoid on labeled paths
  Re-label based on the result of Hough Transform

In the final step, the skeletons isolated with different labels as shown in Fig. 2(d) are merged with the sinogram mask to create distinct sinogram masks for each defect. We use the stored radius of non-zero regions to dilate the skeleton accordingly.

C. Defect Analysis

When X-rays are projected onto a detector along a single beam direction, the resulting projection width can be used to define an area encompassing the object. In two dimensions, performing a similar operation in a different direction produces a different area, the intersection of which indicates the object's position. This intersection forms the minimum enclosing parallelogram of the object, and the center is found at the intersection of the midlines of these two projections. These midlines are what we get from the reclassification, which gives the coordinates of the center. The lengths of these projections allow the overlapping area to be calculated and the size of the defect to be estimated.

However, it's important to note that with different pairs of projections, we can have multiple enclosing parallelograms, possibly with different centers. For non-circular defects, one approach is to pair projections and average the results, but this increases computational complexity. Alternatively, the computation can focus on the shortest projection and its perpendicular counterpart. For circular defects, the centerlines typically intersect at the center of the circle. Here, the properties of the Radon transform indicate that the center aligns with a sine wave created by the projections of the center points of the sinogram. Once the sinogram mask for the defect is obtained, this sine wave can be calculated to determine the center coordinates of the circle, as shown in Figure 3.

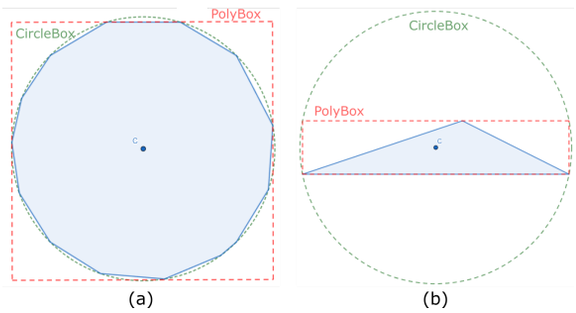


Fig. 3. Appropriate strategies vary with defect shape. (a) For circular defects, the CircleBox method outperforms the overlap method of two different projections in terms of accuracy. (b) For non-circular defects, the CircleBox approach is inappropriate.

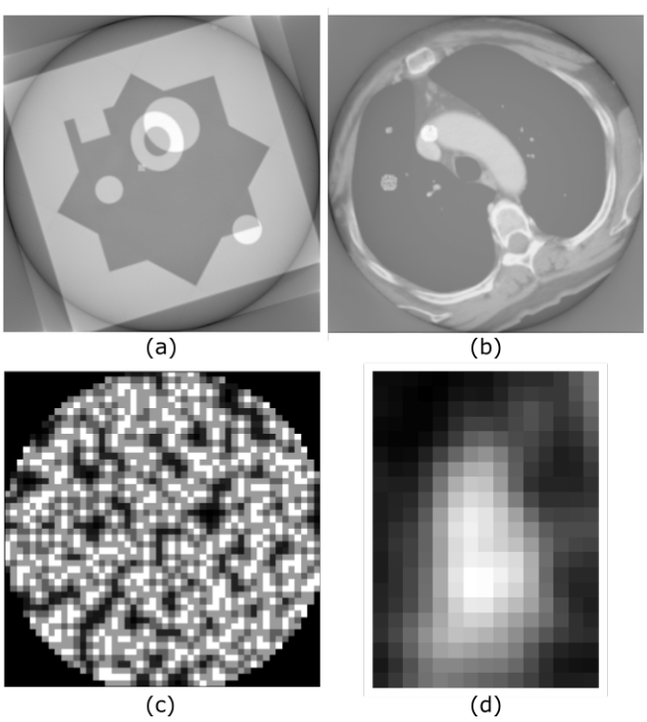


Fig. 4. (a) Sample from the custom dataset. (b) Sample from DeepLesion dataset. (c) Sample of a defect used in this paper which has a radius of 10 pixels. (d) Sample of a lesion extract from the DeepLesion datasets

III. Experiment

The U-Net architecture was developed using Pytorch and Pytorch Lightning. Network parameters were randomly initialized, followed by training using backpropagation and the Adam optimizer, starting with a learning rate of 0.001. The system was operated on Windows 10, using an NVIDIA 3060 graphics card.

We used two different datasets with two different types of volumes as the speciments: a custom MagicCube and the DeepLesion [13] (see Fig. 4). The custom MagicCube contains various shapes and forms. The Deeplesion are reconstructed patient iamges. For this study's focus on 2D parallel ray projections, we applied this technique to each slice of both volume types, creating simulated 2D sinogram datasets.

For the custom MagicCube, simulated sinograms were generated using the PYRO-NN toolkit [14] for 2D parallel ray projection. The sinograms covered a rotation range from 0 to 180 degrees, with a total of 512 sinograms and a 1D detector size of 512 pixels. The dimensions of the MagicCube are $512 \times 512 \times 512$, which corresponds to 512 slices of 2D phantoms, each 512×512 in size. We allocated 80% of these slices for training, and the remainder for validation and testing.

For data augmentation, one defect was introduced in each 2D phantom, and the training phantoms were rotated between 0 and 90 degrees in 1-degree increments. This process resulted in 40,500 sinograms for training, each containing a single defect, and 5,580 for testing, each containing up to three defects.

For defect simulation, we manually introduced two types of defects: small patches of circular or square Gaussian noise, as shown in 4(c), the radius of these patches is randomly chosen from 8 to 30 pixels. The other type of defects is the lesion patterns extract from the DeepLesion dataset, as shown in 4(d). During training data generation, the original images were randomly rotated, and the defects were randomly positioned within these images to ensure no overlap in the case of multiple targets. Each training set used only one defect type. Because our defects were artificially added, generating label masks was straightforward. We simply calculated the difference between the original object and the object with the added defects to generate the label masks.

We evaluated our pipeline using different metrics tailored to each step. For sinogram segmentation, the impact of false positives on subsequent instance segmentation algorithms was significant. To address this, IoU and precision were chosen as the primary metrics.

Additionally, recall and F1 score were considered as supplementary metrics. When IoU values were comparable, recall was useful for identifying more true positives, and F1-score assisted in balancing precision and recall at the pixel level.

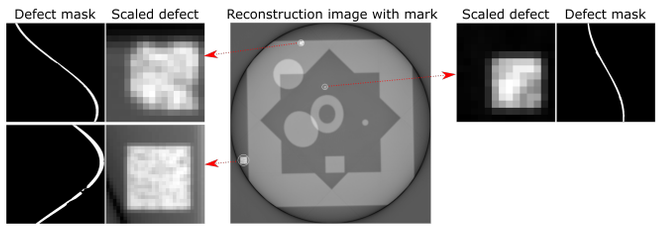


Fig. 5. Our method can detect and analyze the defects before reconstruction and facilitate defect identification in reconstructed domain.

IV. Results

We first demonstrate the practical effectiveness of our method for extracting defect information from sinograms, as shown in Fig. 5. This figure effectively demonstrates the successful isolation of three different defects within a sinogram. Furthermore, by analyzing the sinogram mask individually, we can determine the positions and sizes of these defects, which will aid in their identification in the reconstructed domain, and importantly, this analysis is performed prior to image reconstruction. The results in Fig. 5 confirm the achievement of our goal.

Since our method involves three steps, each of which not only affects the others, but also provides opportunities for isolated evaluation, we have performed tests on the entire process as well as on each individual step. The results of these tests are detailed in Table I and Table II.

TABLE I
SINOGRAM SEGMENTATION METRICS

Dataset	IoU	Recall	F1	Precision
MagicCube+Gauss	0.9392	0.9396	0.9683	0.9995
DeepLesion+Gauss	0.8571	0.8693	0.9267	0.9996
MagicCube+Lesion	0.7476	0.7499	0.8495	0.9950

As shown in the results, our method shows a remarkably high average IoU of 0.9352 on the MagicCube-derived sinograms. This high IoU value indicates the efficiency of our method in accurately isolating the defect-containing portions from the sinogram data. But we can observe a decrease in results, with IoU dropping to 0.7476 when we used the Lesion patterns as the simulated defects.

TABLE II
INSTANCE SEGMENTATION METRICS

Dataset	Correct rate	time (seconds)	Distance relative error	Area relative error
MagicCube	98.27%	0.033	0.0055	0.0197
DeepLesion	95.93%	0.052	0.0054	0.0139

In the following instance segmentation step, we introduced a metric named segmentation rate, which means the percentage of correct instance segmentation. We achieve an accurate segmentation rate of 98.27% in 0.033 seconds. Errors are mainly due to defect overlap or inaccuracies in the results of the previous step. For defect localization, we use the relative error to show the accuracy, the relative error is the ratio of the pixel error to the width of the detector. The average relative error is 0.0055, implying that on a 512-pixel wide detector, our method computes an average error of about 1.3 pixels for defect positions. The average relative error for the radius is 0.0197.

On the DeepLesion dataset, the performance is slightly worse, with an average IoU of only 0.8571 and an accurate segmentation rate of 95.93% in 0.052 seconds. This is mainly due to the more complex pixel distribution in medical images. The most severe impact comes from the first step of semantic segmentation, as subsequent steps rely on the previous step's results. Consequently, there is a slight decrease in the overall performance.

V. DISCUSSION AND CONCLUSION

This work presents an innovative approach to defect detection and localization in parallel beam computed tomography using a specialized U-Net-based architecture tailored for the analysis of CT sinograms. This methodology consists of sinogram segmentation, instance segmentation, and defect analysis, allowing for direct detection of defects within sinograms, and eliminating the need for image reconstruction.

Our method not only efficiently detects defects, but also offers the potential for integration with reconstruction processes in parallel beam CT. As a result, we have shown the average IoU of our method can be 0.9392, with an average position error of 1.3 pixels for defect detection on a 512-pixel-wide detector.

Experimental results demonstrate the effectiveness of our approach, with high IoU values in sinogram segmentation and accurate defect localization even in complex scenarios. While the performance on the DeepLesion dataset showed a slight decrease due to the challenging nature of medical images, the overall impact of our method on the efficiency and accuracy of defect detection is still remarkable.

Currently, our method is tailored for 2D parallel ray projection analysis, especially for the defects with different attenuation like metal or air bubble. Due to the limitations of our current experimental setup, we have used only simulated projection data and have omitted physical projection testing. Our primary focus for future research is to extend our methodology from 2D to 3D analysis and to transition from parallel to cone beam projection scenarios.

REFERENCES

[1] M. Usman, B.-D. Lee, S.-S. Byon, S.-H. Kim, B.-i. Lee, and Y.-G. Shin, "Volumetric lung nodule segmentation using adaptive roi with multi-view residual learning," *Scientific Reports*, vol. 10, no. 1, p. 12839, 2020.

[2] X. Dong, S. Xu, Y. Liu, A. Wang, M. I. Saripan, L. Li, X. Zhang, and L. Lu, "Multi-view secondary input collaborative deep learning for lung nodule 3d segmentation," *Cancer imaging*, vol. 20, pp. 1–13, 2020.

[3] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[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] G. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, ser. Advances in Computer Vision and Pattern Recognition. Springer London, 2009. [Online]. Available: https://books.google.de/books?id=BhtGTkEjkOQC

[6] M. Pham, Y. Yuan, A. Rana, S. Osher, and J. Miao, "Accurate real space iterative reconstruction (resire) algorithm for tomography," *Scientific Reports*, vol. 13, no. 1, p. 5624, 2023.

[7] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1289–1296, 2016.

[8] Q. De Man, E. Haneda, B. Claus, P. Fitzgerald, B. De Man, G. Qian, H. Shan, J. Min, M. Sabuncu, and G. Wang, "A two-dimensional feasibility study of deep learning-based feature detection and characterization directly from ct sinograms," *Medical physics*, vol. 46, no. 12, pp. e790–e800, 2019.

[9] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Communications of the ACM*, vol. 27, no. 3, pp. 236–239, 1984.

[10] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[11] W. A. Falcon, "Pytorch lightning," *GitHub*, vol. 3, 2019.

[12] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[13] K. Yan, X. Wang, L. Lu, and R. M. Summers, "Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning," *Journal of medical imaging*, vol. 5, no. 3, pp. 036 501–036 501, 2018.

[14] C. Syben, M. Michen, B. Stimpel, S. Seitz, S. Ploner, and A. K. Maier, "Pyro-nn: Python reconstruction operators in neural networks," *Medical physics*, vol. 46, no. 11, pp. 5110–5115, 2019.

Deep Learning Computed Tomography based on the Defrise and Clack Algorithm

Chengze Ye, Linda-Sophie Schneider, Yipeng Sun, and Andreas Maier

***Abstract*—This study presents a novel approach for reconstructing cone beam computed tomography (CBCT) for specific orbits using known operator learning. Unlike traditional methods, this technique employs a filtered backprojection type (FBP-type) algorithm, which integrates a unique, adaptive filtering process. This process involves a series of operations, including weightings, differentiations, the 2D Radon transform, and backprojection. The filter is designed for a specific orbit geometry and is obtained using a data-driven approach based on deep learning. The approach efficiently learns and optimizes the orbit-related component of the filter. The method has demonstrated its ability through experimentation by successfully learning parameters from circular orbit projection data. Subsequently, the optimized parameters are used to reconstruct images, resulting in outcomes that closely resemble the analytical solution. This demonstrates the potential of the method to learn appropriate parameters from any specific orbit projection data and achieve reconstruction. The algorithm has demonstrated improvement, particularly in enhancing reconstruction speed and reducing memory usage for handling specific orbit reconstruction. Code is available at https://github.com/ChengzeYe/Defrise-and-Clack-reconstruction.**

***Index Terms*—CT Reconstruction, Deep Learning, Known Operator, Specific Orbit.**

I. Introduction

COMPUTED tomography (CT) is a medical imaging technique that uses X-ray imaging to produce high-resolution sectional images of the human body. It is essential for medical diagnosis, surgical planning, and scientific research. CT technology is also crucial in the industrial sector, especially in non-destructive testing, allowing engineers to detect and analyze minor internal defects in materials, ensuring quality and safety.

Traditional CT scans usually involve a rotating structure with fixed orbits for the X-ray source and detector. In contrast, recent C-arm CT systems provide greater flexibility by allowing free movement of these components. Flexible CT orbits are especially advantageous in interventional imaging, where they help to avoid metal artifacts and optimize image acquisition speed and radiation dose [1]. In the industrial sector, this approach speeds up inspections, improving the efficiency of identifying and analyzing structural defects or anomalies. However, if the orbit of the X-ray source follows a specific non-circular path, this presents a challenge to the reconstruction process.

Previous research has proposed various methods for reconstructing specific orbits. For example, Zeng [2] described the algebraic iterative reconstruction algorithm, which can reconstruct cone beam computed tomography (CBCT) projections of arbitrary orbits.

Grangeat [3] proposed a reconstruction method based on the relation between cone-beam data and a function associated with the 3D Radon transform of the images. This algorithm can accurately reconstruct any CT orbit. However, in practical applications, the algorithm requires the prior construction of a matrix containing samples of the intermediate function, which is then subjected to subsequent processing.

The reconstruction algorithm mentioned above can successfully reconstruct any given orbit. However, it has a slow reconstruction speed and high memory consumption. To overcome this limitation, Defrise and Clack [4] introduced a filtered backprojection type (FBP-type) algorithm based on the Grangeat method. The algorithm is customized for different orbits by designing specific redundancy weights, ultimately achieving analytic reconstruction for any orbit data.

Calculating specific redundancy weights for different orbits in practical applications can be challenging. However, this problem can be approximated and solved through deep learning techniques. While deep learning is adept at modeling complex functions, it may lead to overly complex models with excessive parameters when dealing with intricate problems such as the inverse problem in CT reconstruction.

Maier *et al.* [5] proposed a method to incorporate known operators into machine learning algorithms. This approach significantly reduces the number of neural network parameters, enabling faster training and reducing maximum error bounds.

Syben *et al.* [6] proposed a neural network based approach that utilizes known operators in the filtered backprojection (FBP) framework. It demonstrates the feasibility and effectiveness of optimizing parameters in the specific layer of the reconstruction pipeline.

We present a data-driven methodology for reconstructing CBCT projections for a specific orbit. The method belongs to the category of FBP algorithms and incorporates a sequence of elementary 2D operations as a filter, including weightings, differentiations, 2D projection, and 2D backprojection. A reconstruction framework is established using known operators, where only the redundancy weight associated with the CT orbit are treated as trainable parameters. They are obtained through a training process, where parameters are fitted based on a given orbit, ultimately enabling rapid reconstruction of specific orbit CBCT scans.

All authors are with the Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg, Erlangen, Germany (e-mail: chengze.ye@fau.de)

II. Methods

The principle of CT is based on measuring the absorption of X-rays at different angles. Algebraic algorithms are then used to transform this projection data into an image of the object's interior.

A. Filtered Backprojection

Algebraic reconstruction involves solving $p = Ax$ to recover image or volume x from projection data p. The relationship between projection and image domain is captured by matrix A. However, computing the inverse of A directly is challenging. To address this, the filtered backprojection algorithm is used [2].

To reconstruct the original object from its projection data $p(s, \theta)$ in a parallel beam geometry, a convolution with a Ramp filter $h(s)$ should be first performed. The goal of this process is to filter the projection data in order to enhance high-frequency components while suppressing lower-frequency components.

$$q(s, \theta) = h(s) * p(s, \theta) \tag{1}$$

$$h(s) = \int |\omega| e^{2\pi\omega s} d\omega \tag{2}$$

The filtered data $q(s, \theta)$ is then subjected to a backprojection operation. This step maps the filtered data back into the image space from different angles.

$$f(x, y) = \int_0^{\pi} q(s, \theta) \Big|_{s = x \cdot \cos\theta + y \cdot \sin\theta} d\theta \tag{3}$$

Finally, the reconstructed image or volume $f(x, y)$ can be obtained through this comprehensive procedure.

When using cone beam geometry for projection data, it is necessary to introduce cosine weight to correct the geometry.

B. Grangeat's Inversion

Grangeat [3] introduced a reconstruction method that established a connection between cone-beam projections and the first derivative of the Radon transform by identifying an intermediate function. This method achieves rebinning from the Cartesian coordinate system of cone-beam geometry to the spherical coordinate system of the Radon domain using a plane as an information vector. Zeng [2] describes a process where numerous lines are sampled on a two-dimensional cone-beam detector. The projection is integrated along these lines, resulting in a line integral that can be viewed as a plane integral weighted by $\frac{1}{r}$. This plane is characterized by the distance $l \in [-B, +B]$ to the origin and a vector θ orthogonal to the plane. Furthermore, it has been observed that the derivative of a plane integral along the tangential direction dt is equivalent to the derivative of a $\frac{1}{r}$ weighted plane integral along $d\alpha$. Based on this observation, Grangeat's intermediate function (4) is derived.

$$S(\theta, \lambda) = -\int_{s^2} d\beta \delta'(\beta, \theta) g(\beta, \lambda) \qquad \theta \in S^2, \lambda \in \Lambda \tag{4}$$

Equation (4) uses $\lambda \in \Lambda$ as a parameter for the source position $a(\lambda)$ and $\beta \in S^2$ to indicate the direction of the line integral. The function $g(\beta, \lambda)$ represents the cone-beam projection, while δ' signifies the derivative of the Dirac delta distribution. By incorporating the intermediate function into the inverse Radon transform formula, one can derive the reconstruction formula (5) for Grangeat's method.

$$f(x) = -\frac{1}{4\pi^2} \int_{s^2/2} d\theta \int_{-B}^{B} dl \delta'(x \cdot \theta - l) S(\theta, \lambda(l, \theta)) \tag{5}$$

C. Defrise and Clack Algorithm

This algorithm is a shift-variant filtered backprojection algorithm derived from Grangeat's method. It can be used to reconstruct projection data of general non-circular orbits. Unlike the Grangeat algorithm, it eliminates the need to store all intermediate functions in a matrix. This enables the projection acquisition and processing to be executed synchronously, which greatly enhances the speed of the reconstruction process. The reconstruction formula is as follows:

$$f(x) = \int_{\Lambda} d\lambda \int_{S^2/2} d\theta - \frac{1}{4\pi^2} \mid a'(\lambda) \cdot \theta \mid \frac{1}{n(\theta, \lambda)} \times \delta'((x - a(\lambda)) \cdot \theta) S(\theta, \lambda). \tag{6}$$

The term $n(\theta, \lambda)$ is defined as the number of intersections between orbit $a(\Lambda)$ and the plane orthogonal to θ through orbit point $a(\lambda)$. The term $\frac{1}{n(\theta,\lambda)}$ and $\mid a'(\lambda) \cdot \theta \mid$ are closely related to the orbit's geometry and could be regarded as a redundancy weight. In the next section II-D, we will discuss the parameters obtained through the training of the neural network to approximate this discrete redundancy weight.

The algorithm comprises three main steps. Firstly, for each cone-beam projection $\lambda \in \Lambda$, Grangeat's intermediate function is computed using (4). Secondly, shift-variant filtering to the obtained intermediate function $S(\theta, \lambda)$ is applied:

$$g^F(\omega, \lambda) = \int_{S^2/2} d\theta - \frac{1}{4\pi^2} \mid a'(\lambda) \cdot \theta \mid \frac{1}{n(\theta, \lambda)} \delta'(\omega, \theta) S(\theta, \lambda) \qquad \omega \in S^2, \lambda \in \Lambda. \tag{7}$$

Finally, the filtered cone-beam projections are backprojected into the 3D volume.

$$f(x) = \int_{\Lambda} d\lambda \frac{1}{\mid x - a(\lambda) \mid^2} g^F \left(\frac{x - a(\lambda)}{\mid x - a(\lambda) \mid}, \lambda \right) \qquad x \in \mathbb{R}^3, \mid x \mid \leq B \tag{8}$$

D. Defrise and Clack Neural Network

The objective of this study is to create an end-to-end neural network framework based on Defrise and Clack's algorithm [4]. The framework's data-driven nature enables it to learn autonomously from projection data of any specific CT orbits. This involves determining the appropriate redundancy weight for a specific CT orbit to facilitate effective image reconstruction.

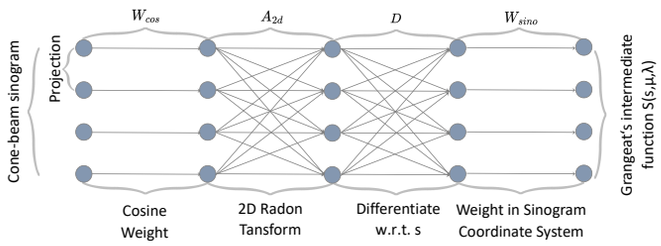


Fig. 1. Grangeat's intermediate function part of the neural network.

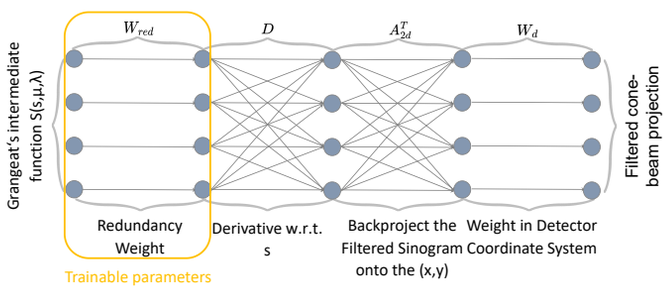


Fig. 2. Filtering part of the neural network.

1) Grangeat's intermediate function: A flat-panel detector is used to collect cone-beam projection data and each straight line in the detector is parametrized as usual by a radial coordinate s and μ. Equation (9) can then be derived from (4) [4].

$$S(s,\mu,\lambda) = \frac{s^2+D^2}{D^2}\int_{+e}^{-e} dv \frac{\partial}{\partial u}\left\{\frac{Dg(x,y,\lambda)}{\sqrt{u^2+v^2+D^2}}\right\}_{u=s} \tag{9}$$

In this equation, u and v represent coordinate transformations defined as $u = x\cos\mu + y\sin\mu$ and $v = -x\sin\mu + y\cos\mu$, while D denotes the distance between the source and the detector. In addition, the variable e represents the radius of the region defined by the cone-beam projection of the field of view.

According to (9), the computation of Grangeat's intermediate function can be broken down into three steps, each represented by one or two layers in a neural network. The first step applies cosine weight W_{cos} to the projection data. Next, we perform a 2D Radon transform A_{2d} using parallel beam geometry to re-project the weighted projection data. Finally, we differentiate D the sinogram obtained from the 2D Radon transform with respect to s and apply weight W_{sino} in the sinogram domain. By combining the previous steps, we formulated (10).

$$S(s,\mu,\lambda) = W_{sino}DA_{2d}W_{cos}p \tag{10}$$

This equation differs from (9) mainly in the order of integration and differentiation being switched. The purpose of this switch is to reduce the noise caused by the differentiation process. Using (10), we construct the neural network shown in Fig. 1.

2) Filter the cone-beam data: Equation (11) is derived from (7) by Defrise and Clack [4].

$$g^F(x,y,\lambda) = (x^2+y^2+D^2)\int_0^{\pi} d\mu \left\{\frac{\partial}{\partial s}\frac{S_1(s,\mu,\lambda)}{\sqrt{s^2+D^2}}\right\}$$
$$s = x\cos\mu + y\sin\mu$$
$$S_1(s,\mu,\lambda) = -\frac{1}{4\pi^2}\mid a'(\lambda)\cdot\theta\mid M(\theta,\lambda)S(\theta,\lambda) \tag{11}$$

A neural network is constructed based on (11). The redundancy weight W_{red} applied to the projection data is closely related to the orbital geometry and is set as learnable parameters. The weighted sinogram is then differentiated D with respect to s in the sinogram domain. After that, the filtered sinogram is backprojected A_{2d}^T into the detector coordinate system using the same parallel beam geometry. Finally, weight W_d is applied in the detector domain for geometric correction. Following these steps will result in (12).

$$g^F(x,y,\lambda) = W_dA_{2d}^TDW_{red}S(s,\mu,\lambda) \tag{12}$$

The filtering part of the neural network is illustrated in Fig. 2.

3) Backprojection to the 3D image: All of the above steps can be considered as shift-variant filtering. Afterward, the filtered cone beam projection data is back-projected into a 3D volume A_{3d}^T. The output layer of the neural network uses ReLU to enforce non-negativity. By combining (10) and (12), we obtain the complete reconstruction formula (13).

$$x = A_{3d}^TW_dA_{2d}^TDW_{red}W_{sino}DA_{2d}W_{cos}p \tag{13}$$

Based on this formula, the complete network architecture has been constructed, as shown in Fig. 3.

III. Experiments and Results

For our study, we selected the walnut dataset [7], an open collection of X-ray cone-beam CT datasets of forty-two walnuts. The dataset includes raw projection data, scanning geometry details, pre-processing, and reconstruction scripts, making it suitable for tasks such as image reconstruction. We generated 30 simulated data samples to obtain a suitable training and validation dataset. Twenty-four of the sets are used for training, while the remaining sets are used for validation. For each set, we randomly generated 5 to 10 geometric objects of different types, positions, and rotation directions within a blank voxel volume as the ground truth. We followed the same orbit geometry as the walnut dataset and used the forward projection of the software PyroNN [8] to generate the sinogram, which served as input for the neural network. A subset of 10 walnuts was used as the test dataset to evaluate the model's performance on real-world data.

Table I presents the specific parameters of the circular orbit geometry.

In our experiment, we used PyTorch 2.1.1 to construct the neural network's architecture. The 2D Radon transform and the 3D cone-beam backprojection were implemented using operators from the PyroNN library [8]. We chose Mean Squared Error (MSE) as the loss function. Additionally, we utilized the Adam optimizer with an initial learning rate of 1×10^{-5} and adjusted it within a specific range using the Onecycle

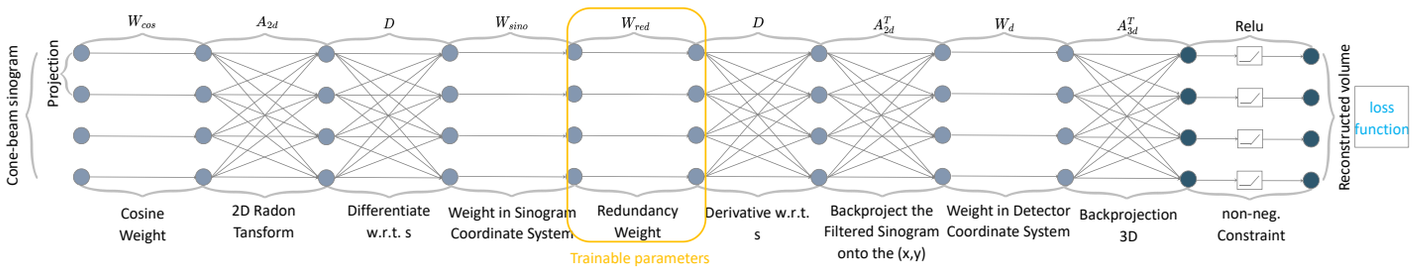
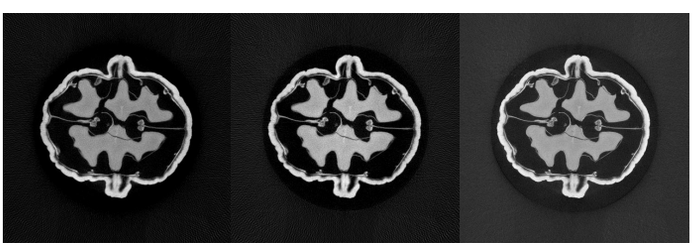


Fig. 3. Grangeat's intermediate function part, filtering part, and backprojection part are combined to form the Defrise and Clack neural network architecture.

TABLE I
PARAMETERS OF THE CIRCULAR ORBIT GEOMETRY

Volume shape	**Volume spacing**	**Number of projections**
501×501×501	0.1mm×0.1mm×0.1mm	400
Angular range	**Source isocenter distance**	**Source detector distance**
2π	66mm	199mm
Detector shape	**Detector spacing**	
972×768	0.1496mm×0.1496mm	



(a) (b) (c)

Fig. 4. Reconstructed results of the network. (a) Reconstruction using learned redundancy weight. (b) Reconstruction using analytic redundancy weight. (c) FDK reconstruction result.

learning rate policy, which enhanced the training process. Random initialization was also applied to the redundancy weight layer W_{red}. The model ultimately converged after undergoing training for 10 epochs on an Nvidia A40 GPU.

In our experiments, we observed unstable convergence of the network which introduced high-frequency noise. This adversely affected the reconstruction results. Therefore, we decided to apply Gaussian filtering to the learned parameters before using them for the reconstruction process.

In Fig. 4, a comparison of the central slice of volumes reconstructed with different redundancy weights is presented. The results show that the neural network, after training, can achieve reconstruction results similar to those obtained through analytical methods.

IV. Conclusion and Discussion

Our research proposes a reconstruction pipeline for specific CBCT orbits using known operators. We have demonstrated the effectiveness of a data-driven approach for optimizing parameters in this reconstruction pipeline. Our findings indicate that the Defrise and Clack neural network can learn parameters for reconstruction based on projection data with circular orbit geometry. This trained neural network performs well in reconstruction. For non-circular orbits, it is necessary to generate projection data specific to the given orbit geometry and subsequently optimize the parameters of the neural network to fit this orbit geometry. In summary, the Defrise and Clack neural network has the potential to reconstruct projection data with any specific orbit. These improvements have implications for medical and industrial imaging, representing a major advancement in C-arm CT imaging technology. Future research should focus on improving reconstructed image quality, optimizing this neural network model to further speed up reconstruction, and exploring further applications.

References

[1] L. S. Schneider, M. Thies, R. Schielein, C. Syben, M. Unberath, and A. K. Maier, "Learning-based Trajectory Optimization for a Twin Robotic CT System," 3rd ed., 2022.

[2] G. L. Zeng, "Medical image reconstruction," *Medical image reconstruction*, vol. 530, Springer, Heidelberg, 2010.

[3] P. Grangeat, "Mathematical framework of cone beam 3D reconstruction via the first derivative of the Radon transform," in *Mathematical Methods in Tomography: Proceedings of a Conference held in Oberwolfach, Germany, 5–11 June, 1990*, Springer Berlin Heidelberg, 1991.

[4] M. Defrise and R. Clack, "A cone-beam reconstruction algorithm using shift-variant filtering and cone-beam backprojection," *IEEE transactions on medical imaging*, vol. 13, no. 1, pp. 186–195, 1994.

[5] A. K. Maier, C. Syben, B. Stimpel, T. Würfl, M. Hoffmann, F. Schebesch, W. Fu, L. Mill, L. Kling, and S. Christiansen, "Learning with known operators reduces maximum error bounds," *Nat. Mach. Intell.*, vol. 1, no. 8, pp. 373–380, 2019.

[6] C. Syben, B. Stimpel, K. Breininger, T. Würfl, R. Fahrig, A. Dörfler, and A. Maier, "Precision learning: Reconstruction filter kernel discretization," *arXiv preprint arXiv:1710.06287*, 2017.

[7] H. Der Sarkissian, F. Lucka, M. van Eijnatten, G. Colacicco, S. B. Coban, and K. J. Batenburg, "A cone-beam X-ray computed tomography data collection designed for machine learning," *Scientific data*, vol. 6, no. 1, pp. 215, 2019.

[8] C. Syben, M. Michen, B. Stimpel, S. Seitz, S. Ploner, and A. K. Maier, "PYRO-NN: Python reconstruction operators in neural networks," *Med. Phys.*, vol. 46, no. 11, pp. 5110–5115, 2019.

Latent Diffusion Reconstruction for Stationary CT

A. Lopez-Montes, T. McSkimming, W. Zbijewski, H. Huang, A. Skeats, C. Delnooz, B. Gonzales, and A. Sisniega

***Abstract*— Generative diffusion models have shown great potential for synthesis of realistic images in unconditional and conditional paradigms, via iterative sampling of a reverse noise diffusion model with a denoising score matching learned operator. Sampling of the training data distribution can be leveraged for inverse problem solving using posterior sampling strategies conditioned on measured data. These methods are particularly suited for applications to CT reconstruction in heavily undersampled scenarios. However, sampling of the reverse diffusion path requires of a large number of steps, with a high computational burden.**

In recent years, latent diffusion models (LDM) have been proposed as an alternative to conventional image-based diffusion. In LDM, a variational autoencoder (VAE) is trained to produce low dimensionality latent representations of the images. Then, the sampling is performed in the latent space. Finally, the image associated to the sampled latent representation is retrieved from the latent domain by the decoding operation of the VAE. While the method can produce fast and accurate results for synthetic image generation, its application for inverse problems remains an open research field. Naïve latent DPS has been shown insufficient since the distance in the latent domain and the image domain is not preserved with conventional VAE designs. To alleviate that problem, in this work we introduced a distance preserving regularization strategy for training of the VAE that promotes consistency between small perturbations in the image and latent spaces. Moreover, we introduced a reconstruction strategy with a modified correction in Latent DPS. Our method was designed for reconstruction in head/brain imaging with stationary CT systems yielding heavily sparse and limited sampling of the CT volume. The approach was assessed in simulation studies including synthetic ICH. The method outperformed MBIR in terms of overall image quality (0.82 vs 0.76 average SSIM for 33 different cases) and bleed detection (0.58 vs 0.4 dice score recovery on bleed lesions). When compared to other strategies for solving inverse problems using LDM, the proposed approach achieved better out-of-domain generalization capacity, while relaxing the computational burden and memory requirements of image-based DPS reconstruction.

***Index Terms*—Diffusion models, Diffusion Posterior Sampling, Latent Diffusion, Stationary CT.**

I. INTRODUCTION

DIFFUSION models [1],[2] pose a powerful approach to generative modeling in multiple domains and are currently recognized as state-of-the-art for image generation tasks, including unconditional and conditional image synthesis. Beyond the potential for generation of synthetic image data, unconditional diffusion models offer a powerful platform for inverse problem solving, via integration of strategies to enforce consistency between the diffusion sampling path and a set of measurements. Within that paradigm, recent work presented diffusion-based approaches to CT image reconstruction, suitable to scenarios yielding sparse and/or limited sampled data [3][4], such as stationary CT systems based on arrays of x-ray source elements, recently proposed for imaging of intracranial hemorrhage (ICH) at the point-of-care [5][6].

Generative diffusion models are based on a forward model that corrupts the distribution of the training data, $p_0(\boldsymbol{x}_0)$, via a series of noise-addition stages acting on an individual element, x, of $\boldsymbol{x}_0$ until it is indistinguishable from pure noise.
The forward diffusion process is commonly represented via a stochastic differential equation (SDE) [1]:

$$dx = f(x,t)dt + g(t)dw \qquad (1)$$

where $f(x,t)$ and $g(t)$ are the so-called drift and diffusion terms that control the evolution of the image (x), with respect to diffusion time, t, and noise, w, respectively. Image synthesis is achieved via the associated reverse diffusion model that samples from $p_0(\boldsymbol{x}_0)$ given a noise distribution. This reverse diffusion can be modeled with a second SDE [1]:

$$dx = \left(f(x,t) - g(t)^2 \nabla_{x_t} log\, p(x_t)\right)dt + g(t)dw' \qquad (2)$$

where $\nabla_{x_t} log\, p(x_t)$ is the score of the reverse process, which is non-tractable and can be approximated via a denoising neural network $\left(\nabla_{x_t} log\, p(x_t) \approx s_\theta(x_t,t)\right)$ [1]. Training of the unconditional synthesis model amounts to learning the time-dependent denoising operator. Diffusion models can be applied to inverse problem solving via modified sampling strategies, such as Diffusion Posterior Sampling (DPS), in which the score term is modified to account for the posterior distribution $\nabla_{x_t} log\, p(x_t|y)$, where y are the measured data

Sampling of the reverse diffusion path often requires a large number of steps, resulting in high computational burden, particularly in inverse problem scenarios. To mitigate the computational burden associated to sampling, several strategies have been proposed for diffusion-based synthesis models, including faster SDE solvers [7]; consistency models with low number of sampling steps [8]; or latent diffusion models (LDM) [9], which have demonstrated state-of-the-art performance in unconditional and conditional image synthesis.

In LDM a variational autoencoder (VAE) consisting of a convolutional encoder $(\mathcal{E})$ and decoder $(\mathcal{D})$ is trained to learn a low-dimensionality latent representation (z) of the training data, x [9]. A generative model can then be obtained by building a diffusion SDE in the latent domain. Image synthesis is performed via reverse diffusion sampling in the latent space followed by decoding of the sampled latent, $\mathcal{D}(z_0)$. The LDM sampling process is depicted in Fig. 1.

Despite the success of LDM for image synthesis tasks,

A. Lopez Montes (e-mail alopezm2@jhmi.edu), T. McSkimming, W.Zbijewski, and A.Sisniega (e-mail asisnie1@jh.edu) are with Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, 21205 USA.
T. McSkimming, A. Skeats, C. Delnooz, B. Gonzales are with Micro-X Ltd, Adelaide, SA, Australia.
T. McSkimming is also with College of Science and Engineering, Flinders University, Adelaide, SA. Australia.

application to inverse problem solving is challenged by the difficulty of enforcing data consistency strategies directly in the latent domain and the lack of trivial correspondence between posterior sampling on the original data domain and on the latent space. To overcome this limitation, alternatives to DPS acting in the latent space have been proposed. One of such approaches, denoted Posterior Sampling with Latent Diffusion (PSLD) [10], used a "gluing" term to minimize the drift between the encoded compensation and the expected solution. Further approaches include the ReSample strategy [11] in which, a minimization problem enforcing consistency between the latent, z_t, and the data, y, is solved in a subset of the sampling steps and the result in used to "resample" the current latent-space solution.

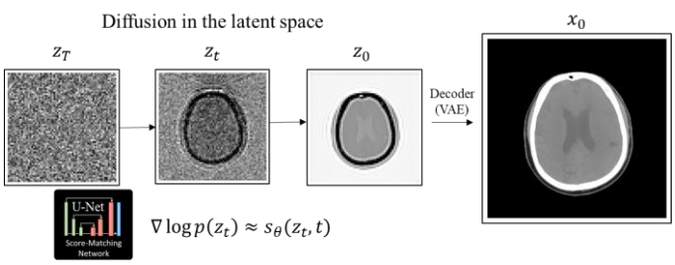


Figure 1. Schematic representation of the LDM reverse sampling process in the latent domain. The synthetic data in the original image space is constructed via decoding of the low-dimension latent tensor using a pretrained VAE.

In this work we propose a novel approach for posterior sampling with latent diffusion via: i) introduction of a distance-preserving autoencoder; and, ii) a novel strategy for latent DPS that combines data-consistent iterative updates with diffusion sampling on the latent space. The proposed strategy is assessed for reconstruction of stationary CT for imaging of intracranial hemorrhage, in comparison with alternative approaches (latent DPS, PSLD, and ReSample).

II. Materials and Methods

A. Distance preserving VAE

The VAE to obtain the latent-space representation of the input data was modeled following the architecture proposed in [9]. Training of the VAE was performed in an adversarial mode including Kullback-Leibler regularization term, to prevent large variations in the magnitude of the latent features.

To better precondition the relation between additive operations in the image and latent domain, the training loss included a regularization term to encourage distance preserving transformations between the pixel and the latent domains. Distance preservation was encouraged via modification of the input data with a small-sized random additive term, as $x_T = x + T(x)$. The operator T was modeled as a combination of: i) blurring with a gaussian kernel with standard deviation randomly sampled between $1 \cdot 10^{-6}$ and 5 pixels; ii) a random sharpening adjustment; iii) a scaling operation with a random scaling factor < 0.3; and, iv) addition of zero-mean gaussian noise with $\sigma = 1$. The distance-preserving loss was posed as:

$$L_{dist_pres} = \frac{1}{2}\left|\left|\mathcal{E}(x + T(x)) - \left(\mathcal{E}(x) + \mathcal{E}(T(x))\right)\right|\right|^2 \quad (3)$$

B. Latent Diffusion Posterior Sampling

In this work we used a Denoising Diffusion Probabilistic Model (DDPM) for equations (1) and (2), acting in the latent space (z). The variance of the noise distribution at time step t is denoted by β_t, yielding the following drift and diffusion:

$$f(z,t) = -\frac{\beta_t}{2};\ g(t) = \sqrt{\beta_t} \quad (4)$$

Within this formulation, unconditional sampling in the latent space can be achieved with the following update [2]:

$$z_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} \hat{z}_{0_t} + \tilde{\sigma}_t w \quad (5)$$

Where $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{i=0}^{t} \alpha_i$, and $\tilde{\sigma}_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$.

The term $\hat{z}_{0_t}$ represents the posterior mean of the latent, and is computed following:

$$\hat{z}_{0_t} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(z'_t - \sqrt{1-\bar{\alpha}_t} \cdot s_{\theta_{t,z}}\right) \quad (6)$$

Within this model, DPS [4] leverages the information of $\hat{z}_{0_t}$ to minimize the distance to the measured data (y), by modifying Eq. (5) to include a correction term enforcing an update in the direction of the gradient of the posterior at each time step. Assuming a linearized forward model of CT image formation ($y = Ax + n$), with n representing additive Gaussian noise. The conditional score can then be approximated as [4]:

$$\nabla_{x_t} \log(p(y|\hat{x}_{0_t}) \approx \lambda \nabla_{x_t} \left|\left|y - A\hat{x}_{0_t}\right|\right|^2 \quad (7)$$

where λ is a hyperparameter that controls the weight of the data enforcement correction. Note that in Eq. 7, $\hat{x}_{0_t}$ is required to be in the image domain to compute the update and it is therefore not directly application in latent diffusion models.

Considering that the latent and image domains are connected through the VAE ($x = \mathcal{D}(z)$), it is possible to reformulate the Latent-DPS (LDPS) update as:

$$\nabla_{z_t} \log(p(y|\hat{z}_{0_t}) \approx \lambda \nabla_{z_t} \left|\left|y - A\mathcal{D}(\hat{z}_{0_t})\right|\right|^2 \quad (8)$$

Combining the update of DDPM in Eq. 5 with the correction term in DPS, the final LDPS update at time t can be posed as:

$$z_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} \hat{z}_{0_t} + \tilde{\sigma}_t w - \nabla_{z_t} \log(p(y|\hat{z}_{0_t})) \quad (9)$$

However, this "vanilla" LDPS formulation has been shown to be insufficient to produce high fidelity reconstructions [10][11]. This is due, in part, to the non-linearity of the VAE [11], i.e. the distance to the data in the image domain is not necessarily preserved in the latent domain. The distance preserving penalty introduced in section II.A aims to alleviate this limitation but does not guarantee linearity of the encoding operation.

Alternatively, the term $\nabla_{z_t} \log(p(\hat{z}_{0_t}|y)$ could be replaced with a more general correction to z_t, denoted Δz_t in this work, that aims to minimize the distance between the sampled solution at time t and the measured data y, but without imposing strict linearity constraints on $\mathcal{E}$ and $\mathcal{D}$.

Following a similar strategy as proposed in [11], the correction term Δz_t can be obtained via several gradient descent iterations to minimize a quadratic distance metric between the current image estimation, $\hat{x}_{0_t} = \mathcal{D}(\hat{z}_{0_t})$, and the measured data y. Differently to the hard constrain proposed in [11] that

involves iteration until convergence, in the proposed approach, only a small number of gradient descent iterations are required per sampling step. The updated $\hat{x}_{0_t}$ is then used to compute an updated latent vector $\hat{z}_{0_t}' = \mathcal{E}(\hat{x}_{0_t})$. Leveraging the distance-preserving properties of the VAE for small perturbations of the input, the update in Eq. 7 can be approximated in the latent space by defining the following correction:

$$\Delta z_t = \lambda \cdot \sqrt{\bar{\alpha}_t}\left(\hat{z}_{0_t}' - \hat{z}_{0_t}\right) \tag{10}$$

where λ acts as a hyperparameter controlling the strength of the correction. In addition to the scalar λ, the proposed approach contains a second hyperparameter, N_{it}, which is the number of iterations used for computation of $\hat{z}_{0_t}'$. The proposed latent diffusion posterior sampling approach is summarized in Algorithm 1.

Algorithm 1.

1. Initialize $z_t \rightarrow N(0,1)$.
2. For $\mathrm{t} = \mathrm{t}_{\mathrm{init}}$: 0 do
3. $\quad \hat{z}_{0_t} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(z_t - \sqrt{1-\bar{\alpha}_t} \cdot s_{\theta_{t,z}}\right)$
4. $\quad \hat{x}_{0_t} = \mathcal{D}(\hat{z}_{0_t})$
5. $\quad$ For n_{it} = 1:N_{it}
6. $\quad\quad \hat{x}_{0_t} = \hat{x}_{0_t} + \frac{A^T(y - A\hat{x}_{0_t})}{A^T A \cdot \mathbb{I}}$
7. $\quad \hat{z}_{0_t}' = \mathcal{E}\left(\hat{x}_{0_t}\right)$
8. $\quad \Delta z_t = \lambda \cdot \sqrt{\bar{\alpha}_t}(\hat{z}_{0_t}' - \hat{z}_{0_t})$
9. $\quad z_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} z_t + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} \hat{z}_{0_t} + \tilde{\sigma}_t w + \Delta z_t$
10. Return $\mathcal{D}(z_0)$

Our method is related to other methods in the bibliography. On the one hand, in PSLD [10] the authors proposed a "gluing" term correction to Eq. 9, yielding:

$$\nabla_{z_t} \log(p(\hat{z}_{0_t}|y) \approx \lambda \nabla_{z_t} \left|\left|\mathrm{y} - \mathrm{A}\mathcal{D}(z)\right|\right|^2 + \gamma \nabla_{z_t} \left|\left|\hat{z}_{0_t} - \mathcal{E}(\hat{x}_{0_t} + \mathrm{A}^{\mathrm{T}}(\mathrm{y} - \mathrm{A}\mathcal{D}(\hat{z}_{0_t})\right|\right|^2 \tag{11}$$

where γ is an additional hyperparameter that controls the gluing strength. This gluing term serves a similar purpose as the proposed correction in Eq. 7, by encouraging the correction to z_t to remain in the manifold of the possible solutions of the diffusion process.

Furthermore, in [11] the authors proposed a ReSample strategy in which an updated latent, $\hat{z}_{0_t}'$, is obtained by minimizing a reconstruction problem with a formulation similar to the update in Line 6 of Algorithm 1. The final latent vector, z_t', is then obtained via a resampling operation using the diffusion forward model, yielding:

$$z_t' \in \mathcal{N}\left(\kappa\sqrt{\bar{\alpha}_t}\hat{z}_{0_t}' + (1-\kappa)z_t, \kappa \cdot (1-\bar{\alpha}_t)\mathbb{I}\right) \tag{12}$$

where κ is a scalar that controls the contribution of the new resampled solution and the previous z_t.

In this work, we adapted these two methods to provide a benchmark for comparison with the proposed approach.

C. Validation Studies

The proposed latent diffusion model was trained for synthesis of brain/head CT using a total of 25000 2D axial slices from the publicly available CQ500 dataset [12], containing images with and without visible ICH. All the images were resized to 256 x 256 pixels with 0.5 mm isotropic size. The VAE and score matching networks were modeled as a modified version of [9]. The VAE was trained using the loss described in section II.A, for a 64x64-features latent space. The unconditional latent diffusion model was trained using 1000 time-steps and a U-Net denoising network, following [9].

The proposed latent diffusion posterior sampling method was assessed for reconstruction of stationary CT for ICH imaging on a set of controllable simulation studies. The simulation setup used a stationary CT acquisition geometry based on a previously reported scanner design [5]. The system integrates a Multi X-ray source Array (MXA) of 31 stationary sources in combination with a curved area detector with a radius of 450 mm. The sources are distributed on a circular arc placed at a distance of 650 mm from the detector, measured from the central source of the array. The radius of curvature of the MXA was set to 450 mm.

For the sake of simplicity, a one-dimensional circular detector with 1650x1 pixels of 0.5 mm size was considered in this study. Stationary CT projection data were obtained via forward projection of randomly selected individual axial slices extracted from brain CT volumes from a testing subset of the CQ500 dataset, containing 33 CT instances. The baseline CQ500 anatomies were augmented with synthetic ICH, modeled as randomly deformed spheres of 3 mm radius and 40 HU attenuation, and introduced at random positions within the brain parenchyma. Notice that this type of lesions were not present in training. Finally, quantum noise (0.5 mAs/projection) was introduced to the simulated the projection data.

Volumetric reconstructions were obtained for 256 x 256 voxels, with 0.5 mm isotropic size. For the proposed approach, we found the optimal hyperparameters for λ following a linear schedule from $\lambda = (0.05, 0.3)$ and a total of N_{it} = 20 iterations per time step. For PSLD we optimized λ and γ to 8 and 4 respectively. For ReSample we used 50 iterations per time step and κ following a linear schedule from 0.75 to 0.95. To reduce the computational burden, for both ReSample and our method, we only applied the correction update to the last 800 time-steps. Solutions of z_t for t > 800 were observed to be highly dominated by noise diminishing the effect of the correction terms [11].

The proposed approach was evaluated in comparison with the latent space DPS alternatives described in section II.B, and with a conventional model-based iterative reconstruction (MBIR) method based on a Penalized Weighted Least Squares (PWLS) algorithm [5] with 200 iterations and Nesterov momentum acceleration. A Huber, edge preserving, penalty was used in the PWLS reconstructions [3].

Quantification of the of image quality was obtained via structural similarity SSIM metrics referred to the ground truth digital phantoms used for simulation. Recovery of the synthetic bleeds was assessed with dice coefficients, computed between the ground truth lesion and a threshold-based segmentation with lower and upper thresholds set to 20 HU and 80 HU and acting on the brain parenchyma.

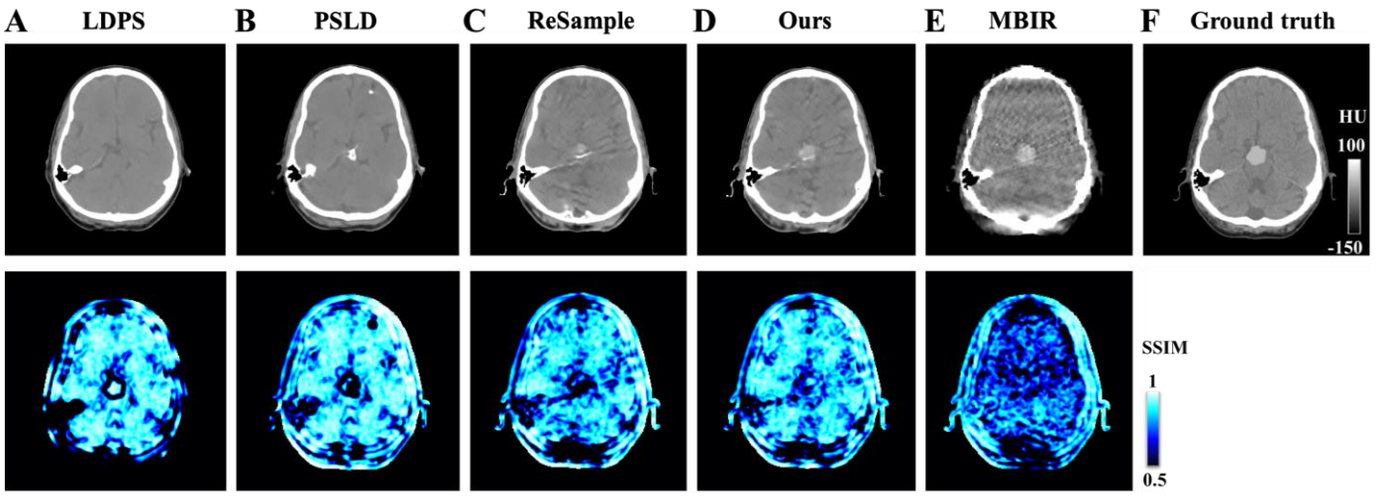


Figure 2. Example of reconstructed images with different approaches for an axial slice close to the skull base. (A) Vanilla latent DPS. (B) PSLD including a gluing strategy. (C) ReSample strategy. (D) Proposed latent posterior sampling approach. (E) Reference standard MBIR with an edge preserving penalty. (F) Digital phantom used as ground truth. Top row: reconstructed images. Bottom row: corresponding SSIM distributions referenced to ground truth.

III. Results and Discussion

Fig. 2 shows reconstructed images with the different methods studied in this work. For this example, a synthetic bleed (5 ml volume) was placed at an approximately central location inside the brain parenchyma.

Fig. 3 shows quantitative evaluation of image quality and ICH delineation via SSIM with respect to the ground truth (Fig. 3A) and dice coefficient (Fig. 3B) for the different studied configurations. Bleeds with DICE<0.1 were deemed undetectable and not included in the distributions in Fig. 3B.

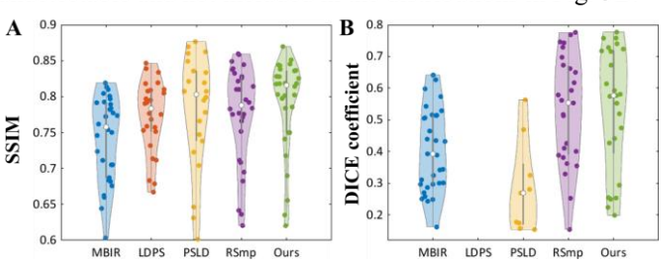


Figure 3. (A) Distributions of SSIM values referenced to the ground truth for the different reconstruction strategies. (B) DICE coefficient computed on the reconstructed bleeds, using the simulated bleed extent as ground truth.

According to the results presented in Fig. 2 and Fig. 3, the LDPS approach with no correction strategies failed at reproducing the underlying ground truth despite the use of a distance preserving VAE. In particular, the synthetic bleeds were not recovered in any of the test cases, consistently yielding DICE < 0.1. In comparison, PSLD produces results that are closer overall to the ground truth, yielding median SSIM = 0.80, compared 0.78 for LDPS. However, bleed DICE coefficients remained very low (DICE<0.3), rendering a significant number of synthetic bleeds unidentifiable (23 out of the total 33). Both ReSample and the proposed approach yielded similar results in terms of bleed recovery with median DICE ~0.58. However, the proposed method resulted in better image quality, reflected in a larger median SSIM (0.82 vs 0.79). Furthermore, the proposed approach required a lower number of iterations per time step (20 vs 50 in ReSample). In comparison to conventional MBIR, the proposed approach resulted in a noticeable reduction of artifacts associated to limited and sparse sampling, reflected in a lower median SSIM = 0.76 for MBIR. Residual artifacts also hampered the detection of synthetic bleeds, yielding median DICE ~0.4, lower than the proposed method.

While running several iterative steps increases the computational burden of diffusion sampling, our approach with 20 iterations per time step resulted in lower runtime compared to DPS in the image domain. In particular, we found that the proposed method provided a 6-fold speedup compared to previous results [3] on image domain DPS. Further acceleration can be achieved via stochastic contraction approaches [13], that proved suitable for acceleration of image-domain DPS.

IV. Conclusions

In summary, this work proposed a novel strategy for posterior sampling with latent diffusion models via combination of a distance preserving VAE, coupled to a correction-based approach to posterior sampling. The proposed method was exercised for image reconstruction in stationary CT, involving heavily sparse and limited sampling of the CT volume, and resulted in promising performance for reconstruction of underlying image features and generalization to out-of-domain scenarios, including imaging of ICH, while relaxing the computational burden and memory requirements in DPS reconstruction.

References

[1]Song, Y., et al. (2020). arXiv preprint arXiv:2011.13456.
[2]Ho, J. et al. (2020). Adv. in neural inf. processing systems, 33, 6840-6851.
[3]Lopez-Montes, A., et al. (2023). In Proc.: Fully 3D. arXiv:2310.16846
[4]Chung, H. et al. (2022). arXiv preprint arXiv:2209.14687.
[5]Lopez-Montes, A. et al.(2022). In Proc.: CT meet. doi: 10.1117/12.2646858
[6]McSkimming, T., et al.(2023). In Proc.: SPIE 2023. doi:10.1117/12.2654383
[7]Jolicoeur-Martineau, A. et al.(2021). http://arxiv.org/abs/2105.14080
[8]Song, Yang, and Prafulla Dhariwal. (2023) arXiv:2310.14189.
[9]Rombach, R., et al. (2022). In Proc.: IEEE/CVF (pp. 10684-10695).
[10]Rout, L., et al. (2023). arXiv:2307.00619.
[11] Song, B. et al. (2023). arXiv preprint arXiv:2307.08123.
[12] S. Chilamkurthy et al., (2018) http://arxiv.org/abs/1803.05854
[13] Chung, H. et al. (2022). In Proc.: IEEE/CVF (pp. 12413-12422)

Learned Tube Current Modulation from Scout Images for Individualized Organ-Level CT Dose and Image Quality

Sen Wang, Maria Jose Medrano, Jennie Jiayi Cao, Wonkyeong Lee, Grant M. Stevens, Justin Ruey Tse, Adam S. Wang

***Abstract*— For optimized dose efficiency, automatic exposure control based on tube current modulation (TCM) is used in modern scanners to reduce dose while maintaining image quality. Conventional TCM uses total exposure from the tube and noise in the center of CT slices as surrogates of dose and image quality, respectively. Scout images are acquired before the CT scan to verify patient positioning and are also used for TCM planning. In this work, we demonstrate that with deep learning, it is possible to enable more personalized, radiation risk-oriented and organ-level exposure control. Our method utilizes scout images as inputs and generates the desired TCM as outputs for the specific task(s). Compared to conventional TCM, the proposed method offers the flexibility to prioritize organs or tasks. The training and evaluation references were established through retrospective optimization on a previously acquired CT dataset. The 2D TCM map is parameterized based on DCT basis images so that the target for the network is to learn the corresponding DCT coefficients. We validated the performance of the proposed method in an in-house dataset consisting of 96 patient cases and showed its applicability to different tasks, for example, optimized TCM for bladder, liver, and lung. The results showed that for the targeted organ, the relative errors in organ noise and dose are around 10% and the relative error for the whole body is around 8%. By comparison, conventional TCM with uniform dose reduction produced errors of 40%. These results demonstrate the potential to directly predict optimized TCM from scout images.**

***Index Terms* — deep learning, automatic exposure control, tube current modulation**

I. Introduction

Automatic exposure control provides automatic voltage selection and tube current modulation[1]–[3] (TCM, both angular and longitudinal) during the whole-body CT scan. This provides a practical way to enhance image quality while maintaining a fixed dose or vice versa. This work focuses on optimizing the TCM, which adjusts the tube current during the scan to accommodate variations in attenuation from lateral/frontal views or changes in body shape along the z-direction.

Typical TCM mainly works around physical quantities such as using total exposure in terms of current-time product (mAs) as the dose surrogate and noise in the center of the image as the quality indicator[4]. It does not assign proper weights regarding organ-level risk and lacks the flexibility to automatically adjust the emphasis on different organs according to disease-specific conditions.

In our previous work [5], [6], we developed a pipeline to retrospectively generate personalized organ-level TCM profiles from patient's CT scans. It is scalable to large clinical datasets so that we can take full advantage of the established ground-truth references and train a model that directly predicts the optimized TCM from scout images, which then enables prospective planning. Thus, in this work, we evaluate the ability of a deep learning model to directly predict the optimized TCM from scout images.

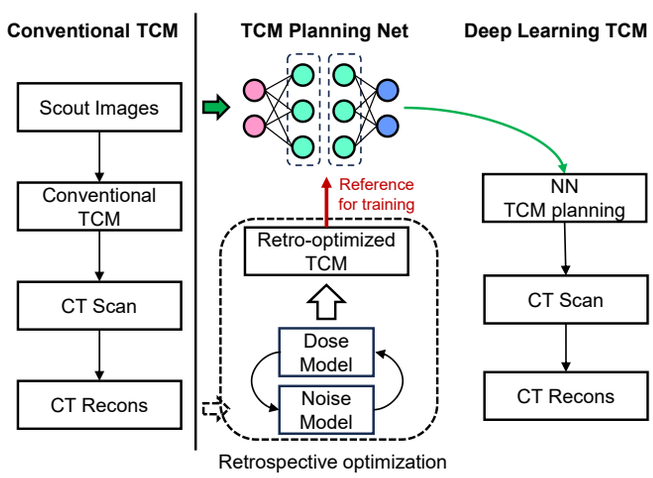


Fig. 1. Overview of the proposed deep learning-based TCM, where the training reference was established from retrospective optimization. After sufficient training and validation, it can be easily incorporated into the conventional CT workflow for more individualized and organ-specific TCM planning.

This work was supported by GE HealthCare.

Sen Wang, Maria Jose Medrano, Jennie Jiayi Cao, Wonkyeong Lee, and Justin Ruey Tse are with the Department of Radiology, Stanford University, Stanford, CA USA 94305. (Email: senwang@stanford.edu, medranom@stanford.edu, jcao19@stanford.edu, uulee@stanford.edu, jrtse@stanford.edu).

Grant Stevens is with GE HealthCare, Waukesha, WI USA 53188 (Email:grant.stevens@gehealthcare.com)

Adam Wang is with the Departments of Radiology and Electrical Engineering, Stanford University, Stanford, CA USA 94305. (Email: adamwang@stanford.edu)

II. Methods

We show the schematic pipeline for the proposed deep learning-based TCM planning using neural networks (nn TCM) in Fig. 1. The training references are from retrospective optimization based on CT scans, which are used to train a network with scout images as the inputs to predict or fit the optimized TCM. We will briefly introduce the retrospective optimization for TCMs and then describe more details about the network.

A. Reference TCM generation from retrospective optimization

We used a 2D TCM map to describe the tube current modulation in a CT scan where the horizontal axis is the gantry angle and the vertical axis is the z increment. There are two key components in our retrospective optimization for TCM:

1) Empirical dose and noise models based on CT scan. Given a specific TCM map, we were able to estimate the corresponding organ doses and noise levels using Monte Carlo simulation (MC-GPU) [7], a previously developed empirical noise model [8], and automatic organ segmentation (TotalSegmentator) [9].

2) Parameterized TCM maps. We parameterized the TCM map using Discrete Cosine Transform (DCT) basis maps so that the goal of the optimization is to find the optimal DCT coefficients that minimize a given task.

Thus, we formulate the optimization as follows:

$$\begin{aligned} &\min_{c_{ij}} \textstyle\sum_L w_L^n n_L \\ &\text{such that:} \end{aligned} \quad \begin{aligned} &\mathrm{mA}_{\min} \le \mathrm{TCM}(z_n, \phi_n) \\ &= \sum\nolimits_{i,j} c_{ij}\, B_{ij}(z_n, \phi_n) \le \mathrm{mA}_{\max}, \\ &\sum\nolimits_L w_L^d \sum\nolimits_{i,j} c_{ij} \sum\nolimits_n B_{ij}(z_n, \phi_n) d_L[n] \\ &\qquad = \chi \sum\nolimits_L w_L^d d_L^{100}, \\ &\sum\nolimits_{i,j} c_{ij}\, B_{ij}(z_n, \phi_n) \le \mathrm{TCM}_0(z_n, \phi_n), \end{aligned} \tag{1}$$

where n_L is the organ-level noise, and w_L^n is the corresponding weighting factor that can be adjusted to shift the image quality focus onto different organs. In the constraints, we set the upper and lower limit of tube current modulation to be $\mathrm{mA}_{\max}$= 800 mA and $\mathrm{mA}_{\min}$= 50 mA. The average dose is fixed at 50%, i.e., χ=0.5, and d_L^{100} are the original organ doses. w_L^d is the weighting factor for the corresponding organ, and we used w_L^d=1 for all six organs and entire body, although it can be easily replaced by other weighting schemes, such as effective dose weights. The final outputs from the optimization framework are the optimal DCT coefficients (c_{ij}), as well as the corresponding dose and noise estimates from dose calculation and the empirical noise model, respectively.

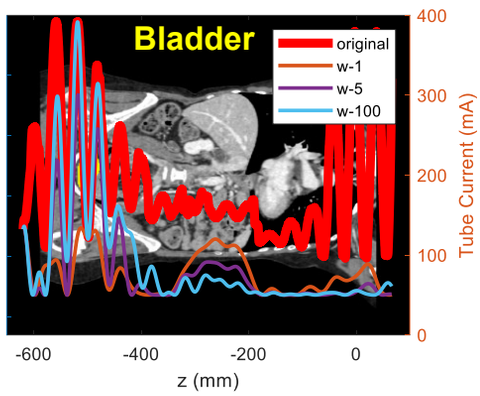


Fig. 2. An example of optimized TCM from retrospective optimization that used Monte Carlo dose calculation and an empirical noise model based on the CT images. Image quality for the bladder is optimized by minimizing noise for fixed dose (50% of original scan).

We illustrated an example of optimized TCM in Fig. 2 where we can find that given the same amount of total dose, with more emphasis on the image quality on the bladder, more and more dose was allocated around it. w-1, w-5, and w-50 denote the weighting factor for bladder noise w_{Bladder}^n as 1, 5 and 50, respectively. For example, w-50 weights the bladder noise 50 times higher than the default w-1.

B. Scout-based TCM planning from neural network.

We illustrate the network-based TCM planning from scout images in Fig. 3. Two independent networks are used to extract features from two scout images (frontal and lateral), which are then merged for decoders to fit/predict DCT coefficients that define TCM maps. In this work, we used ResNet-18 to extract the features but they can be easily replaced by any other

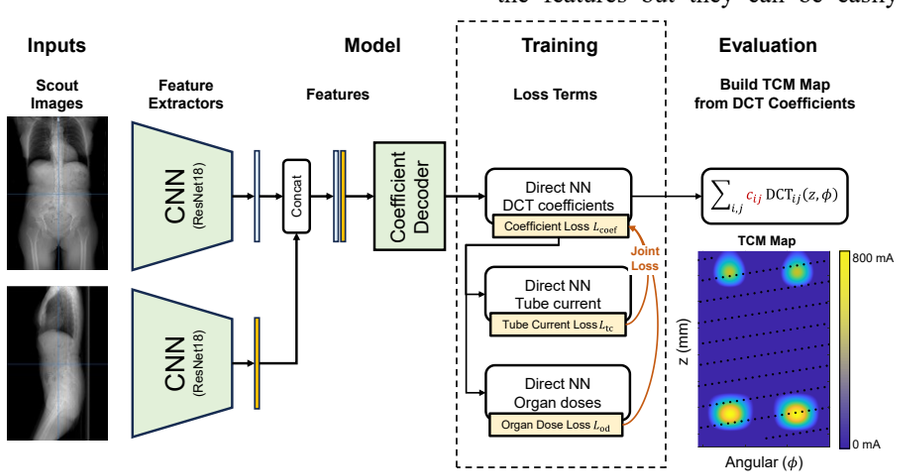


Fig. 3. Overview of network design for TCM planning based on scout images.

networks if better performance can be found. We applied three different losses from different domains to help/stabilize the convergence of the network:

1) Coefficient loss in the DCT coefficient domain (equation (2)). Instead of predicting the whole TCM map, which is a 2D image, we chose to have the network predict the DCT coefficients. This reduces the number of unknown parameters the network needs to learn.

$$L_{\text{coef}} = \sum_{ij} \left(c_{ij}^{NN} - c_{ij}^{REF}\right)^2 \qquad (2)$$

2) Tube current loss (equation (3), p^{REF} is the reference tube current curve). DCT coefficients do not yield direct and clear physical meanings. Once they are determined, it is possible to calculate the corresponding tube current profile based on the DCT basis. Hence, we applied additional tube current loss which is in the tube current domain to help the network learn physically meaningful coefficients, including tube current limits.

$$L_{\text{tc}} = \left\| \sum_{i,j} c_{ij}^{NN} B_{ij}(z_n, \phi_n) - p^{REF}(z_n, \phi_n) \right\|_2 \qquad (3)$$

3) Organ dose loss (equation (4), d^{REF} is the reference dose for organs). Similar to tube current loss, we added an organ dose loss term in the organ dose domain to further enhance the physical understanding of the trained network and match desired dose levels.

$$L_{\text{od}} = \left\| \sum_{i,j} c_{ij}^{NN} \sum_{n} B_{ij}(z_n, \phi_n) d_L[n] - d^{REF}[L] \right\|_2 \qquad (4)$$

An ablation study was conducted to demonstrate the effectiveness of the loss terms.

C. Dataset preparation and network performance evaluation

To train and evaluate the applicability and performance of the proposed method, we utilized 96 clinical CT exams (120 kVp, Chest-Abdomen-Pelvis) from a General Electric (GE) Revolution scanner, which were collected under IRB approval from Stanford Health Care. Scans were processed in the retrospective optimization to optimize the noise performance around different target organs with 50% of their routine clinical dose. The TCM maps were scaled with respect to the rotation speed, to normalize for its impact on dose and image quality.

70 cases (scout images and corresponding TCMs) were used for training and 12 cases were used for validation to save the best model weights during the training. 14 cases were reserved for the test set.

The model was trained for 1000 epochs with the Adam optimizer and a learning rate of 1e-3 for the first 500 epochs, followed by 1e-4 for the remaining epochs.

The performance of the network is evaluated and reported in terms of relative error in organ-level dose and noise with respect to the retrospective optimization.

III. Results

A. Predicting optimized TCM for bladder

We first show the test results of predicting optimized TCM for the bladder with the proposed neural network method in Fig. 4 and Fig. 5. The organ-level dose and noise were evaluated. The reference TCM was from a retrospective optimization which emphasized the noise property of bladder by putting a factor of 50 onto it. We observed from the retrospective reference (white bars in Fig. 4) that a reasonable amount of dose was allocated around the bladder and its image noise was lower than that of other organs.

We can find that conventional TCM with uniform dose reduction is not capable of organ-level optimization, so in the bladder it produced large errors (~40% in organ dose and noise) compared to the TCM from retrospective optimization. By comparison, the proposed neural network is trained using retrospectively optimized TCM (bladder) as the reference. It approaches the performance of retrospective optimization, with a relative error around 10% in both bladder dose and noise.

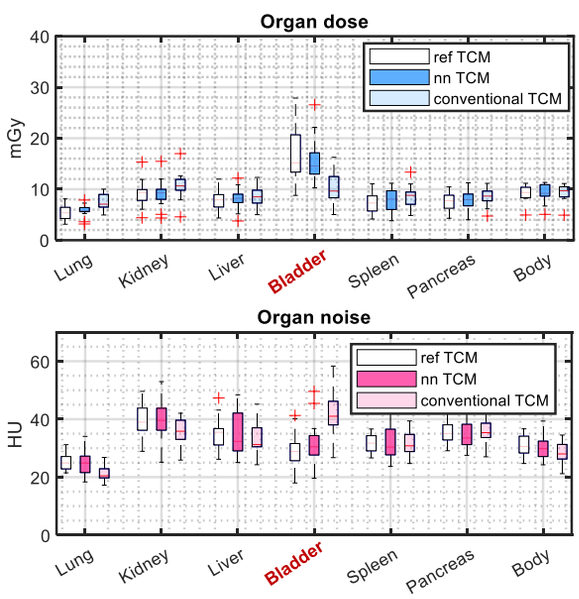


Fig. 4. Dose and noise of the organs from reference TCM (retrospective optimization), conventional TCM, and nn TCM. In the retrospective optimization, a weighting factor of 50 was applied on the bladder noise ($w^n_{Bladder}$=50 in equation (1)).

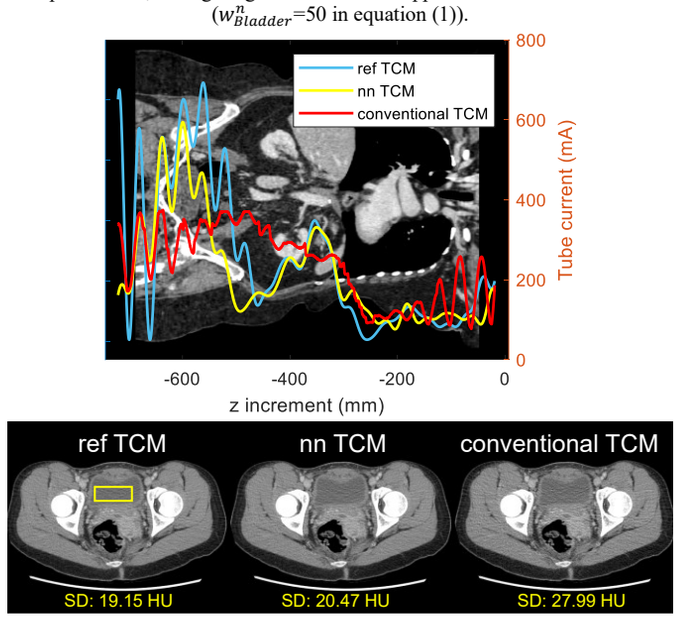


Fig. 5. An example of retrospective TCM (reference), nn TCM, and conventional TCM. nn TCM was able to optimize noise of bladder and is closer to the reference TCM's performance compared to conventional TCM.

TABLE II
APPLICABILITY TO DIFFERENT TASKS

		Relative error in test set						
		Lung	Kidney	Liver	Bladder	Spleen	Pancreas	Body
Bladder-Net	Dose	16.87%	13.58%	12.82%	**13.14%**	15.90%	14.04%	8.39%
	Noise	9.12%	10.58%	7.95%	**9.84%**	9.28%	8.60%	4.96%
Liver-Net	Dose	21.10%	18.79%	**10.00%**	17.65%	13.50%	12.31%	9.26%
	Noise	12.41%	23.91%	**10.90%**	9.07%	9.44%	14.24%	7.77%
Lung-Net	Dose	**14.35%**	19.59%	16.02%	20.48%	14.99%	15.46%	8.62%
	Noise	**11.72%**	17.10%	11.88%	16.32%	10.37%	14.56%	6.16%

B. Ablation study of loss terms

We list the results from an ablation study in TABLE I to show the performance improvement by incorporating loss terms from different domains as was introduced in the Methods section. We can observe a decreasing trend in relative error with more loss terms included in the training.

TABLE I
ABLATION STUDY FOR LOSS TERMS

Loss terms	Relative error in bladder	
	Dose (%)	Noise (%)
Coefficient loss	16.94%	12.20%
+ Tube current loss	15.94%	11.04%
+ Organ dose loss	13.14%	9.84%

C. Applicability to different organs/tasks

TCM planning may require flexibility to adjust the optimization task. For example, one may have different target organs of interest. To meet the need, currently we opt to build different training sets targeting different organs based on retrospective optimization, such a bladder, liver, and lung, and then train a model separately for the specific target. This may allow the interpreting radiologist to tailor the CT to the disease process and prioritize organs depending on the clinical indication.

TABLE II lists the test performances of the three models trained for optimal TCM of bladder, liver, and lung. From the table we can find that the three models were able to output distinct types of TCMs that target different organ tasks, achieving an error rate of approximately 10% for the target organ. All of them relied on the same input: the scout image, which is the only information available before the formal image acquisition.

IV. DISCUSSION AND CONCLUSION

We demonstrated the potential application of deep learning in TCM planning by taking the scout images as input and generating task-specific TCM, enabling more flexible, radiation risk-oriented and patient-specific planning. We used a dataset of 96 patient cases to train and validate its performance, and the results showed that it is capable of designing TCM plans with respect to different organ tasks and generally outperforms conventional TCM, with retrospectively optimized TCM as the reference. We leveraged parameterized TCMs and multi-domain physical constraints, aiming to improve and stabilize the training.

There are several future works in progress. First, our current dataset is relatively small, and we are collecting and generating additional reference data. Second, we currently only utilize physical constraints from tube current and organ doses in the training loss function, rather than as an explicit constraint. We are considering enforcing physical constraints, for example tube current constraints from the generator/tube, and are also aiming to further improve the performance through organ noise loss, etc. Lastly, a multi-reader study is needed to assess the tangible impact of different TCM maps on clinical tasks such as liver lesion detection.

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," *Medical physics*, vol. 26, no. 11, pp. 2235–2247, 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.

[3] M. K. Kalra *et al.*, "Techniques and applications of automatic tube current modulation for CT," *Radiology*, vol. 233, no. 3, pp. 649–657, 2004.

[4] L. Klein *et al.*, "Patient-specific radiation risk-based tube current modulation for diagnostic CT," *Medical Physics*, vol. 49, no. 7, pp. 4391–4403, 2022.

[5] S. Wang, M. Medranoa, G. Stevens, J. R. Tse, and A. S. Wang, "Retrospective Tube Current Modulation Optimization of Individualized Organ-Level CT Dose and Image Quality," in *Medical Imaging 2024: Physics of Medical Imaging*, SPIE, 2024.

[6] M. Medrano, Wang, Sen, Stevens, Grant, Tse, Ruey, Justin, and Wang, Adam, "Personalized, Scout-Based Dose Estimation for Prospective Optimization of CT Tube Current Modulation," in *Medical Imaging 2024: Physics of Medical Imaging*, SPIE, 2024.

[7] 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.

[8] A.-A.-Z. Imran, D. Pal, S. Wang, S. Dutta, E. Zucker, and A. Wang, "Personalized CT organ noise estimation from scout images," in *Medical Imaging 2022: Physics of Medical Imaging*, W. Zhao and L. Yu, Eds., SPIE, 2022, p. 120310R. doi: 10.1117/12.2610986.

[9] J. Wasserthal, M. Meyer, H.-C. Breit, J. Cyriac, S. Yang, and M. Segeroth, "TotalSegmentator: robust segmentation of 104 anatomical structures in CT images," *arXiv preprint arXiv:2208.05868*, 2022.

CT Material Decomposition using Spectral Diffusion Posterior Sampling

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

***Abstract*—In this work, we introduce a new deep learning approach based on diffusion posterior sampling (DPS) to perform material decomposition from spectral CT measurements. This approach combines sophisticated prior knowledge from unsupervised training with a rigorous physical model of the measurements. A faster and more stable variant is proposed that uses a "jumpstarted" process to reduce the number of time steps required in the reverse process and a gradient approximation to reduce the computational cost. Performance is investigated for two spectral CT systems: dual-kVp and dual-layer detector CT. On both systems, DPS achieves high Structure Similarity Index Metric Measure(SSIM) with only 10% of iterations as used in the model-based material decomposition(MBMD). Jumpstarted DPS (JSDPS) further reduces computational time by over 85% and achieves the highest accuracy, the lowest uncertainty, and the lowest computational costs compared to classic DPS and MBMD. The results demonstrate the potential of JSDPS for providing relatively fast and accurate material decomposition based on spectral CT data.**

I. Introduction

SPECTRAL CT has enabled a number of applications including density estimation, virtual monochromatic imaging, and contrast agent enhancement. Many of these applications rely on material decomposition, which reconstructs basis material density maps from the spectral projections[1]. Material decomposition is an ill-conditioned nonlinear inverse problem without an explicit solution. Existing material decomposition algorithms can be categorized into three main types: analytical decomposition[2], iterative/model-based decomposition[3], and learning-based decomposition[4].

Deep learning methods have been extensively used in medical image formation including spectral CT[4], [5], [6], [7], [8]. Such approaches leverage prior knowledge learned from large datasets and can generally surpass classic approaches in terms of image quality. However, many deep learning methods do not directly leverage a physical model of the measurements, raising concerns about robustness, network hallucinations, etc. Some methods integrate a physical model as part of network training to improve the data consistency[9], [10]. However, these trained networks are tailored to specific system models, requiring network retraining for system changes in protocol, technique, device, etc.

Diffusion Posterior Sampling (DPS)[11] has established a new framework to integrate a learned prior and physics model. Specifically, DPS starts with unsupervised training of a score-based generative model (SGM) to capture the target domain distribution, then application of a reverse process to estimate image parameters - alternating between SGM reverse sampling to drive the image towards the target distribution and model-based updates to improve the data consistency with measurements. As a result, the final output adheres to both the prior distribution and the actual measurements. A major advantage of DPS is that its network training is not specific to any physical model, allowing for application across different imaging systems. Recently, we have applied DPS to nonlinear CT reconstruction[12], demonstrating its effectiveness for both low-mA and sparse-view single-energy CT reconstruction.

In this work, the original DPS framework is expanded to Spectral DPS (SDPS) specifically for material decomposition in spectral CT. To further enhance this method, we propose a "jumpstarted" sampling strategy with gradient approximation, which significantly stabilizes the sampling process and reduces computational costs. The performance of both SDPS and its jumpstarted variant (JSDPS) are demonstrated in dual-kVp[13] and dual-layer CT[14].

II. Methodology

A. Spectral Deep Posterior Sampling

1) Score-based Generative Model (SGM) for Inverse Problems: A SGM[15] defines a forward process which continuously perturbs the target-domain sample with time-dependent noise. The corresponding reverse process generates new samples by inverting the perturbation process. Both forward and reverse processes are described by a stochastic differential equation (SDE). Specifically, the SDEs for a Denoising Diffusion Probabilistic Model (DDPM)[16] have the following form at time t:

$$Forward: \mathrm{d}\mathbf{x} = -\frac{\beta_t}{2}\mathbf{x}\mathrm{d}t + \sqrt{\beta_t}\mathrm{d}\mathbf{w} \tag{1a}$$

$$Reverse: \mathrm{d}\mathbf{x} = [-\frac{\beta_t}{2}\mathbf{x} - \beta_t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)]\mathrm{d}t + \sqrt{\beta_t}\mathrm{d}\mathbf{w} \tag{1b}$$

where β_t is the time-dependent noise variance and $\mathrm{d}\mathbf{w}$ is the standard Wiener process. The unknown score function $\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)$ is approximated by a deep neural network $\mathbf{s}_\theta(\mathbf{x}_t, t)$. A SGM is also capable of posterior sampling from conditional distribution $p(\mathbf{x}|\mathbf{y})$:

$$\mathrm{d}\mathbf{x} = [-\frac{\beta_t}{2}\mathbf{x} - \beta_t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t|\mathbf{y})]\mathrm{d}t + \sqrt{\beta_t}\mathrm{d}\mathbf{w} \tag{2}$$

In the context of CT, $\mathbf{y}$ and $\mathbf{x}$ are the projection measurements and the image volume, respectively. Leveraging Bayes rule $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{x})p(\mathbf{y}|\mathbf{x})$, the reverse sampling (2) may be reformulated as:

$$\begin{aligned}\mathrm{d}\mathbf{x} =&[-\frac{\beta_t}{2}\mathbf{x} - \beta_t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)]\mathrm{d}t + \sqrt{\beta_t}\mathrm{d}\mathbf{w}\\ &- \beta_t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{y}|\mathbf{x}_t)\mathrm{d}t\end{aligned} \tag{3}$$

The conditional distribution $p_t(\mathbf{y}|\mathbf{x}_t)$ can be approximated[11] as

$$p_t(\mathbf{y}|\mathbf{x}_t) \approx p_t(\mathbf{y}|\hat{\mathbf{x}}_0), \text{where } \hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t + (1-\bar{\alpha}_t))\mathbf{s}_\theta(\mathbf{x}_t, t) \tag{4}$$

Based on this approximation, the deep posterior sampling process could be expressed as[12]:

$$\begin{aligned}\mathrm{d}\mathbf{x} =&[-\frac{\beta_t}{2}\mathbf{x} - \beta_t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)]\mathrm{d}t + \sqrt{\beta_t}\mathrm{d}\mathbf{w}\\ &- \beta_t\nabla_{\hat{\mathbf{x}}_0}\log p_t(\mathbf{y}|\hat{\mathbf{x}}_0)\nabla_{\mathbf{x}_t}\hat{\mathbf{x}}_0\mathrm{d}t\end{aligned} \tag{5}$$

Note that the first two term is exactly same as the unconditional DDPM sampling (1b), and $p_t(\mathbf{y}|\hat{\mathbf{x}}_0)$ is the likelihood term for the measurements. Therefore, the reverse sampling (5) integrates the prior information captured by DDPM and the physical measurement model provided by the likelihood function.

The forward process is discretized into T time steps [16]:

$$\begin{aligned}&\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}\\ &\text{where } \bar{\alpha}_t = \prod_{i=1}^{t}(1-\beta_t),\ \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I}),\ t = 1, 2...T\end{aligned} \tag{6}$$

Since $\boldsymbol{\epsilon} = -\sqrt{1-\bar{\alpha}_t}\ \log p_t(\mathbf{x}_t)$, DDPM trains a network $\boldsymbol{\epsilon}_\theta(\mathbf{x}, t)$ to predict the noise $\boldsymbol{\epsilon}$. That is, network parameters, θ, are estimated

$$\theta^* = \text{argmin } \mathbb{E}_{\mathbf{x}_0}\mathbb{E}_{\boldsymbol{\epsilon},t}\|\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) - \boldsymbol{\epsilon}\|_2^2 \tag{7}$$

Using the trained network and physical model, the image $\mathbf{x}_0$ may be reconstructed by solving the reverse SDE[16], [17].

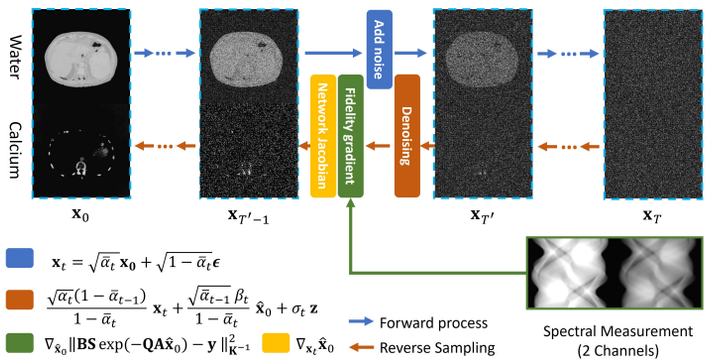


Fig. 1. SDPS Workflow: Basis material images are concatenated to form $\mathbf{x}_0$. The forward process progressively perturbs images with noise, and a noise prediction network is trained. The reverse process uses DDPM sampling to progressively denoise the image and uses a likelihood-based update to enforce data consistency.

2) Diffusion Posterior Sampling for Material Decomposition: This work aims to apply the DPS for the material decomposition problem, which estimates basis material densities from the spectral CT measurements[1]. We adopt a general spectral CT model[14] where measurements $\mathbf{y}$ are assumed to follow a multi-variate Gaussian distribution with covariance $\mathbf{K}$: $\mathbf{y} \sim \mathcal{N}(\bar{\mathbf{y}}, \mathbf{K})$ and mean:

$$\bar{\mathbf{y}} = \mathbf{BS}\exp(-\mathbf{QAx}) \tag{8}$$

where $\mathbf{x}$ are densities for each basis volume, $\mathbf{Q}$ are mass attenuation coefficients for each basis, $\mathbf{A}$ represents (channel-specific) projection matrices, $\mathbf{S}$ are channel-specific spectral response, and $\mathbf{B}$ captures overall gain effects. Substituting Eq.(8) into Eq.(5), we have the spectral DPS (SDPS) for material decomposition:

$$\begin{aligned} \mathrm{d}\mathbf{x} =& [-\frac{\beta_t}{2}\mathbf{x} - \beta_t\nabla_{\mathbf{x}_t}\log p_t(\mathbf{x}_t)]\mathrm{d}t + \sqrt{\beta_t}\mathrm{d}\mathbf{w} \\ & - \beta_t\nabla_{\hat{\mathbf{x}}_0}\|\mathbf{BS}\exp(-\mathbf{QA}\hat{\mathbf{x}}_0) - \mathbf{y}\|^2_{\mathbf{K}^{-1}}\nabla_{\mathbf{x}_t}\hat{\mathbf{x}}_0\mathrm{d}t \end{aligned} \tag{9}$$

The SDPS pseudo-code is shown below and illustrated in Fig. 1.

Algorithm 1 Spectral Diffusion Posterior Sampling (SPDS)

T: diffusion steps
η_t: step size
$\mathbf{y}$: spectral CT measurement
$\mathbf{x}_T \sim \mathcal{N}(0, \boldsymbol{I})$
for $t = T$ to 1 **do**:
  $\mathbf{z} \sim \mathcal{N}(0, \boldsymbol{I})$
  $\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t))$
  $\mathbf{x}'_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \sigma_t\mathbf{z}$
  $\mathbf{x}_{t-1} = \mathbf{x}'_{t-1} - \eta_t\nabla_{\mathbf{x}_t}\|\mathbf{BS}\exp(-\mathbf{QA}\hat{\mathbf{x}}_0) - \mathbf{y}\|^2_{\mathbf{K}^{-1}}\nabla_{\mathbf{x}_t}\hat{\mathbf{x}}_0$
end for

B. Strategies for fast and stable SPDS

1) Instability of SDPS: The SDPS framework has the advantage of combining a learned prior and a physics model. However, we have observed that the stability of SDPS is highly dependent on the step size, η_t, and that even with careful tuning, decompositions present large variations over posterior samples and hallucinations. Careful design of a step size scheduler is required to balance the prior information and measurements, particularly for the earlier steps in the reverse process. In early steps, $\mathbf{x}_t$ is far from the solution. While large step sizes may be favorable to enhance data consistency, this risks disrupting diffusion updates - making it challenging to pick optimal step sizes.

2) Jumpstarted Sampling: For many imaging problems, an initial estimate is readily available through fast computation. For example, in spectral CT, projection-domain or image-domain material decomposition [18], [19] can be used as the first-pass estimate, which we denote as $\hat{\mathbf{x}}_0^f$. Forward diffusion can be directly performed on $\hat{\mathbf{x}}_0^f$ as $\hat{\mathbf{x}}_t^f = \sqrt{\bar{\alpha}_t}\hat{\mathbf{x}}_0^f + (\sqrt{1-\bar{\alpha}_t})\boldsymbol{\epsilon}$. The difference between the distribution of $\hat{\mathbf{x}}_t$ and $\hat{\mathbf{x}}_t^f$ can be quantified by the KL divergence:

$$D_{KL}(p(\mathbf{x}_t|\mathbf{x}_0) \parallel p(\hat{\mathbf{x}}_t^f|\hat{\mathbf{x}}_0^f)) = \frac{\bar{\alpha}_t}{2(1-\bar{\alpha}_t)}\|\hat{\mathbf{x}}_0^f - \mathbf{x}_0\|_2^2. \tag{10}$$

In forward diffusion, $\bar{\alpha}_t/(1-\bar{\alpha}_t)$ progressively diminishes, ultimately converging towards zero. Consequently, Eq.(10) implies that we can expect for sufficiently large t, the difference between the distribution of $\mathbf{x}_t$ and $\hat{\mathbf{x}}_t^f$ is small enough to be ignored, then the reverse sampling can start from $\hat{\mathbf{x}}_0^f$ instead of pure noise. In effect, we can skip early time steps by using an approximate solution to which an appropriate amount of noise has been added - yielding a more stabilized decomposition by using an initialization closer the solution as well as a faster solution having "jumpstarted" over many early time steps.

3) Gradient Approximation: SDPS computes the data fidelity gradient term via chain rule: $\nabla_{\mathbf{x}_t}\log p_t(\mathbf{y}|\hat{\mathbf{x}}_0) = \nabla_{\hat{\mathbf{x}}_0}\log p_t(\mathbf{y}|\hat{\mathbf{x}}_0)\nabla_{\mathbf{x}_t}\hat{\mathbf{x}}_0$. The $\nabla_{\mathbf{x}_t}\hat{\mathbf{x}}_0$ term may be expanded as:

$$\nabla_{\mathbf{x}_t}\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(1 - \sqrt{1-\alpha_t}\nabla_{\mathbf{x}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)) \tag{11}$$

The computation of the Jacobian $\nabla_{\mathbf{x}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ is both time- and memory-intensive, particularly for high dimensional images and deep neural networks. We have experimentally observed that the Jacobian is well-approximated by a diagonal matrix for each time step, and the diagonal elements are of limited range. Therefore, we approximate the Jacobian by a constant (which could be further absorbed into the step size). Incorporating both the jumpstarted sampling and the gradient approximation, we propose the jumpstarted spectral DPS (JSDPS), which is outlined in Algorithm 2 below.

Algorithm 2 Jumpstarted Spectral DPS (JSDPS)

T, T': training steps, sampling steps, $T' \ll T$
η_t: step size
$\mathbf{y}$: spectral CT measurement
$\hat{\mathbf{x}}_0^f$: image-domain decomposition
$\boldsymbol{\epsilon} \sim \mathcal{N}(0, I)$
$\mathbf{x}_{T'} = \hat{\mathbf{x}}_0^{T'} = \sqrt{\bar{\alpha}_{T'}}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_{T'}}\boldsymbol{\epsilon}$
for $t = T'$ to 1 **do**:
  $\mathbf{z} \sim \mathcal{N}(0, \boldsymbol{I})$
  $\hat{\mathbf{x}}_0 = \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t))$
  $\mathbf{x}'_{t-1} = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{x}_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}\hat{\mathbf{x}}_0 + \sigma_t\mathbf{z}$
  $\mathbf{x}_{t-1} = \mathbf{x}'_{t-1} - \eta_t\nabla_{\hat{\mathbf{x}}_0}\|\mathbf{BS}\exp(-\mathbf{QA}\hat{\mathbf{x}}_0) - \mathbf{y}\|^2_{\mathbf{K}^{-1}}$
end for

C. Unconditional Multi-Material Generation

1) Dataset Generation: This work focuses on two material decomposition with water and calcium as bases. To build spectral CT and material density dataset, we collected 32000 clinical chest CT slices from the CT Lymph Nodes dataset[20]. Those images are pre-processed to convert Hounsfield Units to attenuation coefficients and remove patient beds. An upper bound value (2000HU) was also applied to remove large attenuation values from metal. Two soft threshold functions were applied to approximate water and calcium densities (unit: g/ml):

$$\rho_w(\mu) = \begin{cases} k_w\mu, & \text{if } \mu \le \mu_w \\ k_w\mu_w - k_{wc}(\mu - \mu_w), & \text{if } \mu_w < \mu < \mu_c \\ 0, & otherwise \end{cases} \tag{12a}$$

$$\rho_c(\mu) = \begin{cases} k_c(\mu - \mu_c) + k_{cw}(\mu_c - \mu_w), & \text{if } \mu \geq \mu_c \\ k_{cw}(\mu - \mu_w), & \text{if } \mu_w < \mu < \mu_c \\ 0, & otherwise \end{cases} \quad (12b)$$

In this work, k_w, k_{wc}, k_{cw}, k_c were empirically set to 5.18, -8.77, 5.69, 2.12 g/cm^2, while μ_w and μ_c were 0.22 cm^{-1} and 0.35 cm^{-1}. The water and calcium densities were used as ground truth to simulate spectral measurements according to Eq.8.

2) Network Training: DDPM[16] is employed as the SGM framework, with the Residual Unet[21] as the backbone network. Paired water and calcium images are concatenated to form the 2 channel image $\mathbf{x}_0$. The discretized diffusion process uses $T = 1000$ time steps with a linear variance scheduler from $\beta_1 = 1e^{-4}$ to $\beta_{1000} = 0.02$. Training and validation datasets contain 25000 slices and 5000 slices, respectively. The loss function (7) is minimized by the Adam optimizer with a batch size of 16 and a learning rate of 10^{-4}. Training terminated after 200 epochs.

D. Conditional Posterior Sampling for Different Spectral Systems

The unconditional DDPM training captures the material bases distribution without specification of the spectral CT device. The trained model can be used in material decomposition for arbitrary spectral CT systems. Here we investigate the SDPS/JSDPS performance on a simulated dual-layer CT [14] and dual-kVp CT system [13]. The dual-layer CT uses a detector with $300\mu m$ CsI and $600\mu m$ CsI scintillators in the top and bottom layers, respectively (with a 5mm gap between layers). The x-ray tube operates at 120 kVp. The dual-kVp system uses single-layer detector with $600\mu m$ CsI. The tube voltage alternates between 80 kVp and 120 kVp every other view. For both systems, 800 projections were simulated with Poisson noise equivalent to an exposure of 0.05 mAs/view. Voxel size and detector pixel size were set to 0.8 mm and 1.0 mm, respectively.

E. Evaluation

We first evaluated the performance of JSDPS with and without gradient approximation. A standard image-domain decomposition from filtered backprojection reconstructions, i.e., the initialization of the JSDPS algorithm, was also shown for comparison. Second, we compared the performance of several material decomposition algorithms including model-based material decomposition (MBMD) [3], SDPS, and JSDPS. MBMD was formulated as the following optimization problem:

$$\hat{\mathbf{x}} = \text{argmin} \|\mathbf{BS}\exp(-\mathbf{QAx}) - \mathbf{y}\|^2_{\mathbf{K}^{-1}} + \lambda_w \mathbf{R}(\mathbf{x}_w) + \lambda_c \mathbf{R}(\mathbf{x}_c) \quad (13)$$

where $\mathbf{R}$ is a quadratic gradient roughness penalty. Regularization strengths, λ_w, λ_c, for water and calcium were set to 10^{-4} and 4×10^{-4}, respectively. We used 10000 iterations of separable paraboloidal surrogates updates, and final change in loss is $\sim 0.01\%$ from the previous iteration. SDPS was implemented according to Algorithm 1, and the step size scheduler follows [11]: $\eta_t = \eta / \|\mathbf{BS}\exp(-\mathbf{QAx}) - \mathbf{y}\|_{\mathbf{K}^{-1}}$. We swept η from 0.1 to 10, and the optimal η was determined as the one with minimal MSE from ground truth. JSDPS used a constant step scheduler with $T' = 150, \eta_t = 0.02$. Performance was evaluated for computation time (calculated for eight parallel decompositions), and image quality - quantified by Structural Similarity Index Measure(SSIM) and Peak Signal-to-Noise Ratio (PSNR). Additionally, to compare variability and biases in the posterior samples provided by SDPS and JSDPS, we compute sample bias and standard deviation over an ensemble of 16 outputs:

$$\text{Bias} = \|\mathbb{E}\{\hat{\mathbf{x}}\} - \mathbf{x}\|_2 \quad \text{Std} = \|\mathbb{E}\{(\hat{\mathbf{x}} - \mathbb{E}\{\hat{\mathbf{x}}\})^2\}\|_2 \quad (14)$$

III. Results

A. Illustration of JSDPS variants

Fig.2 displays JSDPS results as compared with standard image-domain decomposition (initialization) and the ground truth. Image-domain decomposition shows significant image noise and streaking. JSDPS without gradient approximation effectively captures most of the high-contrast structures, including smaller pulmonary vessels. However, as seen in the zoomed-in ROIs, low-contrast soft tissue hallucinations are present. In the contrast enhanced areas of the calcium image, such as the aorta, deformations are also observed. By using the gradient approximation, the reverse sampling achieves 20.2% and 29.3% computation time decrease for the dual-layer and dual-kVp system, respectively. Gradient approximation also effectively mitigates the hallucinations and enhances the consistency with ground truth for vascular structures.

B. Comparison of SDPS, JSDPS, and MBMD

1) Image Quality: Figure 3 compares the performance of SDPS, JSDPS, and MBMD in water and calcium bases for the dual layer and dual kV systems. Four zoomed in areas are shown in the (1) lung, (2) heart (iodine-filled aorta in the calcium basis), (3) low-contrast soft tissue, and (4) spine. Comparing across all three algorithms, MBMD produces images with the highest noise (evident in ROIs 2 and 3) at the lowest spatial resolution (evident in ROI 1). The SDPS results have low noise and high spatial resolution, but are prone to hallucinations observed as erroneous anatomy in all four ROIs. While SDPS produces results that are more closely aligned with the prior distribution, the instability of the algorithm is problematic for the low fidelity (low dose) data simulated in this work. The JSDPS algorithm outperforms both MBMD and SDPS from both visual observations of anatomical structures faithful to the ground truth and according to SSIM and PSNR. Good decomposition results were obtained for both the dual layer and dual kV system, demonstrating generalizability to different spectral systems.

Figure 4 shows the bias and standard deviation maps computed from an ensemble of 16 individual output samples. SDPS exhibits large variability and bias around edge features, suggesting substantial image uncertainty. In contrast, JSDPS displays markedly reduced bias and standard deviation, demonstrating the capability to effectively stabilize the sampling process.

2) Computational Cost: Table I summarizes the computational cost of different material decomposition algorithms for eight output estimates. MBMD is the slowest due to the large number of iterations (10000) required for convergence. SDPS substantially reduces the number of iterations (1000), thereby shortening the computation time. Leveraging the jumpstarted sampling strategy and the gradient approximation, JSDPS requires only 150 iterations, which results in an additional 86.83% and 88.07% reduction in computation time for dual-layer and dual kVp CT, respectively. Furthermore, a 23.16% and 42.54% memory saving was also achieved compared to SDPS for the two spectral systems due to the absence of Jacobian computation.

TABLE I
COMPUTATIONAL COST OF MATERIAL DECOMPOSITION ALGORITHMS

	Dual layer		Dual kVp	
	Time(s)	Memory(GB)	Time(s)	Memory(GB)
MBMD	6406.3	11.3	2802.5	4.8
SDPS	728.2	17.7	478.6	13.4
JSDPS	95.9	13.6	57.1	7.7

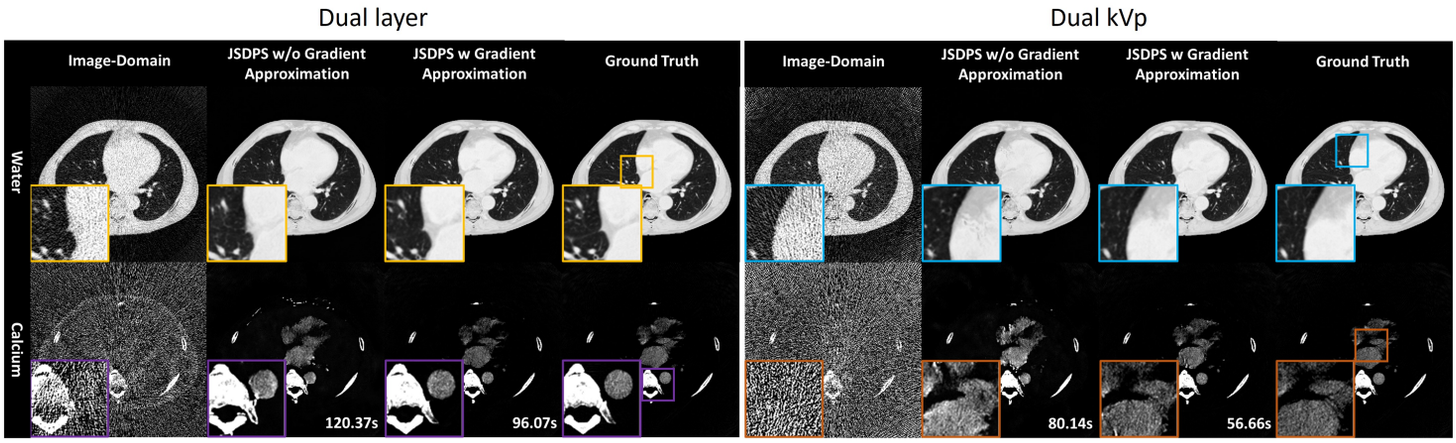


Fig. 2. Decomposed water (top, W/L: $1.2/0.6\ g/ml$) and Calcium (bottom, W/L: $0.05/0.1\ g/ml$) images. Left to right: Image-domain decomposition, JSDPS without gradient approximation, JSDPS with gradient approximation, ground truth. Computational time for eight outputs is provided in the bottom left corner.

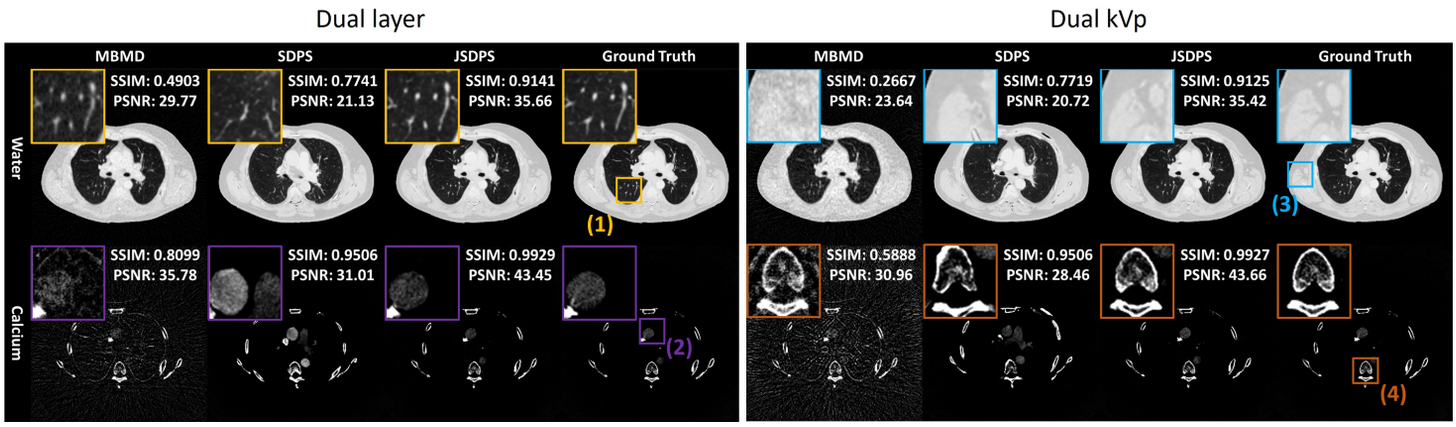


Fig. 3. Decomposed water (top, W/L: $1.2/0.6\ g/ml$) and Calcium (bottom, W/L: $0.05/0.1\ g/ml$) for (left to right): MBMD, SDPS, JSDPS, ground truth.

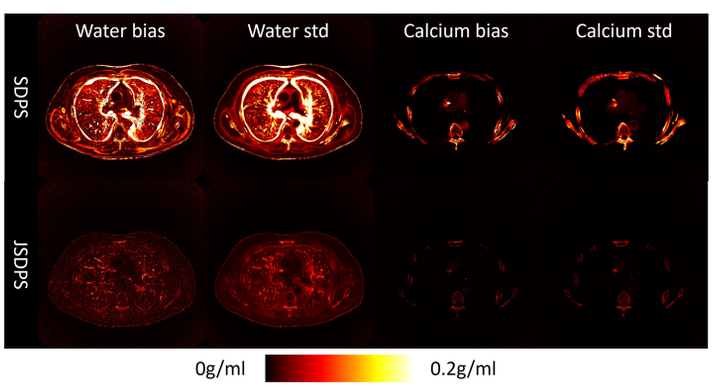


Fig. 4. Bias and standard deviation map for SDPS and JSDPS. Here we only show the results of dual-kVp system, which is similar to that of the dual-layer system.

IV. Conclusion and Discussion

Spectral CT material decomposition is a challenging and ill-conditioned problem, often exhibiting excessive noise and slow convergence. The novel SDPS framework is designed to tackle these issues - combining a sophisticated learned SGM prior and a physical model, to achieve low image noise while enhancing image quality. We additionally adopted jumpstarted sampling and gradient approximation strategies which were found to be effective in stabilizing performance as well as reducing computation time and memory consumption. In evaluations on both dual-layer and dual-kVp systems, JSDPS achieves the highest accuracy, the lowest uncertainty, and the lowest computational costs compared to SDPS and MBMD. This work demonstrated that JSDPS is a promising approach for spectral CT decomposition.

Acknowledgments

This work is supported, in part, by NIH grant R01EB030494.

References

[1] Long, Y and Fessler, J. A, *IEEE transactions on medical imaging*, vol. 33, no. 8, pp. 1614–1626, 2014.
[2] Jiang, X *et al.*, *Medical physics*, vol. 48, no. 9, pp. 4843–4856, 2021.
[3] Tilley, S *et al.*, *Physics in Medicine & Biology*, vol. 64, no. 3, p. 035005, 2019.
[4] Gong, H *et al.*, *Medical physics*, vol. 47, no. 12, pp. 6294–6309, 2020.
[5] Wu, W *et al.*, *Neural Networks*, vol. 144, pp. 342–358, 2021.
[6] Abascal, J. F *et al.*, *IEEE Access*, vol. 9, pp. 25 632–25 647, 2021.
[7] Clark, D. P *et al.*, in *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573. SPIE, 2018, pp. 415–423.
[8] Zhu, J *et al.*, *Physics in Medicine & Biology*, vol. 67, no. 14, p. 145012, 2022.
[9] Fang, W *et al.*, *Physics in Medicine & Biology*, vol. 66, no. 15, p. 155013, 2021.
[10] Eguizabal, A *et al.*, *arXiv preprint arXiv:2208.03360*, 2022.
[11] Chung, H *et al.*, *arXiv preprint arXiv:2209.14687*, 2022.
[12] Li, S *et al.*, *arXiv preprint arXiv:2312.01464*, 2023.
[13] Cassetta, R *et al.*, *Physics in Medicine & Biology*, vol. 65, no. 1, p. 015013, 2020.
[14] Wang, W *et al.*, *Medical physics*, vol. 48, no. 10, pp. 6375–6387, 2021.
[15] Song, Y *et al.*, *arXiv preprint arXiv:2011.13456*, 2020.
[16] Ho, J *et al.*, *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
[17] Song, J *et al.*, *arXiv preprint arXiv:2010.02502*, 2020.
[18] Stenner, P *et al.*, *Medical physics*, vol. 34, no. 9, pp. 3630–3641, 2007.
[19] Mendonça, P. R *et al.*, *IEEE transactions on medical imaging*, vol. 33, no. 1, pp. 99–116, 2013.
[20] Roth, H. R *et al.*, in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014: 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part I 17*. Springer, 2014, pp. 520–527.
[21] Jiang, X *et al.*, in *7th International Conference on Image Formation in X-Ray Computed Tomography*, vol. 12304. SPIE, 2022, pp. 488–492.

Multi-Conditioned Denoising Diffusion Probabilistic Model (mDDPM) for Medical Image Synthesis

Arjun Krishna, Ge Wang, *Fellow, IEEE,* and Klaus Mueller, *Fellow, IEEE*

***Abstract*—Medical imaging applications are highly specialized in terms of human anatomy, pathology, and imaging domains. Therefore, annotated training datasets for training deep learning applications in medical imaging not only need to be highly accurate but also diverse and large enough to encompass almost all plausible examples with respect to those specifications. We argue that achieving this goal can be facilitated through a controlled generation framework for synthetic images with annotations, requiring multiple conditional specifications as input to provide control. We employ a Denoising Diffusion Probabilistic Model (DDPM) to train a large-scale generative model in the lung CT domain and expand upon a classifier-free sampling strategy to showcase one such generation framework. We show that our approach can produce annotated lung CT images that can faithfully represent anatomy, convincingly fooling experts into perceiving them as real. Our experiments demonstrate that controlled generative frameworks of this nature can surpass nearly every state-of-the-art image generative model in achieving anatomical consistency in generated medical images when trained on comparable large medical datasets.**

***Index Terms*—DDPM, Computed Tomography, Generative AI**

I. Introduction

GREAT strides have been made in deep learning-based medical applications; however, their potential remains constrained by the scarcity of specialized, highly accurate, high-resolution annotated images suitable to robustly train these learning models. To address this limitation, researchers have explored image synthesis to augment the existing datasets, demonstrating that such methods can generate convincingly realistic medical images [1], [3], [4], [12].

Yet, the generation of phantom images at full resolution with flawless anatomy remains to be a formidable challenge, particularly when incorporating annotations [3], [4], [12]. This process is prone to introducing anatomical errors as the generation is constrained by these annotations. Existing methods that achieve partial success in generating full-resolution CT images, capable of deceiving radiologists, predominantly rely on applying unconditional state-of-the-art image generative models [1] to large medical datasets. However, these approaches lack purposeful and diversified generative capabilities, merely producing more non-annotated raw medical images adding to the already abundant general datasets.

In this paper, we introduce a methodology that enables the dynamic application of a series of annotations and constraints during the generation process. Our approach simultaneously generates state-of-the-art, full-resolution CT images, passing our Visual Turing Test and exhibiting superior performance compared to other unconditional state-of-the-art image generative models. To our knowledge, our work represents the first endeavor capable of producing full-resolution CT images with accompanying annotations that maintain anatomical accuracy across all clinically relevant Hounsfield Unit (HU) windows.

In our prior work [4], we detailed a method for generating unique and diverse annotated CT lung images to construct balanced datasets. It depended on the independent modeling of annotations, with the generative GAN-based models intricately connected to these annotations. Recently, DDPMs [8] have emerged as an alternative to traditional image generative models. Given that trained DDPMs sample images through denoising, researchers have devised unique methods to iteratively guide the sampling process towards specific areas of underlying image distributions [5], [9], [11]. This approach resembles the conditional generation in GANs but offers the added benefit that such generative models are not tethered to an underlying modality for condition or guidance.

In this paper, we explore a form of conditional generation [9] and extend it to encompass multiple annotations/conditions simultaneously as guidance and control. We demonstrate that not only does combining such annotations not depreciate synthesis quality, it also surpasses certain state-of-the-art unconditional image generative models. We show that this new method eradicates most anatomical inaccuracies and successfully passes our previously designed [4] Visual Turing Test.

II. Methods

We start out with a large dataset comprising low dose CT images from various scanners and train a DDPM based on the refinements proposed by Nichol et al. [10]. Subsequently, we investigate the sampling strategy of these trained DDPMs suggested by Choi et al. [9], and extend it to incorporate multiple conditional or guidance images. Our findings highlight the significance of this strategy for the purpose of synthesizing medical imaging datasets that are not only highly accurate but also annotated. Moreover, as these guidance techniques are not bound by annotations, they can be effectively employed to enhance annotated images featuring rare anatomies and pathology, thereby fostering the development of a more comprehensive and diversified dataset.

A. DDPM

The DDPM we implemented is a Markov Chain model which iteratively converts an isotropic Gaussian distribution

Arjun Krishna is in the Computer Science Department, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: arjkrishna@cs.stonybrook.edu).

Ge Wang is with the Biomedical Imaging Center, Center of Biotechnologies and Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY 12180 USA (e-mail: wangg6@rpi.edu).

Klaus Mueller is with the Computer Science Department, Stony Brook University, Stony Brook, NY 11794 USA (e-mail: mueller@cs.stonybrook.edu).

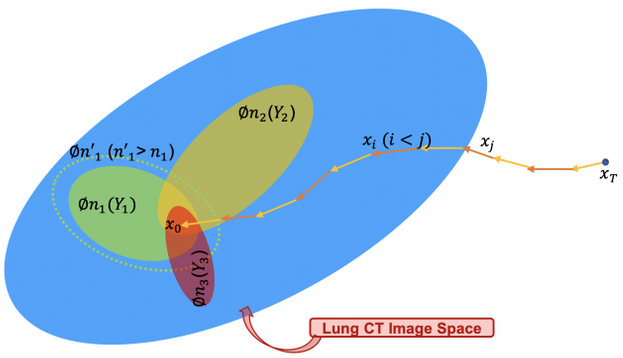


Fig. 1. Multi-Conditioned Guided Sampling. The blue area represents the image space for all CT lung images; the yellow, green and red circles represent the image space closer to the three guidance images y1, y2 and y3, the size of the circles depends on the images themselves and the downsampling factors n1, n2, n3 of the filter used corresponding to these images.

into a full Hounsfield window lung CT image data distribution. The Markov Chain model learns the reverse of the forward diffusion process, a fixed Markov Chain that gradually adds noise to the data in the opposite direction of sampling until the signal is destroyed. This forward process is described as:

$$q(x_t|x_{t-1}) := N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

where x_1,...,x_T are the latents produced by the addition of noise and β_1,...,β_T follow a fixed variance schedule. Eq. 1 can be decomposed by the reparameterization trick and x_t can be further derived in terms of the image x_0 as:

$$x_t = \sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon \tag{2}$$

where α_t := 1 - β_t and $\overline{\alpha}_t := \prod_{i=1}^{t} \alpha_i$ and the added noise $\epsilon \sim N(0, I)$ has the same dimensionality as the image and the sampled latents during training. The reverse diffusion process that our model needs to learn is expressed [8] as:

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I) \tag{3}$$

where p_θ is a neural network to predict μ_θ and μ_θ is further decomposed [8] in terms of noise approximator ϵ_θ:

$$\mu_\theta = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)) \tag{4}$$

By formulating the loss function [8] as the log likelihood of x_0 and computing a variational lower bound (similar to the case of variational auto-encoders) as KL divergence between q and p, the authors [8] decided to frame the loss function as the L2 distance between actual mean of the image(μ) and μ_θ which can be further simplified to as the L2 distance between the predicted noise ϵ_θ and added noise ϵ at any given time t

$$Loss = \|\epsilon - \epsilon_\theta(x_t, t)\|^2 \tag{5}$$

or

$$Loss = \|\epsilon - \epsilon_\theta(\sqrt{\overline{\alpha}_t}x_0 + \sqrt{1-\overline{\alpha}_t}\epsilon, t)\|^2 \tag{6}$$

Eqs. 2 and 6 are used to train our DDPM, incorporating refinements from Nichol et al. [10]. Our DDPM was trained on a large dataset of 5,000 lung CT scans, with images extracted at full HU width of 2000. This ensures that the generated images span the entire width during sampling and can be visualized at other clinically relevant windows, including lung, bone, and soft-tissue. Utilizing Eq. 3 and the reparameterization trick, x_{t-1} can be sampled as:

$$x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)) + \sqrt{\beta_t}\epsilon \tag{7}$$

Using the above equation repeatedly, we can sample lung CT images starting from random noise after training a DDPM on our large dataset. Both training and sampling steps are outlined in prior works related to DDPMs [8]. Next, we will focus on the sampling algorithm of our DDPM to facilitate multi-annotations guidance during our lung CT image generation.

B. Multi-Condition Guidance

As mentioned earlier, various methods exist for guiding the sampling process of a trained DDPM. In this section, we delve into the guidance techniques presented by Choi et al. [9] and leverage them for precise control over the generation of lung CT images across all HU windows. Choi et al. posit that it should be feasible to guide the sampling process to a subset of image distributions around a reference image y if we can ensure similarity between the downsampled reference image y and the downsampled generated image x_0.

Algorithm 1: Sampling

Input: Conditional / guidance images $y_1, ...y_M$
Output: Generated image x
Filter-scales: $\phi_{n_1}, ...\phi_{n_M}$
$x_T \sim N(0, I)$
initialization;
for $t = T$ **to** 1 **do**
  $z \sim N(0, I)$
  if t = *1* **then**
    z = 0
  $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\epsilon_\theta(x_t, t)) + \sigma_t z$
  for $s = 1$ **to** M **do**
    $y_{s_{t-1}} \sim q(y_{s_{t-1}}|y_s)$
    $x_{t-1} \leftarrow \phi_{n_s}(y_{s_{t-1}}) + x_{t-1} - \phi_{n_s}(x_{t-1})$
return x_0

In order to approximate this condition in every Markov transition during sampling, Choi et al. continuously refine the downsampled latent variable x_t to be similar to the corresponding downsampled noisy version of reference image y_t, to ensure that both x_t and y_t share low frequency contents. y_t is computed from reference image y using Eq. 2 during sampling for every Markov transition. Specifically:

$$p_\theta(x_{t-1}|x_t, c) \approx p_\theta(x_{t-1}|x_t, \phi_N(x_{t-1}) = \phi_N(y_{t-1})) \tag{8}$$

where $\phi_N(\dots)$ is a low-pass linear filter (with N as the downsampling factor), and the term is approximated by ensuring the latent x_{t-1} captures the missing low-frequency contents of y_{t-1} after sampling from the unconditional DDPM.

$$x_{t-1} = x_{t-1} + \phi_N(y_{s_{t-1}}) - \phi_N(x_{t-1}) \tag{9}$$

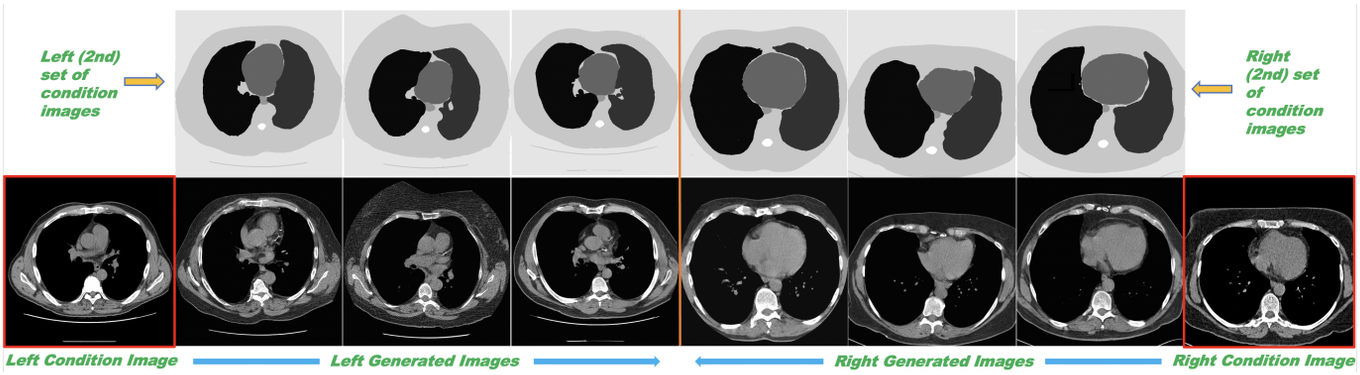


Fig. 2. This figure shows six examples of lung CT soft-tissue window image generations with two conditional images. Both left and right sections display three generated images for three different anatomy / segmentation maps for the same reference (conditional) CT image, shown in the red boxes. The generations follow the anatomy of the segmentation maps above but exhibit the slice of the heart generation corresponding to the referenced CT images. The results are displayed in the soft-tissue window to highlight the similarity and accuracy of the generated anatomy w.r.t guidance images.

We contend that by controlling the extent of low-pass filtering (factor N) of a linear filter ϕ for a given set of conditional or guidance images $y_1, y_2, \ldots, y_m$, we can fine-tune our algorithm using a set of integers $n_1, n_2, \ldots, n_m$. Here, each integer represents the extent of downsampling for a linear filter corresponding to each conditional image. This allows for valid image generation through a trained DDPM that shares low-level features (or similarity) with each of the conditional images. We modify Eq. 9 as:

$$x_{t-1} = x_{t-1} + \sum_{s=1}^{M} (\phi_{n_s}(y_{s_{t-1}}) - \phi_{n_s}(x_{t-1})) \qquad (10)$$

The downsampling factor n_s for a conditional image will depend on the purpose and the nature of the conditional image in the generation of final images. In practice, the above strategy may only work well for a maximum of three or four conditional images. Fig. 1 visualizes our multi-conditional guidance and the steps in sampling where with each step the generated image gets closer to the desired super-subset of the image distribution. As is evident from the visualization; if integers $\{n_1, n_2 \ldots n_m\}$ are not chosen carefully, there may not be a significant overlap between the subset distributions of conditional images in which case the image samplings may start generating inaccuracies in generated images. Steps 11 - 13 in Algorithm 1. illustrate the above process in our sampling of the synthetic lung CT images.

III. Experiment Setup and Results

We trained our DDPM [10] on a dataset of (low-dose) lung CT-Scans of 5,000 patients. The images from the scans were extracted from the mid-abdomen regions, clearly showing the lungs along with the heart. The images were extracted in the entire relevant width of 2000 HU (-1000 HU to 1000 HU) for training our model; which enables the generation of images in the same HU range during the sampling process post training. That way, the images can be viewed at any HU window during their evaluation in our Visual Turing Test.

Figs. 2 and 3 showcase images generated with our model. Fig. 2 illustrates sets of guidance images, each comprising an anatomy map and a CT image. These sets serve as conditional (guidance) images for each of the 6 generated images, shown across the center bottom of the figure (see the caption for more detail). The results reveal that diverse anatomically accurate versions of a single CT image can be generated when annotations for the anatomy are available. Here, we generate anatomy maps using B-splines, as detailed in our prior work [4]. It is noteworthy that this approach can be easily extended to simulate pathology/pathology types given a few annotated examples of CT images depicting such pathology. Fig. 3 presents the original full HU window generated images outlined with a red box, along with their decomposition in other clinically relevant windows. Visual inspections affirm their anatomical accuracy in each window, underscoring the effectiveness of DDPMs when assisted by guidance images in learning the nuances of anatomical structures across the entire HU range of lung CT images.

A. Comparisons with Other Generative Models

Table 1 shows comparative quantitative evaluations of a set of 10k generated full HU window lung CT images with the state-of-the-art image generative models namely NVIDIA's

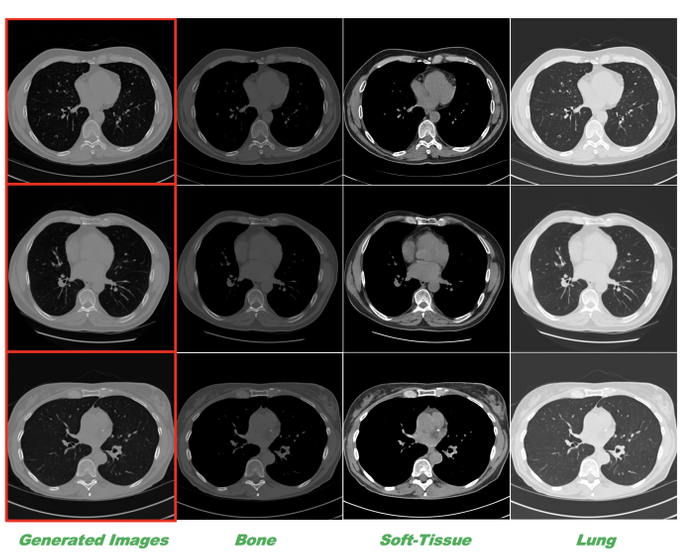


Fig. 3. Left-most column (outlined with a red box): images generated with our multi-conditional sampling algorithm, shown at full HU range. Other three columns: these images in their respective bone, soft-tissue, and lung windows.

TABLE I
COMPARING GENERATIVE MODELS.

	FID	Set-level SSIM
PGGAN	146.09	0.33
StyleGAN	81.57	0.31
StyleGAN2	72.31	0.30
Unguided Sampling	83.24	0.27
Guided Sampling	**69.85**	**0.45**

StyleGAN [6], StyleGAN2 [7] and PGGAN [2] that were trained on the same large dataset. We also compared these with the images generated from unguided sampling via the same trained DDPM to evaluate whether guided sampling has an effect on anatomical consistency apart from it being an annotated dataset generator. We chose the FID score because it measures the "realism" of a set of generated images.

The FID scores in Table 1 show that our model is at least almost as good quantitatively as the state-of-the-art StyleGAN2 if not better, and considerably better than the StyleGAN and the PGGAN. Additionally, unlike these models, our method is focused not only on just generating raw data but also its annotations. As such it could easily be expanded to generate CT scans with annotated pathology which is not possible in either of the above state-of-the-art models.

We also performed an exhaustive set-level comparison where we gauged the Structural Similarity Index (SSIM) of the generated images against the large training dataset to measure the overall similarity of the generated images with respect to the training set images. As shown in Table 1, our model scores higher in SSIM than both of the StyleGANs. On visual inspection, the set of images generated via unconditional state-of-the-art-models can produce accurate generations but are prone to generating odd anatomies due to the absence of an anatomy controlled generation framework. This could explain at least partially the reason for the lower SSIM scores. Finally, our sampling strategy also outperforms the unguided sampled generations via the trained DDPM.

B. Visual Turing Test

We reran our Visual Turing Test [4] with the assistance of three radiologists to evaluate the realism of our generated lung CT images. As previously, the test was administered to the radiologists by presenting them (via a web browser app) with a randomly selected lung CT image from a balanced set of 30 real and generated images, one at a time, in random order. The images were randomly chosen from bone, lung, and soft-tissue windows. For each image there were two options: "Real" or "Fake."

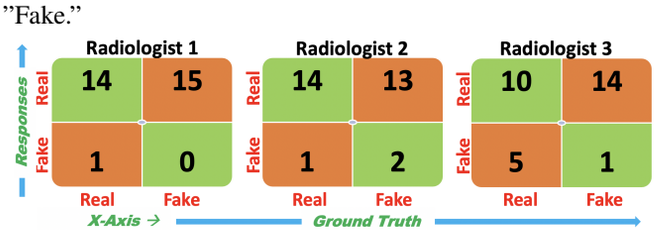


Fig. 4. Confusion matrices for the responses of the 3 radiologists. The overall accuracy of the responses is 45.56% which is close to 50%; a requirement for passing our Visual Turing Test. Proportion of 'True Negatives' (fake images identified as fake) is 6.67% whereas proportion of 'False Negatives' (real images identified as fake) is 15.56%

The test assesses if our model is successfully able to generate medically accurate images. This is determined by measuring the number of times the model is able to fool the experts into thinking that a model generated image is a medical image obtained from a real patient. When experts are unable to separate the images into real or fake at least 50% (chance baseline) of the time, the model is said to have passed the visual Turing test. The test was taken by the same three radiologists as in our previous study. Their responses are compiled in Fig. 4.

The numbers in Fig. 4 indicate that the generative framework presented in this paper (unlike the one presented in our previous work [4], [12]) has passed the Visual Turing Test, as expert radiologists could not identify most of the fake (synthesized) lung CT images from the real ones. Upon analyzing the responses, we found that most responses were marked as 'Real' since the radiologists did not know that half of the shown images were 'Fake' and only marked an image as 'Fake' if they thought there was an anatomical/texture anomaly in the generated image. Even then, most of their responses labeled 'Fake' were, in fact, for the real images, showing that our guidance-based DDPM sampling scheme clearly passed the Visual Turing Test (see the caption for more details).

IV. CONCLUSIONS AND FUTURE WORK

Having demonstrated that our methodology can synthesize realist CT images with anatomical guidance, future work will extend this guidance to the synthesis of realistic pathology.

ACKNOWLEDGMENT

This research was funded in part by NIH grant R01EB032716. We thank Dr. David Yankelevitz and Dr. Yeqing Zhu from the Icahn School of Medicine at Mount Sinai, NY for their help in image evaluation.

REFERENCES

[1] H. Y. Park et al., Realistic High-Resolution Body Computed Tomography Image Synthesis by Using Progressive Growing Generative Adversarial Network: Visual Turing Test, JMIR Medical Informatics **9** (2021).
[2] E. L. Denton et al., Deep generative image models using a laplacian pyramid of adversarial networks, NIPS **28** (2015).
[3] Han, K. et al., MedGen3D: A Deep Generative Framework for Paired 3D Image and Mask Generation, MICCAI 2023.
[4] A. Krishna, S. Yenneti, G. Wang and K. Mueller, Novel Lung CT Image Synthesis at Full Hounsfield Range With Expert Guided Visual Turing Test, in *Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2023.
[5] H. Chung et al., Solving 3D Inverse Problems Using Pre-Trained 2D Diffusion Models, CVPR 2023.
[6] T. Karras, S. Laine, and T. Aila, A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR , 4396–4405 (2018).
[7] T. Karras, S. Laine, and T. Aila, Analyzing and improving the image quality of stylegan, CVPR , 8110–8119 (2020).
[8] J. Ho, A. Jain, and P. Abbeel, Denoising diffusion probabilistic models, NIPS (2020).
[9] J. Choi, S. Kim, Y. Jeong, Y. Gwon, and S. Yoon, ILVR: Conditioning Method for Denoising Diffusion Probabilistic Models, ICCV (2021).
[10] A. Quinn Nichol and P. Dhariwal, Improved Denoising Diffusion Probabilistic Models, openreview.net/forum?id=-NEXDKk8gZ, 2021.
[11] J. Ho and T. Salimans, Classifier-Free Diffusion Guidance, arxiv.org/abs/2207.12598, 2022.
[12] A. Krishna, S. Yenneti, G. Wang and K. Mueller, Image factory: A method for synthesizing novel CT images with anatomical guidance, in *Medical Physics*, 2023.

Integration of Prior Knowledge in Application of Neural Networks for Metal Artifact Reduction in CT

Nele Blum[1], Thorsten M. Buzug[1,2] and Maik Stille[1]

***Abstract*—Metal artifacts remain one of the major challenges in computed tomography. Despite decades of research, today's state of the art procedures are often unable to adequately correct the artifacts and frequently cause secondary artifacts. With the increasing success of deep learning methods, they are also gaining popularity in the field of metal artifact reduction. However, many methods neglect information, that are present in the area of the metal object, which increases the risk of replacing the artifacts with seemingly plausible but incorrect image information. With the development of a new weighted partial convolution in the network layers, we have succeeded in integrating this information into the network application. Moreover, we were able to significantly improve the results by applying a normalization step, which allows additional structural information to be integrated into the networks training.**

***Index Terms*—CT, metal artifact reduction, deep learning.**

I. Introduction

COMPUTED tomography (CT) has been a fundamental medical imaging technique for decades, with ongoing efforts to improve image quality while reducing radiation exposure and acquisition time. However, challenges persist, especially in scans involving high-density objects like metal implants. Many metal artifact reduction (MAR) techniques, including projection completion methods [1], are often not able to reduce the artifacts sufficiently or introduce inconsistencies and unwanted artifacts due to limited information. Recently neural networks, integrated into iterative reconstructions or used in pre- or post-processing, have gained popularity for metal artifact correction. While existing methods often rely on the use of ground truth data to train the networks, we propose an approach for MAR, which allows learning the distribution of reconstructed images in a completely unsupervised setting by integrating a stochastic component that introduces an explicit dependency between the input data and the generated output in the learning process based on the method proposed by Prajot et al. [2]. First results on simulated data showed compared to a supervised setting that similar results could be achieved [3]. However, major drawbacks are often the complete loss of information in the metal trace and the difficulty of integrating prior information into the learning process. In contrast to many inpainting problems in image processing, where there is no information left in the data which is to be replaced, in the metal trace this information is some times still available. Particularly, for projections that pass only through a small area of the object or through thin objects, a significant amount of information about surrounding anatomical structures may still remain in the data. Some iterative methods [4] already use prior information of the metal object in the reconstruction process, yet they often have high computation times, especially for large 3D-volumes. At the same time conventional convolutions in CNNs are increasingly being adapted to specific problems. One example is the partial convolution for the inpainting task developed by Liu et al. in 2018 [5]. By applying a binary mask, only image values outside the missing areas are considered as valid. During the convolution, the mask is gradually filled, such that at the end the mask is entirely set to one and all data are used. A previously published work by us was able to show that partial convolution can also be used in the preprocessing of CT projection data. This approach achieves better results than conventional convolutions [6]. However, to use the residual information in the metal trace, it has to be adjusted. We were able to integrate this information in the learning process, such that the data in the metal trace is not weighted with zero, but depending on the prior knowledge about the path of the projections through the metal object. In addition, methods, such as normalized metal artifact reduction (NMAR) [7], already use prior information in the form of a prior image. By integrating this information, the results of sinorgram completion methods could be significantly improved. To use this information, we have integrated the approach into our network training.

N.Blum and M.Stille are with the Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering IMTE, Lübeck, 23556 Germany, e-mail: (see nele.blum@imte.fraunhofer.de).

T.M.Buzug is with the Fraunhofer Research Institution for Individualized and Cell-Based Medical Engineering IMTE, Lübeck, 23556 Germany and the Institute of Medical Engineering, University of Lübeck, 23556 Germany

II. Materials and Methods

We solve the problem of replacing the metal trace in the projection data by applying a generative adversarial network (GAN) architecture. The generator receives the incomplete sinogram data and the corresponding metal trace as an input. Discriminator networks than assign a probability between $[0, 1]$, indicating whether the data are real or artificially generated. Compared to the original input data or the data from corresponding neighbor slices, respectively. The unsupervised training approach is based entirely on the methodology already presented by us in Blum et al. [3]. However, it is possible to replace it with other training methods and network architectures, which makes it easy to integrate the proposed approach into already developed methods.

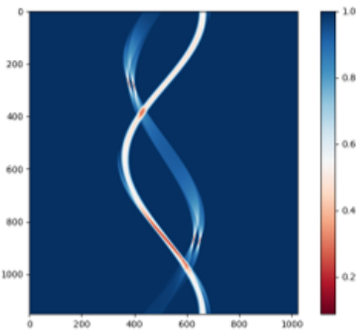


Fig. 1: Example weighted mask of the metal trace used in the weighted PC layer of the neural network.

In our previous work, the corrupted data in the metal trace were considered as missing and were entirely replaced by the network. To integrate the information from the metal trace into the networks training, the convolutional layers of the generator were replaced by partial convolutional (PC) layers. In contrast to a conventional convolution, the PC multiplies the previous feature map X with a binary mask M, resulting in the new output value

$$X' = \begin{cases} W^\intercal (X \cdot M) r(M), & \text{if } ||M||_1 > 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

with

$$r(M) = \frac{\text{size}(M)}{M} \tag{2}$$

of the convolutional layer where $r(M)$ is a correction factor depending on the size of the binary mask M. After each convolution M is updated to one at the positions for which x' is greater zero. Thus, the number of invalid values is reduced and the missing data are replaced layer by layer.

However, since the PC layer described before still neglects the information in the metal trace by setting the values within to zero, it was adapted to the specific problem of using the remaining information in the metal trace. Figure 1 shows an example of the weighting mask, which can be used instead of using a binary mask in the regular partial convolution. Here, the values within the metal trace were additionally weighted in the range between $[0, 1]$. Prior information about the path of the projections through the metal object were used for the weighting in the metal trace via a previous generated segmentation of the object in the image space and a subsequent forward projection. The new mask M' is now calculated by

$$M' = \begin{cases} 1, & \text{if } \frac{W^\intercal(M)}{\max(W^\intercal(M))} > \epsilon \\ \frac{W^\intercal(M)}{\max(W^\intercal(M))}, & \text{otherwise} \end{cases} \tag{3}$$

where ϵ is a threshold, which controls the degree to which values in X' can be considered as valid for the application of subsequent convolution, and thus the number of PCs required to fill in the missing values in the input data. Figure 2 indicates how the data in the mask is updated by applying the weighted partial convolution, and thus gradually contributing more to the result of each the convolutional layer.

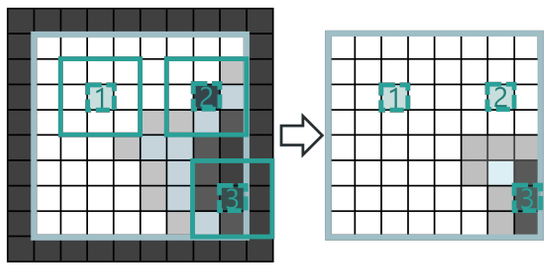


Fig. 2: Update of the weighted mask during the weighted partial convolution. Over each network layer more values becoming greater than ϵ and are considered as valid values in the end of the convolution process.

A. *Normalized GAN*

We assume that a simplified correction of the data can be performed if the data are more homogeneous. For this purpose, similar to NMAR, the data were first normalized using a forward projection of a previously calculated prior image consisting exclusively of the three base materials air, soft tissue, and bone. The resulting sinograms were used for training instead of the original data. To obtain the correct value range in the result, the data were denormalized using the forward projection of the prior image after applying the neural network. In the following the method supplemented by the normalization step is referred to as NGAN.

III. Training data

For the evaluation, simulated and measured data is used. Multiple software phantoms of $515 \times 512 \times 512$ with a voxel size of 1 mm and slice thickness of 1 mm are generated with varying parameter settings using the XCAT software phantom and a simulation of a 140 kVp polyenergetic beam [8]. Afterwards, a forward projection in a cone beam geometry is applied to generate artificial sinogram data with 1152 views over 360 degree. The source to isocenter distance, as well as the detector to isocenter distance are specified according to a clinical Somatom Definition AS scanner. Furthermore, different artificial objects were randomly added. The generated data were divided into training (90), validation (15), and test (15) data. Afterwards, the pre-trained networks on the simulated training data were fine-tuned on a smaller set of measurement data. For an analysis of the network performance on real measured data, a multi-energy CT Phantom (MECT) is used. The hardware phantom can be fitted with a various amount of metal rods at different positions. Hence, an image without metal rods can be acquired to serve as a ground truth in the image domain.

IV. Evaluation methods

Optimization of the network parameters was performed exclusively on the simulated sinogram data, as the ground truth of the artefact free projection data is available only here. Parameters included the patch size, learning rate, batch size, data augmentations, and early stopping. Since the best results were achieved for a batch size of 512×512, it was used as the standard for network training and application. Consequently,

TABLE I: training parameters

Parameter	Value
number of training data	183 000
number of validation data	31 000
batch size	12
patch size	512×512
number of epochs	250
learning rate	$1 \cdot 10^{-5}$

the batch size was limited by 12 by the GPU memory, which gave the best results compared to smaller batch sizes. Training was performed on an Nvidia RTX A6000 with 48 GB of memory. Training on the simulated data set was performed using the final training parameters listed in table I.

Since the ground truth in projection domain was only available for the simulated data, we focused on the evaluation of the reconstructed images. In addition to a simple linear interpolation in the projection domain, we implemented NMAR to evaluate the effectiveness of our proposed algorithm. For the evaluation two commonly used metrics, the Mean Squared Error (MSE) and Structural Similarity Index (SSI), were analysed. The MSE is a measure of the average squared difference between the ground truth and metal artifact reduced pixel values in an image. The SSIM is a metric that takes additionally into account luminance, contrast, and structure, providing a more comprehensive assessment of the image quality.

V. Results

A. Simulation Study

Figure 3 shows an example of the simulated phantoms from the test data set. A linear interpolation (LI), NMAR, the developed network, and the NGAN were applied to the data. Results are shown in the bone (top line) and soft tissue window (middle line), in addition to difference images (bottom line) in relation to the ground truth. In all representations, the image artifacts could be qualitatively reduced by both developed procedures compared to the LI and NMAR. This becomes particularly apparent in the difference image, where the NGAN shows a clear improvement compared to the other methods.

This is additionally confirmed in Figure 4 by the quantitatively calculated values of the MSE and SSIM. Again, the calculated values for LI (red) and NMAR (green) are often below and above the developed methods. However, while the GAN (blue) fluctuates to some extent in the calculated values, the NGAN (light green) is significantly better in both metrics. Table II shows the corresponding average values over the complete test images, where NGAN, followed by the original GAN, performs better than NMAR and LI across both metrics.

B. Hardware Phantom

Figure 5 illustrates the measured phantom with two metal inserts on opposite sides. This was first reconstructed by a weighted filtered back projection (WFBP) without applying

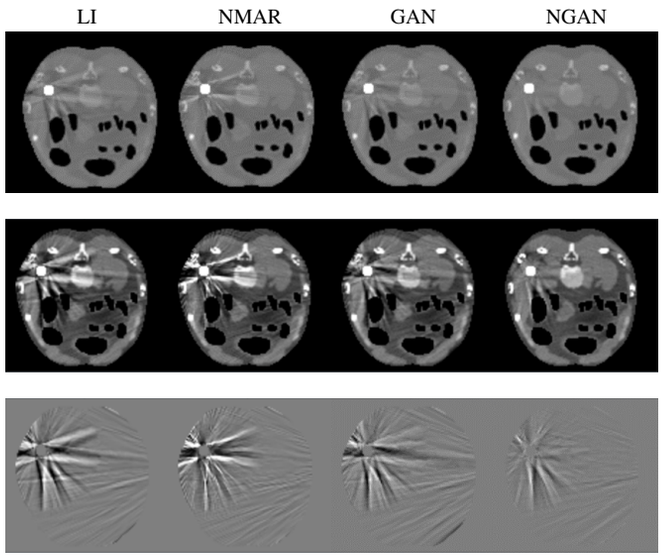


Fig. 3: Reconstructed images for the simulated XCAT phantom. Top row bone window(L: 300 HU, W: 2000 HU), middle row soft tissue window (L: 60 HU, W: 360 HU), last row difference images in relation to the ground truth.

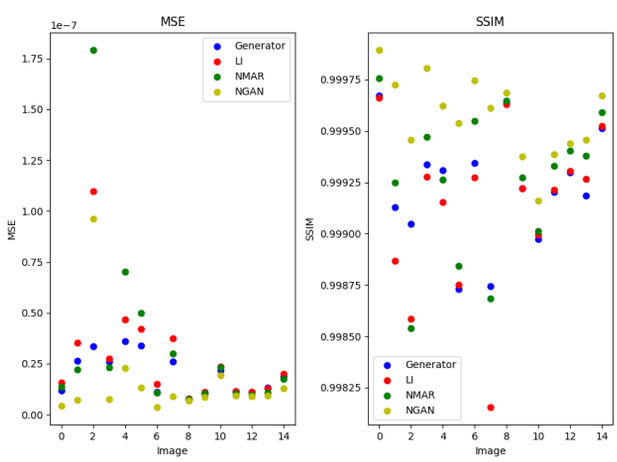


Fig. 4: Validation results for simulated test data.

any metal artifact correction (first column). Subsequently, LI, NMAR, the developed network, and the NGAN were applied to the data. Similar to the simulated phantom, the results are shown in the bone (top line) and soft tissue window (middle line). In addition, difference images (bottom line) were calculated in relation to the ground truth. All correction methods show a clear improvement compared to the uncorrected WFP. In particular the present black streak artifact in the middle of both objects could be corrected by all methods. Especially, LI, NMAR, but also to a lesser extent the network, show secondary streaking artifacts due to the correction method. These could be significantly reduced by the NGAN, where only a few remaining streaking artefacts can be detected in the difference image.

Figure 6 shows the mean MSE and SSIM values for the individual reconstructed images in the test data set. A clear

TABLE II: Final validation values on test data sets.

Data	Metric	LI	NMAR	GAN	NGAN
Simulation	MSE	$2.85 \pm 2.48 \cdot 10^{-8}$	$3.25 \pm 4.25 \cdot 10^{-8}$	$1.99 \pm 0.95 \cdot 10^{-8}$	$\mathbf{1.59 \pm 2.20 \cdot 10^{-8}}$
	SSIM	0.9991 ± 0.0003	0.9992 ± 0.0003	0.9992 ± 0.0002	$\mathbf{0.9995 \pm 0.0001}$
Measure	MSE	$3.55 \pm 1.15 \cdot 10^{-7}$	$3.63 \pm 3.24 \cdot 10^{-7}$	$2.70 \pm 0.85 \cdot 10^{-7}$	$\mathbf{2.27 \pm 2.27 \cdot 10^{-7}}$
	SSIM	0.989 ± 0.0013	0.990 ± 0.0013	0.991 ± 0.0009	$\mathbf{0.992 \pm 0.0008}$

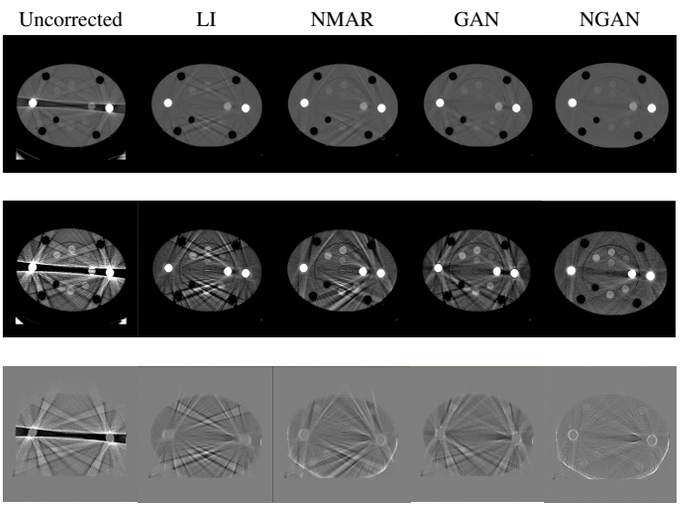


Fig. 5: Reconstructed images for the hardware phantom. Top row bone window(L: 300 HU, W: 2000 HU), middle row soft tissue window (L: 60 HU, W: 360 HU), last row difference images in relation to the metal free phantom.

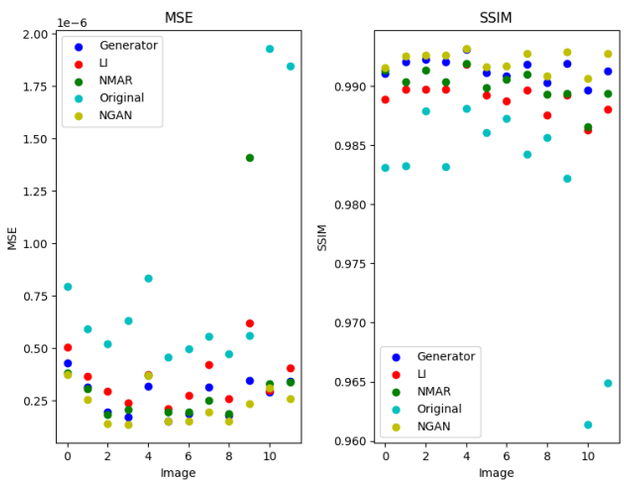


Fig. 6: Validation results for measured test data.

trend can be observed, both for the calculated MSE and SSIM value. While all methods are significantly better than the uncorrected artifact image (light blue), the LI (red) performs worst followed by NMAR (green) and the network (dark blue) for almost all test images. It can be observed that for the MSE NMAR is to some extent even better, while for the SSIM the network clearly achieves the better values. However, our final developed approach with the NGAN (light green) is clearly the best in terms of both metrics. Table II shows the corresponding average values over the complete test images. The NGAN, followed by the original GAN, shows improved values for both metrices compared to NMAR and LI.

VI. Discussion and Conclusion

By developing a new training approach, it is possible to integrate preliminary information into the learning process. Considering the data in the metal trace, reduces the risk of introducing artificially incorrect data in the area of the metal object. Moreover, this reduces the dependency on an exact segmentation of the metal object by using the original projection data as an input for the network. However, to use the information in the metal trace effectively, more detailed studies need to be conducted. Particularly, the exact weighting of the mask within the metal trace is crucial. It should not be too high in order to avoid secondary artifacts, but not too low, as otherwise there is a risk of neglecting remaining information. Further, the decrease of the values in the metal trace based on the forward projection of the metal objects can be investigated in more detail by using different functions that map a value range of $[0, 1]$.

In conclusion it could be shown that the results can be significantly improved compared to state of the art methods for MAR by integrating additional information in the training in particular by the developed NGAN.

References

[1] W. J. H. Veldkamp, R. M. S. Joemai, A. J. van der Molen, and J. Geleijns, "Development and validation of segmentation and interpolation techniques in sinograms for metal artifact suppression in CT," *Medical Physics*, vol. 37, no. 2, pp. 620 – 628, 2010. [Online]. Available: https://doi.org/10.1118/1.3276777

[2] A. Pajot, E. de Bézenac, and P. Gallinari, "Unsupervised adversarial image inpainting," *CoRR*, vol. abs/1912.12164, 2019. [Online]. Available: http://arxiv.org/abs/1912.12164

[3] N. Blum, T. M. Buzug, and M. Stille, "Data-driven metal artifact correction in computed tomography using conditional generative adversarial networks," in *The 7th International Conference on Image Formation in X-Ray Computed Tomography*, 2022, pp. 214–217.

[4] M. Stille, M. Kleine, J. Haegele, J. Barkhausen, and T. Buzug, "Augmented likelihood image reconstruction," *IEEE transactions on medical imaging*, vol. 35, 07 2015.

[5] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," 2018.

[6] L. Hellwege, N. Blum, T. M. Buzug, and M. Stille, "Partial convolution network for metal artifact reduction in ct preprocessing: Preliminary results," in *Medical Imaging and Deep Learning 2021*, 07 2021.

[7] E. Raithel, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (nmar) in computed tomography," *Medical Physics*, vol. 37, pp. 5482–5493, 10 2010.

[8] W. Segars, M. Mahesh, T. Beck, E. Frey, and B. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Medical Physics*, vol. 35, pp. 3800–3808, 2008. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.2955743

Beam hardening correction in diagnostic CT using a convex combination of water and bone

F. Noo, Z. Guo, K. Stierstorfer, G. S. K. Fung

***Abstract*—Beam hardening is an important source of error in diagnostic CT that can impact detection of various diseases and typically prevents reliable use of X-ray attenuation values as biomarkers. The classical approach for beam hardening correction is to map measurements to lengths of water, possibly with an iterative refinement when the region-of-interest is close to bones or iodine-filled blood vessels. In this work, we investigate a rather different concept for beam hardening correction in non-contrast single energy CT. In this concept, water and bone are immediately involved together in a convex combination, rather than sequentially. A computer simulation emulating non-contrast brain CT imaging is carried out to demonstrate the potential of this new concept. The outcome is promising as the results with the new concept are nearly free of artifacts and much more accurate than those obtained with correction based on water lengths only.**

I. Introduction

Beam hardening is an important source of error in diagnostic CT. It is known to affect the detection of brain hemorrhage and other signs of disease in the proximity of complex arrangements of bones or blood vessels that carry large amounts of iodine, like the aorta. It also prevents reliable use of X-ray attenuation values as biomarkers, as discussed in [1]. While beam hardening is most often talked about in the context of single energy CT, beam hardening also affects dual energy CT and photon counting CT whenever the energy decomposition is performed in the image domain, which is often the case.

The classical approach for beam hardening correction is to map the measurements to lengths of water [2], [3]. This approach works well for regions of interest that lie away from complex bone sites or large vessels filled with iodine. For other regions, an iterative refinement based on segmentation of bone or iodine is often applied [2]. Both water mapping and iterative refinement can be performed using polynomials in the data domain as originally proposed, or using polynomials in the image domain as more recently presented [4], [5].

In this work, we investigate a rather different concept for beam hardening correction in non-contrast single energy CT. In this concept, water and bone are immediately involved together in a convex combination, rather than sequentially. Hereafter, we present this concept in more details, and present as well an early demonstration of potential using computer simulated data.

F. Noo and Z. Guo are with the Department of Radiology and Imaging Sciences, University of Utah, USA. K. Stierstorfer and G. S. K. Fung are with Siemens Healthineers. This work was partly supported by NIH grants R01 HL153964-01 and R21 EB034337-01, as well as by Siemens Medical Solutions, USA. The concepts and information presented in this paper are based on research and are not commercially available. Its contents are solely the responsibility of the authors and do not represent the official views of the NIH.

II. Theory

Let $\mu(\vec{x}, E)$ represent the linear attenuation coefficient at energy E for the tissue located at position $\vec{x}$. Then, any CT measurement can be written as

$$g = -\log\left(\int_0^{E_{max}} W(E)\, e^{-\int_{\mathcal{L}} \mu(\vec{x},E)\, d\vec{x}}\, dE\right), \quad (1)$$

where $\mathcal{L}$ is the line along which the measurement is taken, E_{max} is the X-ray tube voltage; and $W(E)$ is the energy response, which is obtained from the source spectrum, $S(E)$, and the energy absorption of the detector, $D(E)$, as

$$W(E) = \frac{S(E)\, D(E)}{\int_0^{E_{max}} S(E)\, D(E)\, dE}. \quad (2)$$

The classical beam hardening correction based on water lengths amounts to making the following approximation

$$\mu(\vec{x}, E) \approx \rho(\vec{x})\, \mu_{H_2O}(E), \quad (3)$$

where $\mu_{H_2O}(E)$ is the linear attenuation coefficient of X-ray for water. That is, all tissues are assumed to behave similarly to water in terms of energy dependence, with the differences observed in Hounsfield units being due to differences in mass density. This approximation leads to

$$g = -\log\left(\int_0^{E_{max}} W(E)\, e^{-\mu_{H_2O}(E)\, L}\, dE\right), \quad (4)$$

where

$$L = \int_{\mathcal{L}} \rho(\vec{x})\, d\vec{x}. \quad (5)$$

Assuming that $W(E)$ is known, L can be computed from (4). Once L is known for all measurement lines, image reconstruction of $\rho(\vec{x})$ can be performed and the final reconstructed image can be expressed as

$$\hat{\mu}(\vec{x}) = \rho(\vec{x})\, \mu_{H_2O}(E_r), \quad (6)$$

where E_r is some reference energy, typically chosen as 70 keV. Typically, it is preferred to create $\hat{g} = L \cdot \mu_{H_2O}(E_r)$. Then, image reconstruction from $\hat{g}$ immediately yields $\hat{\mu}(\vec{x})$. When $W(E)$ is not known, (4) can be estimated using projection data through known lengths of water.

Now, we explain our new concept. Let Ω be the region where $\mu(\vec{x}, E) \neq 0$, and let

$$\chi(\vec{x}) = \begin{cases} 1 & \text{if } \vec{x} \in \Omega \\ 0 & \text{otherwise} \end{cases}. \quad (7)$$

We suggest the following approximation

$$\begin{aligned} \mu(\vec{x}, E) &\approx \chi(\vec{x})\Big(\omega(\vec{x})\, \mu_1(E) + \big(1 - \omega(\vec{x})\big)\, \mu_2(E)\Big) \\ &= \chi(\vec{x})\, \mu_2(E) + \chi(\vec{x})\, \omega(\vec{x})\, \Delta\mu(E) \end{aligned}, \quad (8)$$

where $\Delta\mu(E) = \mu_1(E) - \mu_2(E)$, and where $\mu_1(E)$ and $\mu_2(E)$ are the linear attenuation coefficients of X-ray for cortical bone and water, respectively. The rationale for this approximation is as follows. First, (8) may be a better model for soft tissues than (3) to account for the fact that the mass attenuation coefficients of soft tissues are not exactly the same; by including a bone component using a small (positive or negative) value for ω, a more correct expression may be obtained. Second, cortical bone can be exactly represented using $\omega = 1$. Third, trabecular bone, which comes as various combinations of bone marrow and compact bone, can be accurately modeled. One downside of our model is that air (vacuum) cannot be well-represented at all energies, unlike with (3); hence the need to include $\chi(\vec{x})$ in our model.

Approximation (8) leads to

$$g \approx -\log\left(\int_0^{E_{max}} \hat{W}(E)\, e^{-\Delta\mu(E)\int_{\mathcal{L}} \chi(\vec{x})\,\omega(\vec{x})\,d\vec{x}}\, dE\right), \tag{9}$$

where

$$\hat{W}(E) = W(E)\, e^{-\mu_2(E)\int_{\mathcal{L}} \chi(\vec{x})\,d\vec{x}}. \tag{10}$$

Assuming that $W(E)$ is known,

$$\int_{\mathcal{L}} \chi(\vec{x})\,\omega(\vec{x})\,d\vec{x} \tag{11}$$

can be computed from (9) for each measurement line. Once these line integrals are known, image reconstruction of $\chi(\vec{x})\,\omega(\vec{x})$ can be performed and a final reconstructed image can be expressed as

$$\hat{\mu}(\vec{x}) = \chi(\vec{x})\Big(\omega(\vec{x})\mu_1(E_r) + \big(1-\omega(\vec{x})\big)\mu_2(E_r)\Big). \tag{12}$$

III. Experimental set-up

We investigated the potential of (8) for beam hardening correction using simulated fan-beam data. The simulation parameters, given in TABLE I, were representative of state-of-the-art CT scanners, except for the detector pixel size, which is twice smaller to account for the typical use of a flying focal spot that was not modeled in our simulations. Three tube voltages were considered, namely 80 kV, 120 kV and 140 kV. Quantities $S(E)$, $D(E)$ and $W(E)$ are shown in Fig. 1.

The simulation emulated non-contrast CT imaging of the head. A modified version of the FORBILD head phantom was used for this purpose, to model the two temporal bones as well as trabecular bones of various strengths surrounded by cortical bone. We chose white matter for the background tissue; and we included disks simulating fresh blood, and strings representing gray matter. Fig. 2 shows the phantom using the compressed grayscale window of 80 HU used by the radiologist, and using a wider window to show the cortical and trabecular bones. All linear attenuation coefficients used for the simulation were taken from [6].

To apply (8), we used the cortical bone composition from the ICRU-44 publication [7], which strongly differs from that later suggested in [6]. As an important consequence, note that none of the materials found in the phantom can be exactly represented as a convex combination of $\mu_1(E)$ and $\mu_2(E)$.

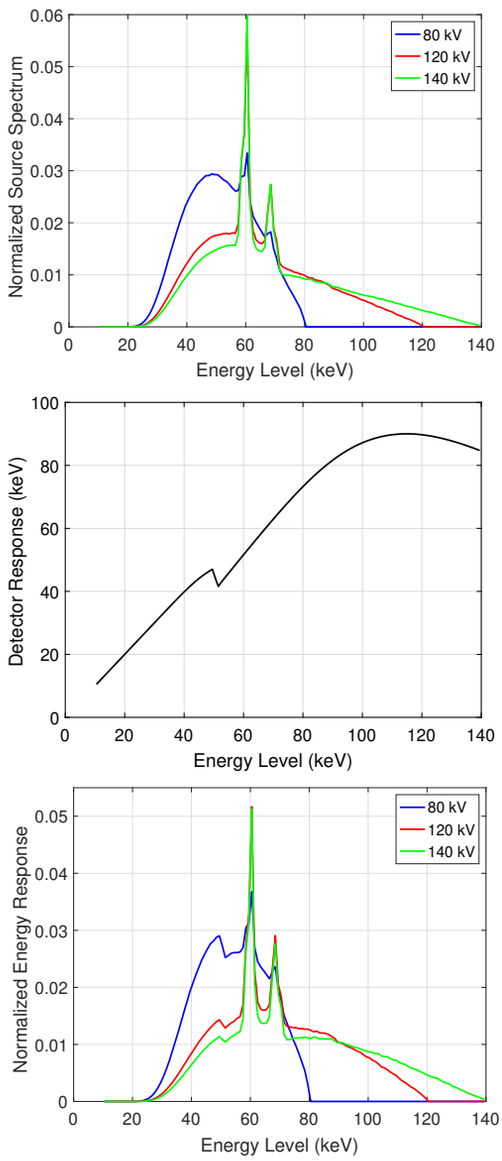


Fig. 1: Top: source spectrum, $S(E)$, viewed as a probability distribution function. Middle: energy absorbed by the detector, $D(E)$. Bottom: resulting energy response, $W(E)$, for tube voltages of 80 kV, 120 kV and 140 kV.

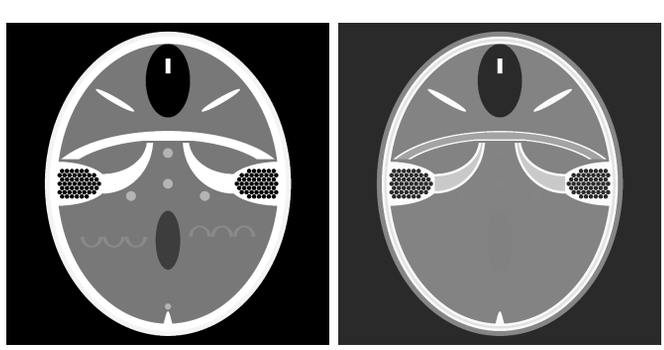


Fig. 2: Modified FORBILD Head phantom shown using two different grayscale windows: (left) $[0, 80]$ HU showing the low-contrast features, and (right) $[-1500, 1500]$ HU showing variations in trabecular bones.

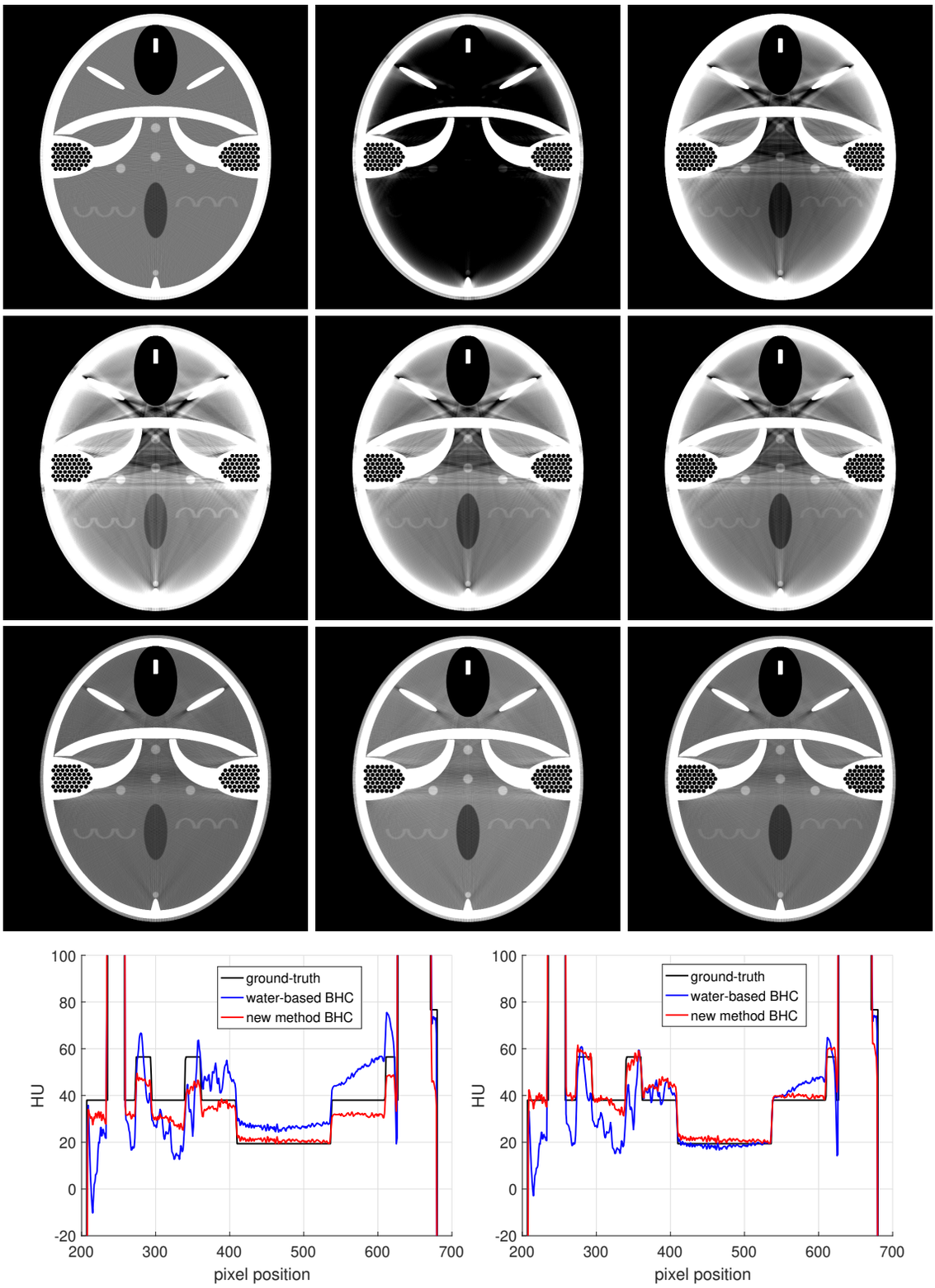


Fig. 3: Top row: (left) reconstruction from mono-energetic data at 70 keV, (middle) reconstruction from poly-energetic data with no correction, (right) same as middle, but displayed with another grayscale window. Second row: from left to right, reconstructions from poly-energetic data with water-based beam hardening correction at tube voltages of 80 kV, 120 kV and 140 kV. Third row: same as second row but using our new concept. All images are displayed over $[0, 80]$ HU, except the top right image where $[-40, 40]$ HU is used. Fourth row: central vertical profile through the 80 kV (left) and 120 kV (right) reconstructions with beam hardening correction; the profile starts at the bottom of the air cavity and goes through two fresh-blood disks and then through the CSF ellipse.

source-isocenter distance (R)	59.5 cm
source-detector distance (D)	108.56 cm
number of views	1152
number of channels	1472
detector pixel size	0.064 cm
# of sub-lines per detector pixel	5
FOV radius (r)	13 cm
image pixel size	0.01875 cm (followed by 2x2 binning)
ramp filter apodization	Hamming window

TABLE I: Simulation parameters

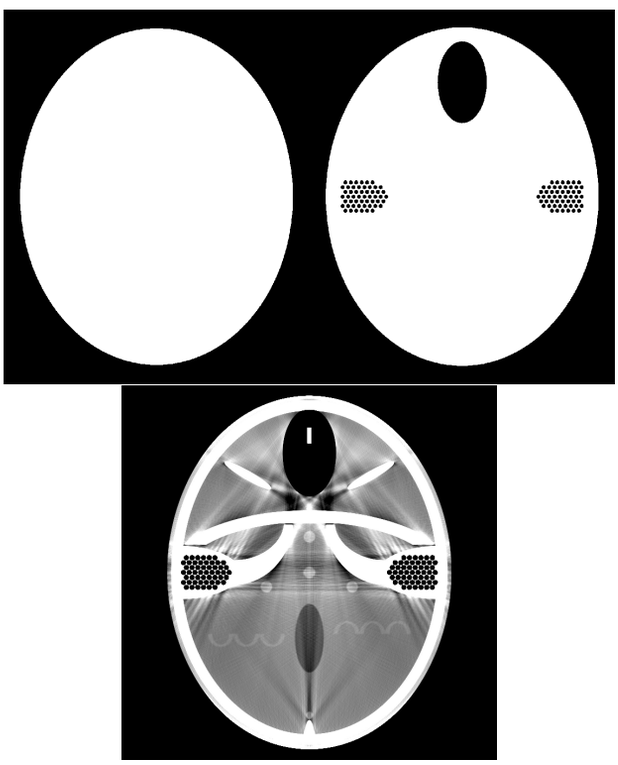


Fig. 4: Top left: coarse version for $\chi(\vec{x})$ defined as a slightly blurred version of the outer ellipse forming the phantom. Top right: slightly blurred version of the true expression for $\chi(\vec{x})$, which was used for the results in Fig. 3. Bottom: 120 kV reconstruction obtained when using the coarse version shown above (top left) for $\chi(\vec{x})$ in (8); grayscale window: $[0, 80]$ HU.

IV. Results

Fig. 3 shows reconstructions obtained with our new concept, (8), versus reconstructions with water correction only, namely (3). This figure also includes, for further comparison, a mono-energetic reconstruction at E_r, and a reconstruction without beam hardening correction. Fig. 3 also compares profiles at 80 kV and 120 kV between using (8) and (3). The results of Fig. 3 assume knowledge of a slightly blurred version of $\chi(\vec{x})$. To test the importance of knowing $\chi(\vec{x})$, we repeated the experiment using a coarse estimate of $\chi(\vec{x})$ that was equal to 1 everywhere inside the phantom and 0 otherwise. The results for this test are shown in Fig. 4.

V. Discussion and conclusion

We have introduced a new concept for beam hardening correction in diagnostic CT that jointly involved bone and water in a convex combination. We tested the potential of this concept using computer-simulated fan-beam data emulating non-contrast brain CT imaging. We find the results shown in Fig. 3 very promising, as the reconstructions with the new concept are nearly free of artifacts and much more accurate than those obtained using beam hardening correction based on water lengths. The observed residual errors may be attributed to the mismatch between the model (i.e., the chosen basis functions, water and cortical bone from ICRU-44, and their convex combination) and the chemical formulations used for tissues and bones in the phantom, which are from [6].

We observed early on that the new concept is unable to accurately represent air regions, hence a mask, $\chi(\vec{x})$, corresponding to the region of non-zero attenuation must be known. The results in Fig. 4 clearly showed that image quality depends on the accuracy with which $\chi(\vec{x})$ is known, although we note that the result obtained with the coarse mask is still more accurate than that obtained with water correction. We believe that it should be relatively straightforward to identify $\chi(\vec{x})$ from an initial reconstruction, using for example a simple threshold of -500 HU.

It is not clear if (8) could be obtained from calibration measurements or absolutely requires knowledge of $W(E)$. However, we showed in [1] that robust analytical models are now available to estimate $W(E)$, with the advantage that scatter can be corrected separately, either using a simple model as shown in [1] or using deep-learning based advances [8].

We are actively working on testing our new concept on real data and evaluating its value for various clinical applications, including non-contrast brain CT and bone marrow edema imaging for, e.g., early detection of arthritis. We are also actively working on comparing this new concept with corrections based on iterative refinements [2], as used by many vendors.

References

[1] Haase, V., Hahn, K., Schöndube, H., Stierstorfer, K., Maier, A., and Noo, F., "Single-material beam hardening correction via an analytical energy response model for diagnostic CT," *Medical Physics* **49**(8), 5014–5037 (2022).

[2] Joseph, P. M. and Spital, R. D., "A method for correcting bone induced artifacts in computed tomography scanners.," *Journal of computer assisted tomography* **2**(1), 100–108 (1978).

[3] Herman, G. T., "Correction for beam hardening in computed tomography," *Physics in Medicine & Biology* **24**(1), 81 (1979).

[4] Kachelrieß, M., Sourbelle, K., and Kalender, W. A., "Empirical cupping correction: A first-order raw data precorrection for cone-beam computed tomography," *Medical physics* **33**(5), 1269–1274 (2006).

[5] So, A., Hsieh, J., Li, J.-Y., and Lee, T.-Y., "Beam hardening correction in CT myocardial perfusion measurement," *Physics in Medicine & Biology* **54**(10), 3031 (2009).

[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] "ICRU Report 44." https://www.icru.org/report/tissue-substitutes-in-radiation-dosimetry-and-measurement-report-44/.

[8] Maier, J., Sawall, S., Knaup, M., and Kachelrieß, M., "Deep scatter estimation (DSE): Accurate real-time scatter estimation for X-ray CT using a deep convolutional neural network," *Journal of Nondestructive Evaluation* **37**, 1–9 (2018).

Quantitative metal artifact reduction algorithm for spectral CT thermometry

Leening P. Liu[1,2], Kevin M. Brown[3], Amy E. Perkins[3], Michael C. Soulen[1], Peter B. Noël[1]

Abstract— **Spectral CT thermometry can non-invasively monitor internal temperatures to reduce local tumor recurrences caused by insufficient heating/treatment of the tumor and its surrounding safety margin. For its clinical translation, the applied metal artifact reduction algorithm requires quantitative accuracy to ensure the accuracy of generated temperature maps. The newly developed Spectrally Obtained Needle Artifact Reduction (SONAR) algorithm leverages the known shape of the applicator and spectral CT's material decomposition capabilities to isolate the applicator in projections. Projections with long path lengths through metal were then corrected by replacing them with modeled projections of an angled cylinder. To evaluate the accuracy of SONAR, a liver-mimicking phantom embedded with an ablation applicator and thermometers was scanned with a dual-layer spectral CT at phantom temperatures of 35 and 80 °C. Using spectral CT thermometry, temperature maps at 80 °C were generated for image slices with and without the applicator. SONAR significantly decreased streaks along the axis of the applicator. It also eliminated underestimated temperatures immediately adjacent to the applicator and overestimated temperatures in the periphery (2-3 cm from applicator). While application of SONAR resulted in minimal absolute difference in the temperature map without the applicator, averaging 1.1 ± 0.8 °C, temperatures decreased 7.0 ± 4.0 and 10.1 ± 2.3 °C at distances of 2-3 and 0.5-1 cm from the applicator, respectively, to better match the expected temperature. SONAR ultimately minimized metal artifacts and lessened overestimation of temperature in spectral CT thermometry maps. These quantitatively accurate maps will facilitate the *in vivo* evaluation of spectral CT thermometry for non-invasive temperature monitoring of thermal ablations in order to reduce local tumor recurrences.**

Index Terms— **Spectral CT, Thermometry, Metal Artifact Reduction**

I. Introduction

THERMAL ablation has become increasingly utilized in the treatment of focal malignant tumors with effectiveness comparable to surgical resection [1-2]. Its effectiveness relies on focal heating of both the tumor and a surrounding safety margin to above the lethal threshold in order to kill tumor cells [3]. Despite improvements in ablation techniques, such as microwave ablation, which have lowered local recurrence rates, these rates are still notably high. Insufficient heating of the tumor and its margin to the lethal 60°C threshold is a key factor in local tumor recurrences [3]. Real-time non-invasive temperature monitoring, which maps spatial temperature distribution, has the potential to serve as an essential feedback tool for interventional radiologists during procedures. By doing so, the procedure can be guided to ensure total coverage of both the tumor and safety margin.

Spectral CT thermometry, a non-invasive method for temperature monitoring, leverages the material characterization and differentiation capabilities inherent in spectral CT technology. From quantitative spectral results, corresponding physical density maps can be generated, which can be directly correlated to change in temperature with the physical principle of thermal volumetric expansion [4]. Compared to existing CT thermometry methodologies, spectral CT thermometry enables the separation of the effect of temperature from the effect of composition changes, thereby improving its accuracy. Prior research has established the feasibility of spectral CT thermometry in both *ex vivo* muscle tissues [5] and gel phantoms designed to emulate liver tissue, specifically in terms of attenuation, physical density, and heat response characteristics [6]. Furthermore, this technique has demonstrated consistent model parameters across varying radiation doses and repeated phantom tests, reinforcing the reliability of spectral CT thermometry for temperature monitoring [6].

The clinical translation of spectral CT thermometry for use in thermal ablation necessitates a robust and quantitatively accurate metal artifact reduction algorithm (MAR) [4]. Existing metal artifact reduction algorithms are typically generalized to deal with a variety of objects and focused on improving image quality to make sure that images are of diagnostic quality. They range from methodologies with inpainting [7-8] to iterative methods [9-11]. However, for spectral CT thermometry, not only is the metal object known, but also quantitative accuracy is crucial to the accuracy of generated temperature maps. We utilize the defined shape of the applicator and the material decomposition capabilities of spectral CT, to introduce a metal artifact reduction algorithm that can ensure the generation of quantitatively accurate temperature maps.

II. Methods

A. Spectrally Obtained Needle Artifact Reduction (SONAR)

The MAR algorithm, SONAR, decomposes spectral CT data into water and stainless steel projections and corrects them by modelling the projections through the applicator. First, low- and high-energy projections were decomposed into a

This study was funded by the National Institutes of Health (R01EB030494) and Philips Healthcare.

L.P. Liu is with the Department of Bioengineering and Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. M.C. Soulen and P.B. Noël are with the Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. A.E. Perkins and K.M. Brown are with Philips Healthcare, OH, USA. (e-mail: leening@seas.upenn.edu).

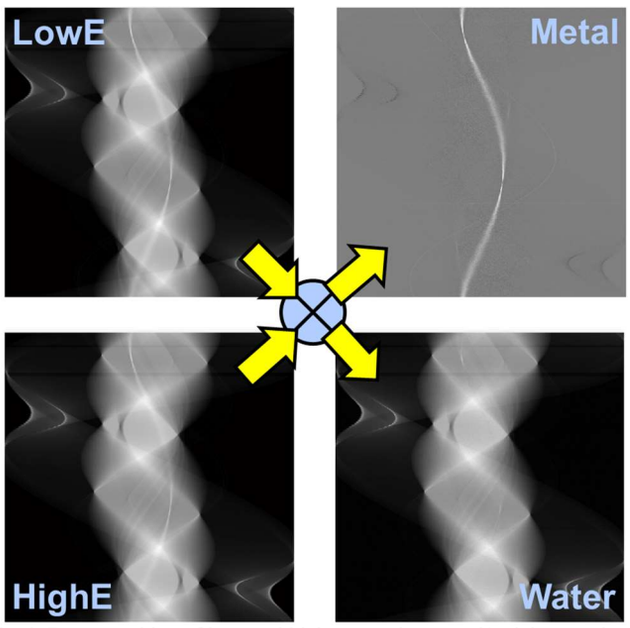


Fig. 1. Decomposition of low- and high-energy into metal/stainless steel and water basis projections. By utilizing metal and water bases, the metal trace associated with the applicator was isolated to enable better correction for projections with long path lengths through metal.

metal/stainless steel and water basis pair [12] to separate the metal trace from other materials (**Fig. 1**). To determine the location and angle of the needle, the peak of the needle trace was identified for each detector row within each projection by taking the midpoint between two full width half max points in the decomposed metal sinogram. The angle γ_v (**Fig. 2**) between the long axis of the needle and the x-axis of the detector for projection v was then calculated from the average slope of the peak movement across the different detector rows as

$$\gamma_v = abs\left(atan\left(slope_v\right)\right) \quad (1)$$

While the decomposition provided correct path lengths through the metal applicator when path lengths through metal are small

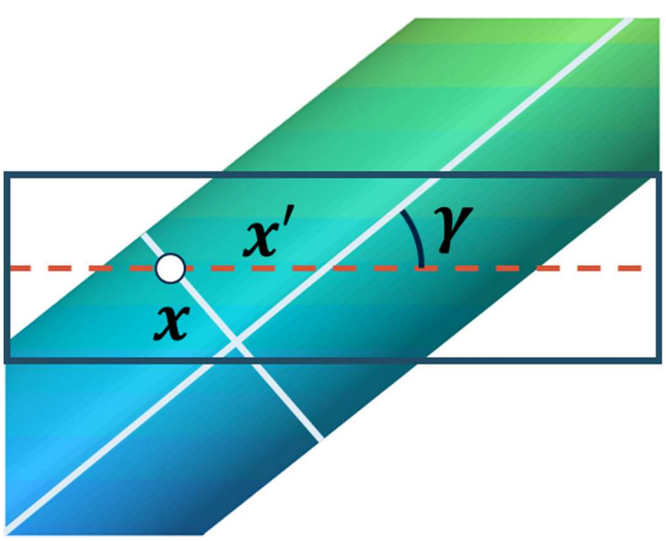


Fig. 2. Model of projection through a thin cylinder at a single angle. For a single detector row (blue bounding box), the path through the thin cylinder at a given point (white circle) is defined by the angle γ between the long axis of the cylinder (white line) and the x axis of the detector (orange dotted line) as well as its relative distances from the long axis of the cylinder.

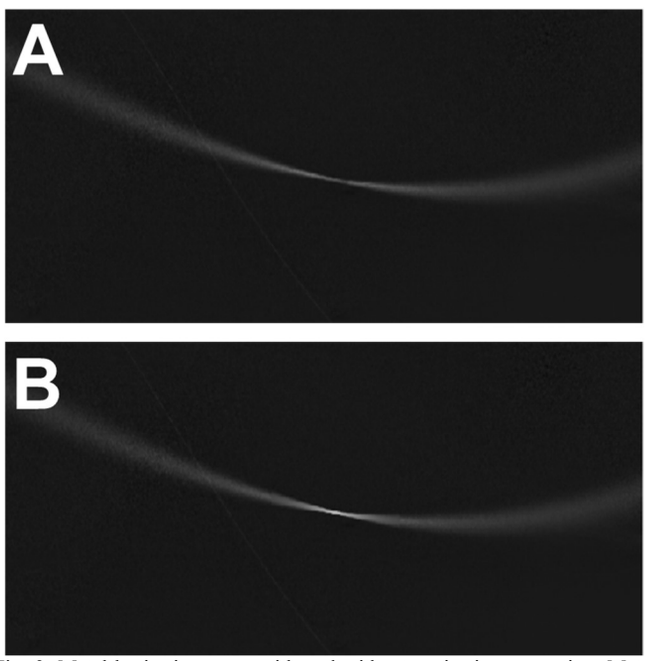


Fig. 3. Metal basis sinograms with and without projection correction. Metal basis sinograms obtained from material decomposition (A) demonstrated errors for projections with long paths through metal. These projections were replaced with modelled projections of a thin cylinder (B) to reduce metal artifacts.

(beam hardening is correctly accounted for), projections with long paths through metal suffered from errors where the measurement becomes dominated by stray photons (i.e. off-focal and scattered radiation, **Fig. 3A**). Projections that suffered from errors were determined by selecting 60 projections on either side of the projection where γ was 90°.

To correct these projections, projections through the known shape of the applicator were modeled as projections through a thin cylinder. Parallel projections through a circle are given by $p(x) = 2\sqrt{R^2 - x^2}$ where R is the circle radius and x is the perpendicular distance of the ray from the circle center. To adapt this to projections through a thin cylinder, two angles of the cylinder were considered. The first angle is γ, the angle between the needle and the x-axis of the CT detector in each projection (Figure 2). By defining $x = x'sin(\gamma)$,, such that x' is the distance of the ray from the circle center along the x-axis of the detector, and subsampling the detector row into s sub-rows, projections through the cylinder at each detector row are given by

$$p_{avg,v}(x) = \frac{1}{s}\sum_s 2\sqrt{R^2 - x_s^2} \quad (2)$$

The second important angle is the angle between long axis of the needle cylinder and the line connecting the x-ray source to the detector. Projections through the cylinder are lengthened when this angle is small and conform closer to Eq (2) when this angle approaches 90° (needle perpendicular to the ray). Fortunately this angle is related to the angle γ from above, and thus projections from Eq (2) were scaled according to

$$p_{final,v}(x) = p_{avg,v}(x)\frac{sin(\gamma_v)}{min(sin(\gamma))} \quad (3)$$

The water sinogram was also corrected for projections with long paths through metal. Between the end points of the modelled metal projection, the water sinogram was linearly interpolated followed by a subtraction of the path lengths of the modeled metal, since metal and water cannot occupy the same physical space. Corrected metal and water sinograms were then reconstructed into metal and water material basis images, and denoised with a version of the algorithm from [13]. Mono-energetic images were then calculated from the denoised metal and water basis images by standard methods [14] using the known attenuation curves of stainless steel and water.

B. Phantom

A homogeneous gel phantom that mimics the attenuation and heating properties of liver was synthesized. With 2 mL of deionized water, 1 g of ammonium persulfate was dissolved and vortexed. Similarly, 5.8 g of calcium chloride dihydrate was dissolved in 10 mL of deionized. Then, 290 mL of deionized water and 200 mL of 40% acrylamide/bisacrylamide were mixed by magnetic stirring. The calcium chloride solution was then added and stirred for at least 60 seconds. Subsequently, the ammonium persulfate and 1 mL of N,N,N',N'-tertramethylethylenediamine were added in quick succession to the water, acrylamide, and calcium chloride solution. During the polymerization of the gel, items were embedded in the gel to prevent formation of bubbles and cracks. These included a fiber Bragg grating (FBG) temperature sensor (FiSpec FBG X100, FiSens GmbH), optical fiber thermometer (Fiber Optic Temperature Sensor, Omega Engineering), and a metallic microwave ablation applicator (Neuwave PR Probe, Ethicon), which were embedded 2, 3, and 4 cm from the center of the phantom, respectively. The phantom was then placed at 4 °C for an hour prior to moving it to room temperature until use.

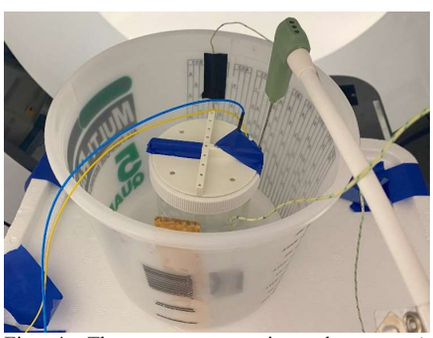


Fig. 4. Thermometry experimental setup. A homogeneous liver-mimicking phantom was embedded with temperature probes and an ablation applicator prior to heating in a water bath with hot water and an immersion heater.

C. Image acquisition and analysis

To expose the phantom to a range of temperatures, the phantom was placed in a water bath with an immersion heater (**Fig. 4**). With the help of the immersion heater, the phantom was heated to 80 °C. During the heating of the phantom, the phantom was scanned with a dual-layer spectral CT (Spectral CT 7500, Philips Healthcare) approximately every 30 seconds. Axial scans were performed at a tube voltage of 120 kVp with a rotation time of 0.75 s and a volumetric CT dose index ($CTDI_{vol}$) of 57 mGy. For scans at 35 and 80 °C, projection data as well as spectral results without SONAR, which utilized material decomposition with photoelectric and scatter bases, were extracted. SONAR was then applied to the projection data to generate spectral results. A commercially available MAR dedicated for use with large orthopedic implants only (O-MAR, Philips Healthcare) was also applied to the data to serve as a comparison [15]. All spectral results had a field of view of 300 mm and a slice thickness of 1 mm.

The output spectral results, particularly mono-energetic images at 70 keV and effective atomic number maps, were then utilized to generate physical density maps with the spectral physical density model [5]. Subsequently temperature maps were calculated based on spectral CT thermometry using the physical density map at 35 °C as the initial physical density. Temperature maps were analyzed with a 4x4 grid of regions of interest (ROI) with diameters of 5 mm (**Fig. 5**). These ROIs were placed 0.5, 1, 2, and 3 cm away from the end of the applicator identified on the SONAR temperature map and -1, 0, 1, 2 cm away from the long axis of the applicator. Mean and standard deviation were measured for each ROI on all temperature maps. Additionally, mean and standard deviations of ROIs along the long axis of the applicator at different distances were calculated and then compared between temperature maps with and without SONAR applied.

III. Results

The implementation of SONAR markedly decreased metal artifacts in temperature maps of the liver-mimicking phantom, maintaining accuracy around 80 °C (**Fig. 5**). While application of SONAR on images without the applicator showed minimal quantitative differences, its use on images with the applicator revealed significant improvements. Notably, with the applicator in the absence of SONAR, a streak artifact created a noticeable void in the temperature map along the applicator's axis. These same artifacts were also apparent on temperature maps with O-MAR in addition to significant underestimations in temperature and changes in temperature distribution throughout the phantom. On the other hand, SONAR effectively addressed this artifact, filling in the previous void with temperatures and enabling better temperature evaluation near the tip of the applicator. In addition, immediately adjacent to the applicator, metal artifacts caused aberrantly low temperatures without SONAR. By correcting projections where path lengths through metal was large, these erroneously low temperatures were eliminated, and the shape of the applicator was more conspicuous as previously present blooming was removed. However, some artifacts were still present, particularly adjacent to the tip of the applicator where temperatures are the hottest during thermal ablation.

With the addition of SONAR, temperature maps exhibited temperature more concordant with temperatures measured with temperature probes. For the temperature maps without the applicator, temperatures ranged 77.2 to 86.3 and 79.2 and 84.1 °C without and with SONAR, respectively, with an average noise of 1.5 ± 0.4 °C. As expected, with no metal present, absolute temperature differences were minimal between temperature maps with and without SONAR, averaging 1.1 ± 0.8 °C, respectively. On the other hand, temperatures in temperature maps containing the ablation

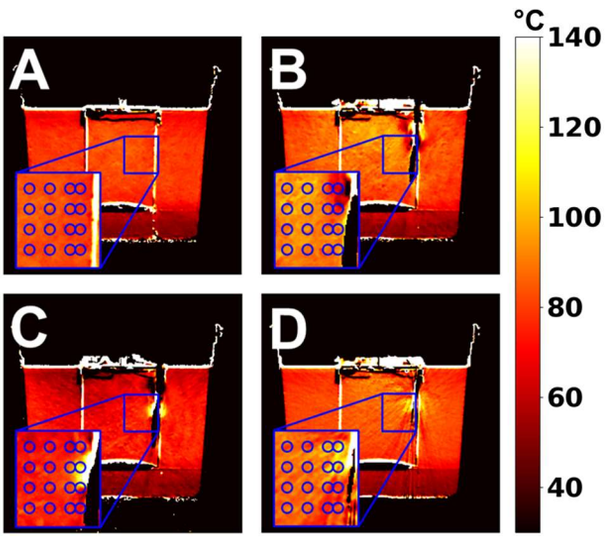


Fig. 5. Temperature maps of liver-mimicking phantom at approximately 80 °C. Temperature maps without the applicator in view demonstrated minimal quantitative differences without and with SONAR (A). In the presence of the applicator, temperature maps without SONAR (B) highlighted errors adjacent to the applicator as well as along the long axis of the applicator. While the temperature maps from images using O-MAR (C) exhibited similar artifacts, the application of SONAR significantly reduced these artifacts (D).

applicator ranged from 63.0 to 97.9 °C with an average noise of 3.8 ± 2.1 °C without SONAR, suggesting both underestimation and overestimation of temperatures as a result of metal artifacts. Temperature maps calculated using images with O-MAR similarly displayed under and overestimation but tended to underestimate temperatures with an average of 73.4 ± 4.42 °C when excluding areas of significant artifact compared to 92.5 ± 3.1 °C without O-MAR or SONAR. However, the addition of SONAR reduced the overestimation of temperatures at areas distant from the applicator by 7.5 and 6.5 °C at distances of 2 and 3 cm from the applicator, respectively. Immediately adjacent to the applicator tip, previously underestimated temperatures were overestimated, measuring 101.4 ± 3.9 °C 0.5 and 1 cm from the applicator. Even so, temperatures adjacent to the streaks decreased 10.1 ± 2.3 °C with SONAR, adjusting the temperature from 90.5 ± 2.8 to 80.3 ± 2.5 °C. Overall, excluding regions with significant artifacts, SONAR measured an average phantom temperature of 84.4 ± 3.7 °C, which better matched the approximate temperature measured by the embedded temperature probe.

IV. Conclusion

SONAR both qualitatively and quantitatively improved metal artifacts, reducing overestimation of temperature in spectral CT thermometry maps. Despite residual artifacts immediately adjacent to the applicator tip that corresponds to regions of high temperature, implementation of SONAR demonstrated reduced blooming and streaks as well as reduced overestimation by an average of 8 °C. While SONAR has shown significant potential for MAR in thermometry, more intensive modeling, i.e. finite focal spot and detector sizes, may further improve its quantitative performance.

This quantitatively accurate MAR enables the generation of accurate temperature maps in the presence in the ablation applicator. With these temperature maps, spectral CT thermometry can identify the ablation zone during ablation to ensure complete treatment of the tumor and its safety margin, thereby reducing local recurrences. Though spectral CT thermometry has demonstrated its utility in phantom studies, *in vivo* evaluations with SONAR are required for its clinical translation for the non-invasive temperature monitoring of thermal ablation.

Acknowledgements

We acknowledge support from National Institutes of Health (R01EB030494), SIR Foundation Allied Scientist Training Grant, and Philips Healthcare.

References

[1] J. Uhlig, C. M. Sellers, S. M. Stein, and H. S. Kim, "Radiofrequency ablation versus surgical resection of hepatocellular carcinoma: contemporary treatment trends and outcomes from the United States National Cancer Database," *Eur Radiol*, vol. 29, no. 5, pp. 2679–2689, Dec. 2018, doi: 10.1007/S00330-018-5902-4.

[2] H. Petrowsky and R. W. Busuttil, "Resection or ablation of small hepatocellular carcinoma: What is the better treatment?," *J Hepatol*, vol. 49, no. 4, pp. 502–504, 2008, doi: 10.1016/j.jhep.2008.07.018.

[3] S. N. Goldberg, G. S. Gazelle, and P. R. Mueller, "Thermal Ablation Therapy for Focal Malignancy," *AJR Am J Roentgenol*, vol. 174, no. 2, pp. 323–331, Feb. 2000, doi: 10.2214/ajr.174.2.1740323.

[4] F. Fani, E. Schena, P. Saccomandi, and S. Silvestri, "CT-based thermometry: An overview," *Int J Hyperthermia*, vol. 30, no. 4, pp. 219–227, 2014, doi: 10.3109/02656736.2014.922221.

[5] L. P. Liu *et al.*, "Non-invasive mass and temperature quantifications with spectral CT," *Sci Rep*, vol. 13, no. 1, pp. 1–11, Apr. 2023, doi: 10.1038/s41598-023-33264-2.

[6] L. P. Liu *et al.*, "Reproducible spectral CT thermometry with liver-mimicking phantoms for image-guided thermal ablation," *medRxiv*, p. 2023.10.04.23296423, Oct. 2023, doi: 10.1101/2023.10.04.23296423.

[7] X. Duan, L. Zhang, Y. Xiao, J. Cheng, Z. Chen, and Y. Xing, "Metal artifact reduction in CT images sinogram TV inpainting," *IEEE Nuclear Science Symposium Conference Record*, pp. 4175–4177, 2008, doi: 10.1109/NSSMIC.2008.4774201.

[8] C. Peng *et al.*, "An irregular metal trace inpainting network for x-ray CT metal artifact reduction," *Med Phys*, vol. 47, no. 9, pp. 4087–4100, Sep. 2020, doi: 10.1002/MP.14295.

[9] F. E. Boas and D. Fleischmann, "Evaluation of two iterative techniques for reducing metal artifacts in computed tomography," *Radiology*, vol. 259, no. 3, pp. 894–902, Jun. 2011, doi: 10.1148/radiol.11101782.

[10] T. Gilat Schmidt, R. F. Barber, and E. Y. Sidky, "Spectral CT metal artifact reduction with an optimization-based reconstruction algorithm," in *Medical Imaging 2017: Physics of Medical Imaging*, SPIE, Mar. 2017, p. 101321B. doi: 10.1117/12.2249079.

[11] R. A. Nasirudin *et al.*, "Reduction of Metal Artifact in Single Photon-Counting Computed Tomography by Spectral-Driven Iterative Reconstruction Technique," *PLoS One*, vol. 10, no. 5, p. e0124831, May 2015, doi: 10.1371/JOURNAL.PONE.0124831.

[12] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in X-ray computerised tomography," *Phys Med Biol*, vol. 21, no. 5, p. 733, Sep. 1976, doi: 10.1088/0031-9155/21/5/002.

[13] K. M. Brown, S. . Zabic, and G. Shechter, "Impact of spectral separation in dual-energy CT with anti-correlated statistical reconstruction. ," in *Proceedings of the 13th fully three-dimensional image reconstruction in radiology and nuclear medicine*, 2015, pp. 491–494.

[14] N. Shapira, Y. Yagil, N. Wainer, and A. Altman, "Spectral Imaging Technologies and Apps and Dual-Layer Detector Solution," *Spectral, Photon Counting Computed Tomography*, pp. 3–16, Jul. 2020, doi: 10.1201/9780429486111-1.

[15] J. Y. Huang *et al.*, "An evaluation of three commercially available metal artifact reduction methods for CT imaging," *Phys Med Biol*, vol. 60, no. 3, pp. 1047–1067, Feb. 2015, doi: 10.1088/0031-9155/60/3/1047.

Evaluation of Data Consistency Condition-based Beam Hardening and Scatter Corrections for Clinical C-arm CBCT

Shiras Abdurahman, Robert Frysch, Daniel Punzet, Michael Manhart, Hannes Lücking, Stefan Lang, Arnd Dörfler and Georg Rose

***Abstract*—Data Consistency Conditions (DCC) have been successfully employed for model-based artifact reduction in Cone-Beam Computed Tomography (CBCT). The model parameters for the optimal correction can be estimated by enforcing DCC on the uncorrected projection data or geometry. Previous works demonstrated that the polynomial coefficients for beam hardening correction could be optimized using pair-wise cone-beam DCC based on Grangeat's fundamental relation. This paper evaluates the efficacy of the DCC-based water and bone corrections for interventional C-arm CBCT systems employing short-scan and sine-spin trajectories. Our studies with the head scan datasets demonstrate the significant reduction of artifacts, and the results are comparable with that of calibration-based correction. In addition, the results show that polynomial correction can be employed to reduce the artifacts caused by scatter radiation. The data-driven algorithm facilitates artifact reduction for various CBCT system configurations (with and without scatter grids and different X-ray filters) and acquisition parameters (low- and high-energy imaging) since it does not require prior knowledge or re-calibrations.**

***Index Terms*—Beam hardening correction, scatter correction, data consistency conditions.**

I. Introduction

THE polychromatic nature of X-ray and attenuation spectra cause beam hardening, resulting in cupping and streak artifacts in the reconstructed images [1]. The underestimation of projection values can also be caused by scatter radiation, causing artifacts similar to beam hardening [2]. Beam hardening artifacts are corrected during the pre-processing of projections, where the underestimation errors are compensated by polynomial correction (first-order correction) [3]. For clinical imaging, the polynomial coefficients are estimated during the scanner's calibration using homogeneous water phantoms [4]. Hence, the correction is also known as water correction. The bone-induced beam hardening artifacts are corrected (higher-order correction) after prior reconstruction and the forward projection of segmented bone volume [5] [6]. A similar scheme can also employed for metal-induced beam hardening artifacts. There are plenty of methods published in the literature to prevent and compensate for the artifacts due to scatter [7]. The studies have shown that the polynomial correction is a viable method for reducing scatter artifacts [4] [8].

Pair-wise cone-beam Data Consistency Conditions (DCC) based on Grangeat's fundamental relation have been successfully employed for model-based artifact reduction in CBCT [9] [10]. Previous studies showed that the polynomial coefficients for water correction could be estimated by enforcing DCC on the uncorrected projection data [11] [12]. The major steps of water correction using DCC are summarized below:

1) Formulate the expression of artifact-corrected projections using the polynomial model.
2) Compute the inconsistency metric from the corrected projections parameterized by polynomial coefficients. In our studies, it is formulated as the sum of dissimilarity measures between the intermediate functions (first derivative of the plane integrals) computed from a set of cone-beam projection pairs.
3) Optimize the polynomial coefficients by minimizing the inconsistency metric.
4) Estimate the corrected projections using the optimized polynomial coefficients.

Similarly, the polynomial coefficients for bone or metal-beam hardening correction can also be optimized using DCC [11] [13].

This paper evaluates the efficacy of DCC-based beam hardening and scatter corrections for clinical C-arm CBCT systems employing short-scan and sine-spin trajectories. Unlike circular short-scan acquisition, the sine-spin trajectory employs an additional sinusoidal motion in the cranio-caudal direction during the rotation of the source-detector pair, reducing cone-beam artifacts in the reconstructed images [14]. Consequently, the improvement of image quality is more pronounced after beam hardening and scatter corrections. We present the results after DCC-based water correction and compare them with the scanner's scatter and beam hardening corrections. We also demonstrate that additional DCC-bone correction removes the spill-over and streak artifacts, enhancing the quality of low-contrast head scan images. We also discuss the advantages and challenges of integrating DCC-based corrections with the image processing pipeline.

Shiras Abdurahman, Robert Frysch, Daniel Punzet, and Georg Rose are with the Institute for Medical Engineering and Research Campus STIMULATE, Otto-von-Guericke-Universität, Universitätsplatz 2, 39106 Magdeburg, Germany.

Michael Manhart is with Siemens Healthcare GmbH, 91301, Forchheim, Germany.

Hannes Lücking, Stefan Lang, and Arnd Dörfler are with the Department for Neuroradiology, Universitätsklinikum Erlangen, Schwabachanlage 6, 91054 Erlangen, Germany.

Corresponding author: shiras.abdurahman@ovgu.de

II. Method

The optimization of univariate and quadratic water correction polynomial $\alpha_2 (g_p)^2 + \alpha_1 g_p + \alpha_0$ (g_p is the uncorrected polychromatic projection value) using DCC is described in detail in [12]. Under the constraints that the function is convex and the area under the polynomial is constant, the linear coefficient can be expressed as a function of the non-linear coefficient. The bound-constrained optimization of the non-linear coefficient can be expressed as:

$$\begin{aligned} &\min_{\alpha_2} \psi(\alpha_2) \\ &\text{s.t. } 0 \leq \alpha_2 \leq \frac{3}{2 g_p^{max}} \\ &\alpha_1 = 1 - \frac{2}{3}\alpha_2 \, g_p^{max}, \alpha_0 = 0 \end{aligned} \tag{1}$$

where g_p^{max} is the maximum value of projection dataset. ψ is the quantified inconsistency computed as the dissimilarity between the Grangeat intermediate functions computed from a set of water-corrected projection pairs.

Similarly, the quadratic polynomial for bone correction can be optimized using DCC as described in [11] and [12]. In this study, we employed bone corrections sequentially, assuming that the water correction had already been performed. The bone-corrected projection values can be expressed using water-corrected projections g_α, the forward projections of segmented bone volume f_b and the quadratic bone correction polynomial coefficients $\beta_2, \beta_1, \beta_0 = 0$ as given below:

$$g_\beta = g_\alpha + \beta_2 (f_b)^2 + (\beta_1 - 1) f_b \tag{2}$$

By utilizing the same convexity and area constraints, the non-linear coefficient of the bone-correction polynomial can be optimized as:

$$\begin{aligned} &\min_{\beta_2} \phi(\beta_2) \\ &\text{s.t. } 0 \leq \beta_2 \leq \frac{3}{2 f_b^{max}} \\ &\beta_1 = 1 - \frac{2}{3}\beta_2 \, f_b^{max} \end{aligned} \tag{3}$$

where f_b^{max} is the maximum value of the bone projections.

In our previous studies, we employed the Sum of Squared Differences (SSD) as the dissimilarity measure to quantify the inconsistency. The view-dependent axial truncation of the projection images caused by the sine-spin trajectory yields outliers in the intermediate function values, making the optimization of correction polynomials less robust. Hence, SSD was replaced by an outlier-resistant Geman-Mcclure error function for both sine-spin and conventional short-scan acquisitions [15] [16].

III. Experiments

Clinical projection data was acquired using Siemens Artis Icono interventional C-arm CBCT systems employing a sine-spin trajectory. The study included 21 patient datasets. The projection images were of size 472 (rows) x 648 (columns), and the number of projection views per acquisition was 546. The quadratic polynomials for water correction were estimated from the projection data without applying the scanner's scatter or water corrections.

DCC-based bone correction was performed after prior reconstruction, forward projection of the bone volume, and optimizing the quadratic bone correction polynomial. The prior volume was reconstructed from the projections after processing them with scanners's scatter and water corrections.

To study the efficacy of water correction for reducing the residual scatter artifacts, the projection data of the head phantom was acquired with and without scatter grids as per conventional short-scan trajectory. Before optimizing the polynomial, the projection data was processed with the scanner's scatter correction and beam hardening correction algorithms.

The images were reconstructed using the scanner's proprietary reconstruction algorithm. However, the images displayed in the results section do not represent the scanner's final images since many of the processing steps of the complete image processing pipeline were not applied.

IV. Results and Discussion

Fig. 1 shows that the cupping artifacts due to beam hardening and scatter were significantly reduced after DCC-based water corrections, and the results were similar to those of the scanner's scatter and beam hardening corrections.

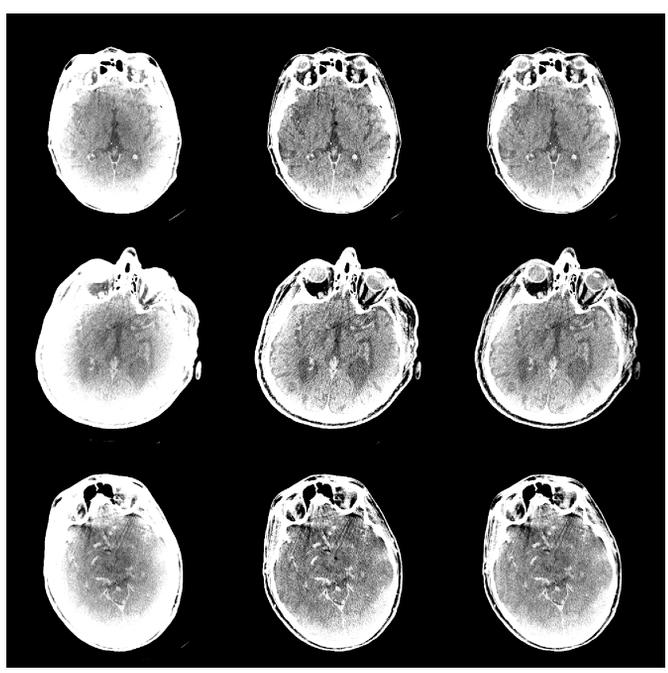


Fig. 1: Water correction for clinical data: left, reconstructed image without scatter or beam hardening correction; middle, reconstructed image after scanner's scatter and beam hardening corrections; right, reconstructed image after DCC-based water correction. Images courtesy of University Hospital, Erlangen, Germany.

Table. I displays the polynomials computed from a patient's projections without any corrections, after the scanner's scatter correction and after the scanner's scatter + beam hardening corrections. The nonlinearity of the optimized polynomial

(the value of the polynomial's non-linear coefficient α_2) is proportional to the degree of projection errors caused by beam hardening and scattering. The results validate our studies with simulated data described in [12], showing that polynomial correction can reduce artifacts due to beam hardening and scatter by adjusting the non-linearity of the polynomial. The composition of two polynomials (each for beam hardening and scatter correction) is another polynomial, and its approximation is a quadratic polynomial of high non-linearity. Since the projection error which can be minimized by the first-order correction was very low after the scanner's scatter and beam hardening corrections, the algorithm could not minimize the quantified inconsistency, and the optimized polynomial was the initial parameter (identity function).

Prior scanner's corrections	Optimized polynomial
no correction	$0.0407\ x^2 + 0.869\ x + 0.0$
scatter correction	$0.0129\ x^2 + 0.955\ x + 0.0$
scatter and beam hardening corrections	$0.0\ x^2 + 1.0\ x + 0.0$

TABLE I: Quadratic polynomials optimized from the projection data with and without the scanner's scatter and/or beam hardening corrections.

Fig. 2 displays the results before and after DCC-based bone correction. The results highlight the reduction of spill-over artifacts between the skull and brain regions, which could not be removed by sole water correction. DCC-bone correction enhances the soft-tissue contrast of C-arm CBCT imaging and facilitates the detection of pathologies (e.g., hemorrhage) near the skull.

Fig. 3 shows the reduction of streak artifacts after bone correction. The artifacts in the posterior fossa region are caused and amplified by multiple factors, including beam hardening, scatter, under- and non-uniform sampling, noise, motion, and cone-beam artifacts. Though the correction of streaks was not prominent compared to the spill-over artifacts, the visible artifact reduction could be achieved by DCC-bone correction.

Similarly, the artifacts due to beam hardening by metal implants can be reduced. As per our studies, the reduction of artifacts caused by small metal implants is marginal. We need to rely on projection inpainting-based Metal Artifact Reduction (MAR) algorithms like NMAR for better artifact reduction [17].

Fig. 4 shows the results after DCC-based water correction for the images acquired with and without scatter grids. All the projections were processed with the scanners' scatter and beam hardening corrections. The parameters of the scatter correction algorithm were calibrated using the scatter grid. Therefore, the correction was not optimal for grid-less imaging, and the residual scatter artifacts are visible as cupping artifacts. The DCC-based algorithm reduced the residual artifacts without recalibrating the parameters of the scatter correction algorithm. Hence, the DCC-based algorithm can be used to reduce the residual artifacts if the scanner's scatter or beam hardening correction is inadequate due to the changes in acquisition parameters or system configuration. In addition, patient-specific

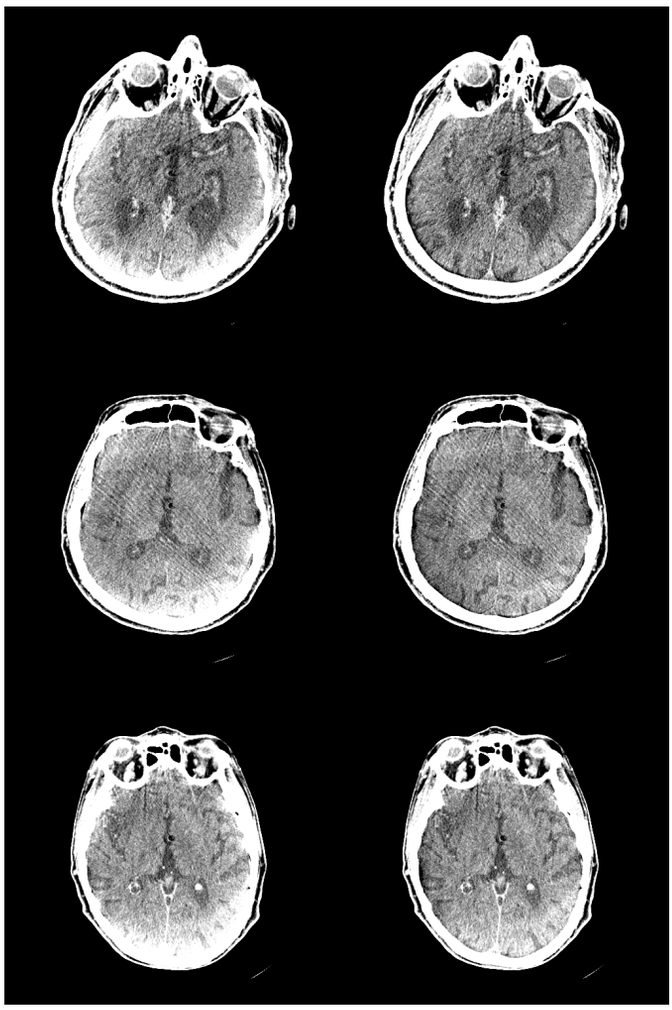


Fig. 2: Bone correction for clinical data: left, reconstructed image before DCC-based bone correction; right, reconstructed image after DCC-based bone correction. The projections were corrected with the scanner's scatter and water corrections. Images courtesy of University Hospital, Erlangen, Germany.

corrections are difficult to achieve using the calibration-based methods.

DCC-based artifact reduction algorithms work well only when the projection data is free from lateral truncations. As per our studies, the polynomial could only be optimized if the lateral truncation is not significant, like in the head scans. Most DCC-based algorithms are less robust for thoracic and abdominal scans if the detector area is small.

Besides reducing cupping artifacts, calibration-based water correction is necessary for maintaining Hounsfield Unit (HU) accuracy in clinical CT imaging. During the routine calibration of the scanner, the operators inspect whether the HU value of the uniform water phantom is 0 HU and the standard deviation of voxel values is very low. In DCC-based corrections, the water correction polynomials are optimized under the constraint that the mean of the projection pixel values are preserved before and after correction. In addition, the primary objective is to optimize the polynomial's non-linear coefficient

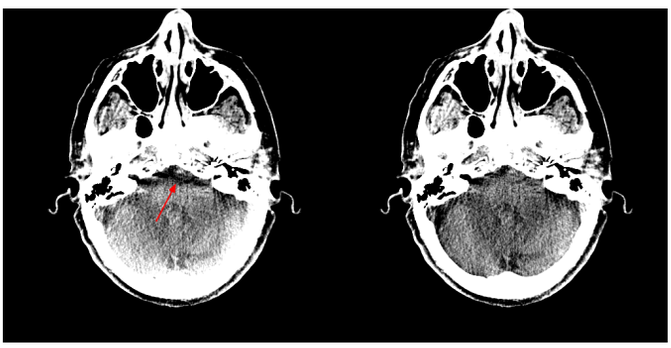


Fig. 3: Reduction of streak artifacts: left, reconstructed image before DCC-based bone correction; right, reconstructed image after DCC-based bone correction. The projections were corrected with the scanner's scatter and water corrections. Images courtesy of University Hospital, Erlangen, Germany.

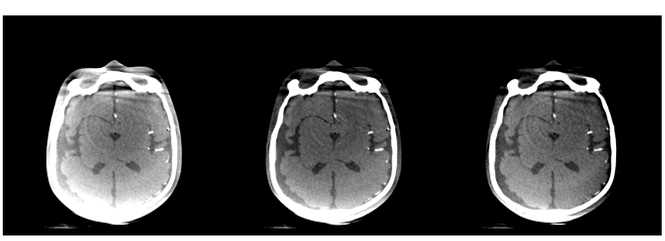


Fig. 4: Reduction of scatter artifacts: left, image reconstructed from the projections acquired without scatter grid; middle, image reconstructed from the projections acquired with scatter grid; right, image reconstructed from the projections acquired without scatter grid but after DCC-based polynomial correction. Images courtesy of Siemens Healthineers, Forchheim, Germany.

since they are mainly responsible for artifact reduction. Along with the linear attenuation coefficient of water (used for HU scaling), adapting the linear coefficient and constant term of the polynomial may be necessary to maintain HU accuracy. Hence, integrating the DCC-based water correction algorithm in imaging workflow requires data-driven HU scaling (e.g., based on histogram) after reconstruction. Our experiments with scatter reduction demonstrate that the additional DCC-based water correction can be used if the calibration-based corrections are insufficient for artifact reduction. HU accuracy was maintained after DCC-bone correction due to the prior calibration-based water correction and the constant area constraint during polynomial optimization.

V. Conclusion

In this paper, we assessed the beam hardening and scatter correction algorithms by employing data consistency conditions based on Grangeat's fundamental relation for C-arm CBCT systems utilizing short-scan and sine-spin trajectories. According to our experiments, cupping, spill-over, and streak artifacts were mitigated, and the results were comparable to those obtained with calibration-based corrections. DCC-based algorithms enable object- or patient-specific corrections without requiring prior knowledge of X-ray or material attenuation spectra. DCC-based algorithms can be additionally employed to reduce the artifacts if the calibration-based algorithms are inadequate due to the changes in system configuration or acquisition parameters, thereby eliminating re-calibrations.

Disclaimer: The concepts and information presented in this paper are based on research and are not commercially available.

References

[1] R. A. Brooks and G. Di Chiro, "Beam hardening in x-ray reconstructive tomography," *Physics in medicine & biology*, vol. 21, no. 3, p. 390, 1976.

[2] P. M. Joseph and R. D. Spital, "The effects of scatter in x-ray computed tomography," *Medical physics*, vol. 9, no. 4, pp. 464–472, 1982.

[3] G. T. Herman, "Correction for beam hardening in computed tomography," *Physics in Medicine & Biology*, vol. 24, no. 1, p. 81, 1979.

[4] 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.

[5] 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.

[6] P. M. Joseph and R. D. Spital, "A method for correcting bone induced artifacts in computed tomography scanners," *Journal of computer assisted tomography*, vol. 2, no. 1, pp. 100–108, 1978.

[7] 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," *Medical physics*, vol. 38, no. 7, pp. 4296–4311, 2011.

[8] M. Bertram, J. Wiegert, and G. Rose, "Potential of software-based scatter corrections in cone-beam volume ct," in *Medical Imaging 2005: Physics of Medical Imaging*, vol. 5745, pp. 259–270, International Society for Optics and Photonics, 2005.

[9] C. Debbeler, N. Maass, M. Elter, F. Dennerlein, and T. M. Buzug, "A new ct rawdata redundancy measure applied to automated misalignment correction," in *Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pp. 264–267, 2013.

[10] R. Frysch and G. Rose, "Rigid motion compensation in interventional c-arm ct using consistency measure on projection data," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 298–306, Springer, 2015.

[11] S. Abdurahman, R. Frysch, R. Bismark, S. Melnik, O. Beuing, and G. Rose, "Beam hardening correction using cone beam consistency conditions," *IEEE Transactions on Medical Imaging*, vol. 37, no. 10, pp. 2266–2277, 2018.

[12] S. Abdurahman, R. Frysch, T. Pfeiffer, O. Beuing, and G. Rose, "Comparative evaluation of cone-beam consistency conditions for monomaterial beam hardening correction in flat detector computed tomography," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 7, no. 4, pp. 354–371, 2022.

[13] S. Abdurahman, R. Frysch, and G. Rose, "Reduction of beam hardening induced metal artifacts using consistency conditions," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072, pp. 305–309, SPIE, 2019.

[14] Z. Guo, M. Manhart, P. Bernhardt, B. Schreiber, J. DiNitto, and F. Noo, "C-arm ct imaging of the head with the sine-spin trajectory: evaluation of cone-beam artifacts from computer-simulated data of voxelized patient models," in *Proceedings of the 17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, pp. 127–130, 2023.

[15] A. Preuhs, A. Maier, M. Manhart, M. Kowarschik, E. Hoppe, J. Fotouhi, N. Navab, and M. Unberath, "Symmetry prior for epipolar consistency," *International journal of computer assisted radiology and surgery*, vol. 14, no. 9, pp. 1541–1551, 2019.

[16] R. Frysch, T. Pfeiffer, and G. Rose, "A novel approach to 2d/3d registration of x-ray images using grangeat's relation," *Medical Image Analysis*, vol. 67, p. 101815, 2021.

[17] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (nmar) in computed tomography," *Medical physics*, vol. 37, no. 10, pp. 5482–5493, 2010.

ComptonNet: A Spare-detector-guided Deep Learning Framework for Scatter Estimation in Multi-source Dual-ring Stationary CT

Yingxian Xia, Zhiqiang chen, Li Zhang, Yuxiang Xing, and Hewei Gao*

***Abstract*—Accurate correction of both forward and cross scatter in multi-source stationary CT is challenging and essential for high image quality of CT reconstruction. Traditional methods based on models or scatter kernels have limited accuracy in multi-source stationary CT. Deep learning methods, particularly Unet, have been popular for scatter estimation due to their high accuracy and computational efficiency. However, estimating multi-cross scatter by common Unets requires high parameter complexity, which may lead to unstable prediction results and reduce the credibility of the architecture. Additionally, dual-ring stationary CT has spare detectors that only receive large-angle compton scatter signal. How to improve the accuracy of deep scatter estimation under the guidance of spare detector requires careful architecture design. In this manuscript, we propose a spare-detector-guided deep learning framework (ComptonNet) for scatter estimation in multi-source dual-ring stationary CT. ComptonNet first uses a 3-channel fusion encoder-decoder deep learning network to estimate cross scatter under the guidance of spare detector data, and then uses a frequency Unet to estimate forward scatter. Overall, it is an end-to-end neural network that can avoid manual feature extraction and make maximum use of spare detector data in dual-ring stationary CT. We introduce the frequency attention module in this framework, which not only helps achieve low parameter complexity of networks but also improves the low-frequency smoothness of the deep scatter estimation results.**

***Index Terms*—Dual-ring stationary CT, Deep scatter estimation, Frequency attention, Spare detector.**

I. Introduction

Multi-source stationary CT, a class of innovative, non-standard CT architecture, undergoing development for decades, offers the potential to significantly improve image acquisition temporal resolution through two key strategies: rapid source switching and simultaneous multi-source emission. However, the latter approach introduces a major challenge for scatter artifact correction, as the primary signal will be contaminated by both forward and cross scattered components. In general, scatter correction methods for CBCT include hardware techniques of scatter rejection and software methods of scatter estimation[1]. Hardware techniques, such as collimators, bowtie filters, and anti-scatter grids, are designed and added to the CT system to reject scatter signal. Software-based scatter estimation methods primarily include three categories: Monte Carlo simulations, scatter kernel superposition (SKS) techniques, and deep learning-based approaches. Monte Carlo simulations are the gold standard for scatter estimation, but they are computationally expensive. SKS techniques are less computationally expensive than Monte Carlo simulations, but they could be inaccurate. Deep learning-based approaches have the potential to achieve high accuracy while being computationally efficient, but they are still a relatively new area of research. Comparing with scatter estimation in CBCT, it is challenged to address both forward and cross scatter correction in multi-source CT (MSCT). For achieving accurate scatter estimation in MSCT, it is essential to combine the extra information into the deep learning methods.

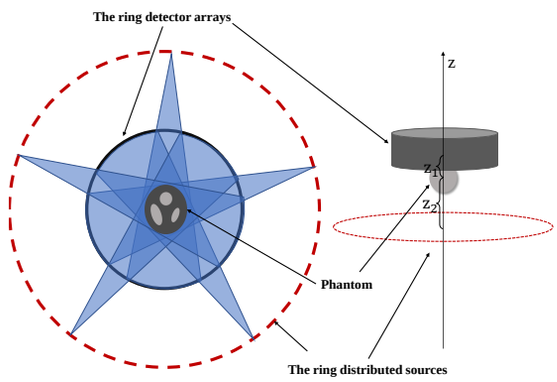


Fig. 1. The architecture of multi-source dual-ring stationary CT.The top view is on the left, and the side view is on the right. There are Z-shifts between ring detector arrays, phantom and ring distributed sources.

As shown in Fig.1, we have proposed a dual-ring stationary CT architecture[2, 3] before. This manuscript presents a novel deep learning framework (ComptonNet) for scatter estimation of this multi-source stationary CT, which leverages large-angle compton scatter signal from spare detector. ComptonNet includes two networks in series: 3-channel Fusion Encoder-decoder network (FED-net) for cross scatter estimation and frequency Unet for forward scatter estimation. And it is an end-to-end network, which could maximize information utilization between the forward signal and large-angle compton scatter signal, where Frequency Attention Modules in the ComptonNet ensure low parameter complexity and smooth low-frequency estimation.

This project was supported in part by the National Key Research and Development Program of China (No.2022YFC3302501), and the National Natural Science Foundation of China (No.12327809).

Yingxian Xia, Zhiqiang chen, Li Zhang, Yuxiang Xing, and Hewei Gao are 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 author. Email: hwgao@tsinghua.edu.cn)

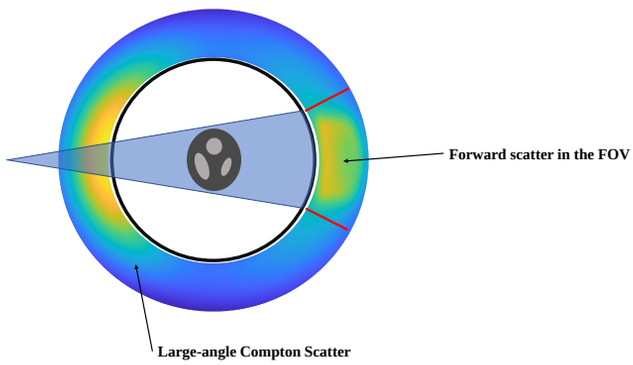


Fig. 2. Scatter signal from one source in the dual-ring stationary CT. The forward scatter signal in the field of view (FOV) includes compton scatter, rayleigh scatter and multiple scatter. The scatter signal received by out-of-FOV detectors is mainly composed of large-angle compton scatter signal, which could help estimate forward scatter signal.

II. Methods

A. Data Generation

In order to acquire raw total signal and the counterpart scatter signal in quantity, datasets are generated using a Monte Carlo simulation software package called MCGPU[4]. The simulation geometry of MCGPU is the common CBCT using a flat panel detector, so geometry changes of MCGPU are made for simulating the geometry of the dual-ring stationary CT. The geometry parameters of the dual-ring stationary CT Monte Carlo simulation are listed in Table I.

TABLE I
Parameters of Monte Carlo simulations

Parameters	Values
Radius of X-ray source ring (R_1)	110 cm
Radius of detector ring (R_2)	40 cm
Height of detector ring	25.6 cm
Z-shift distance of detector ring (z_1)	10.5 cm
Z-shift distance of X-ray source (z_2)	-24.7 cm
Detector element size	1.39 mm × 1 mm
Size of phantom	256×256×128
Voxel size of phantom	1 mm × 1 mm × 1 mm
Photon number of one projection	2e10

The image objects to generate the datasets are the head, thorax and abdomen phantoms, which generate from available CT scans from TCIA[5]. The conversion from CT scans to the MCGPU voxel phantom format is finished according to the document of MCGPU. The simulation totally uses 67 phantoms, including 11 thorax phantoms, 26 head phantoms and 30 abdomen phantoms. For each CT phantom, 60 projections covering 360 deg are simulated respectively. We add together five equi-spaced stacks of signals for simulating CT imaging with five X-ray sources beaming simultaneously. In this way, we generate a multi-source dual-ring stationary CT image dataset containing 804 scatter-contaminated ring-detector images, and the corresponding 804 scatter ring-detector images. Due to the different scatter data scales in cross scatter estimation and forward scatter estimation, the input projection for Cross

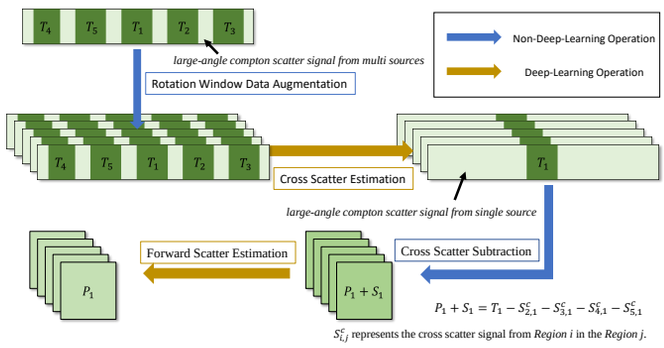


Fig. 3. ComptonNet: the end-to-end deep learning framework for Scatter Estimation in Multi-source Dual-ring Stationary CT.

scatter estimation is rebinning to 720×256 for 360 deg range and the input projection for forward scatter estimation is rebinning to 250×256 for 50 deg range. We simulate CT images with five X-ray sources beaming simultaneously, so they are 22-deg-wide spare detector between the field-of-views (FOVs). The spare detectors receive pure scatter signal in the simulation, which are used for guiding the scatter estimation in the FOVs.

B. Scatter Estimation

The workflow of ComptonNet is shown in Fig.3. When there are five sources beaming together in the dual-ring stationary CT, there will be five FOVs in the raw multi-source global CT image. First, the global CT images could be rotated and divided with different FOV placed in the center for realizing data augmentation. And then, the five global CT images can be used to generate five corresponding cross scatter signals, via FED-net. In Cross Scatter Subtraction procedure, we could subtract the cross-scatter signal from source 2, 3, 4, and 5 in the FOV 1, and up-sampling the signal in the FOV from 100×256 to 250×256. The rest FOVs can be done in the same manner. Until now, there will be no cross signals in the FOVs. Finally, we generate the primary signal by subtracting forward scatter estimated by the Frequency Unet.

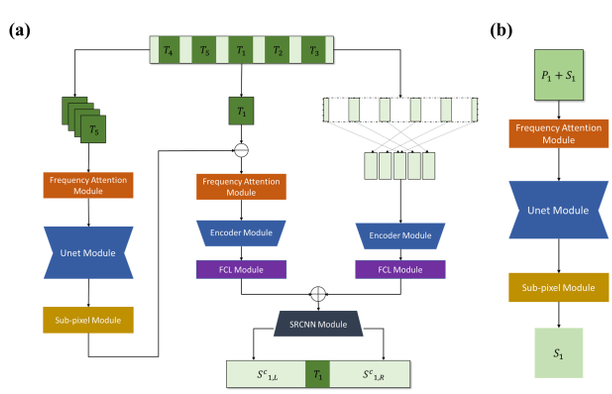


Fig. 4. The architecture of 3-channel FED-net(a) for cross scatter estimation and frequency Unet(b) for forward scatter estimation.

As shown in Fig.4, in cross scatter estimation, we use the FED-net to estimate the global cross scatter signal. Center

FOV channel, spare data channel and other four FOV channel would be used together for estimating the cross-scatter signal. The main channel of FED-net is to estimate global scatter signal from the primary signal in the FOV. The main channel has four kinds of modules: Frequency Attention Module, Encoder Module, FCL Module and SRCNN Module. Frequency Attention Module and Encoder Module realize the low dimensional feature extraction of input signal. FCL Module and SRCNN Module could recover the global cross scatter signal. Besides the signal in the center FOV, the other four signals could also help realize a rough estimation of cross scatter, so we could subtract the output of the other four FOV channel from the input of the main channel. Also, Center FOV channel and spare data channel would fuse both the low dimensional feature extraction together, and then take it as the input of SRCNN Module for generating cross scatter signal for better estimation.

Unet has been proven to be a powerful network architecture to estimate the forward scatter signal for single source CT imaging. However, high parameter complexity and unrobust prediction results reduce the credibility of the architecture. In order to estimate smooth, low-frequency scatter signal, we design the Frequency Attention Module with frequency attention mechanism [6, 7] and use shallow Unet Module avoiding the high parameter complexity. sub-pixel Module realizes the pixel rearrangement and ensures the output size consistent with the input signal.

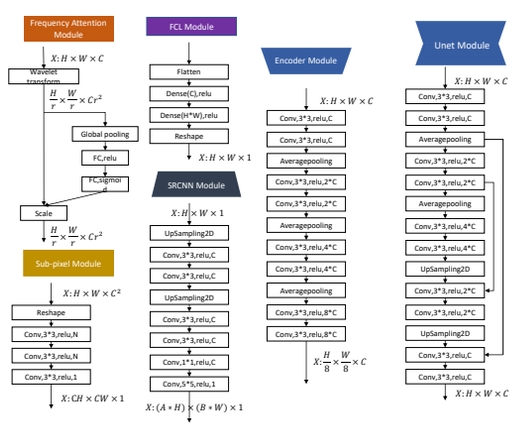


Fig. 5. The specific composition of different modules in ComptonNet.

III. Results

A. Experimental Setup

Due to the two independent networks in our framework, we first carry out pre-learning on FED-net and frequency Unet. For both of them, we all can get the total signal as the inputs and the corresponding scatter signal as the labels. After learning separately, we can take the whole framework as an end-to-end network for training aiming for better predictions of cross and forward scatter. Each type of phantoms has one used as test data, and the other 64 phantoms are using for training, where 20% data is used as validation datasets. The target scatter signals are noisy and vary greatly, so Mean Absolute Percentage Error (MAPE) loss function is usually used for deep scatter estimation. The networks are trained using Adam optimizer to minimize the MAPE to the target MC scatter simulations. To evaluate the performance of proposed deep scatter estimation method, we use MAPE and Mean Absolute Error(MAE) as the evaluation metrics.

B. Cross and Forward Scatter Estimation

Table II shows the parameters, pixel-wise MAPE and MAE of the two sub-networks on the three test phantoms. As shown in Fig.5, APE from the predictions of our framework to the GT are lower than 5% for the whole projection image. FED-net could correct the cross scatter with high precision. In the results of frequency Unet, throax and head phantom have relatively bigger errors in the middle of phantom.

TABLE II
PARAMETER NUMBER AND QUANTITATIVE TEST RESULTS FOR ALL PROPOSED NETWORK ARCHITECTURES.

Networks	Parameters	MAPE	MAE
FED-net	2.8 M	1.97%	0.43
Frequency Unet	1.1 M	2.45%	0.58

Prediction results also remain the property of low-frequency smoothness. ComptonNet is capable to maintain the high accuracy while reducing the number of parameters comparing to the common Unet networks, whose common parameter number is about 10 million. The Quantitative test results of framework with/without spare detector data are shown in Table III. Using spare detector data could decrease MAPE from 1.16% to 0.80% and MAE from 1.32 to 0.90. As shown in Fig.8, under the guidance of spare detector data, the deep scatter estimation would bring better correction, especially in the middle of test phantom.

TABLE III
QUANTITATIVE TEST RESULTS OF COMPTONNET WITH/WITHOUT SPARE DETECTOR DATA GUIDANCE.

Spare detector data guidance	MAPE	MAE
With	0.80%	0.90
Without	1.16%	1.32

IV. Conclusions

This work proposed a novel spare detector data-guided deep learning framework (ComptonNet) for scatter estimation in multi-source dual-ring stationary CT. The framework leverages a 3-channel Fusion Encoder-decoder network (FED-net) to estimate cross scatter under the guidance of spare detector data, followed by a frequency Unet for forward scatter estimation. Frequency Attention Modules ensure both low parameter complexity and smooth low-frequency results. Furthermore, the end-to-end ComptonNet effectively utilizes spare detector data, improving overall scatter estimation accuracy. Compared to the networks without spare detector data, ComptonNet reduces MAPE from 1.16% to 0.80% and MAE from 1.32 to

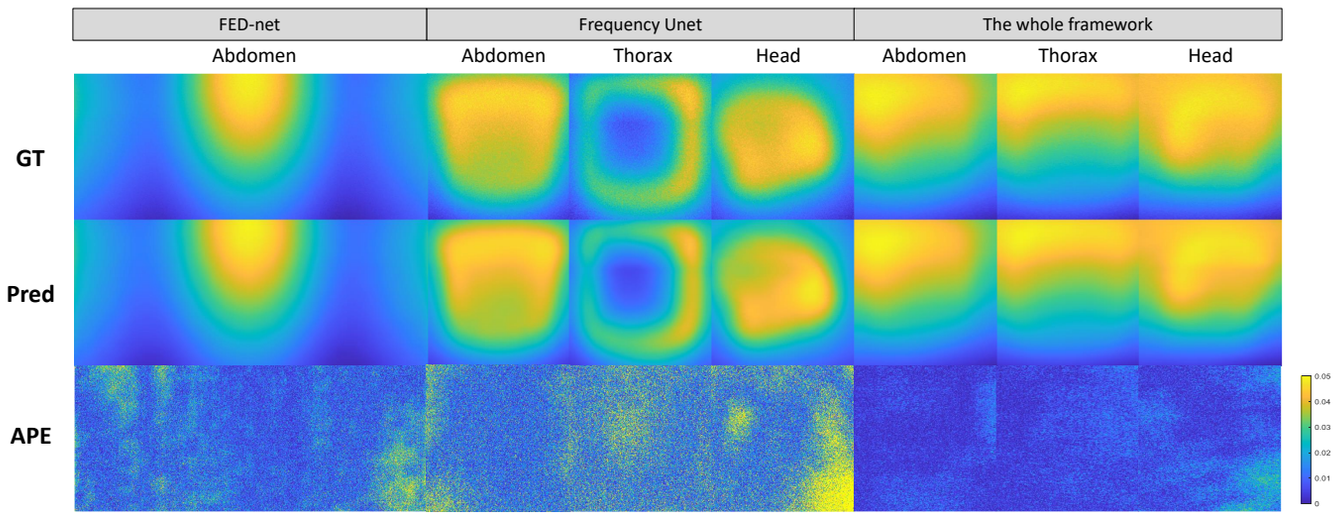


Fig. 6. Qualitative results of FED-net, frequency Unet and the whole framework for three different test phantoms. The images in the first line are the ground truth(GT) and the images in the second line are the predictions(Pred) of networks. They are windowed to give a good visual representation. Absolute percentage errors (APE) are shown in the range of $[0, 5\%]$.

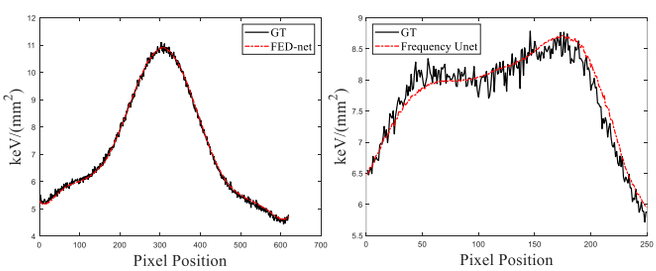


Fig. 7. Center horizontal profile of predictions of test head phantom.

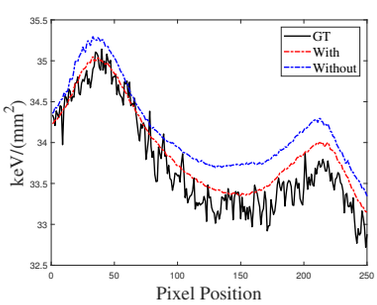


Fig. 8. Center horizontal profile of predictions of the whole framework for the test thorax phantom shown in Fig.6.

0.90. This demonstrates the effectiveness of employing large-angle scatter data from spare detector and the superior performance of our deep learning framework for scatter estimation in multi-source stationary CT.

References

[1] Xu Zhuo, Yuchen Lu, Yuexuan Hua, Hanxi Liu, Yikun Zhang, Shilei Hao, Lin Wan, Qingguo Xie, Xu Ji, and Yang Chen. Scatter correction for cone-beam CT via scatter kernel superposition-inspired convolutional neural network. *Physics in Medicine & Biology*, 68(7):075011, March 2023. Publisher: IOP Publishing.

[2] Yingxian Xia, Li Zhang, Yuxiang Xing, Zhiqiang Chen, and Hewei Gao. Generalized-equiangular geometry ct: Concept and shift-invariant fbp algorithms. *Medical Physics*, 50(8):5150–5165, 2023.

[3] Zhiqiang Chen, Li Zhang, Xin Jin, YANG Hongkai, Xiaofei Xu, Zhenhua Zhao, Siyuan Zhang, and Changyu Chen. Imaging system and method for radiographic inspection, April 20 2023. US Patent App. 17/967,039.

[4] Andreu Badal and Aldo Badano. Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit. *Medical Physics*, 36(11):4878–4880, 2009.

[5] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al. The cancer imaging archive (tcia): maintaining and operating a public information repository. *Journal of digital imaging*, 26:1045–1057, 2013.

[6] Huaibo Huang, Ran He, Zhenan Sun, and Tieniu Tan. Wavelet-SRNet: A Wavelet-Based CNN for Multi-scale Face Super Resolution. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 1698–1706, Venice, October 2017. IEEE.

[7] Kai Xu, Minghai Qin, Fei Sun, Yuhao Wang, Yen-Kuang Chen, and Fengbo Ren. Learning in the frequency domain. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1740–1749, 2020.

Experimental study on charge sharing correction: hardware or software?

Shengzi Zhao, Le Shen, and Yuxiang Xing, *Senior Member, IEEE*

***Abstract*—In computed tomography (CT), the use of photon counting detectors (PCDs) brings valuable advantages, particularly in the capture of spectral information. Nevertheless, charge sharing events clearly limit their widespread applications. Based on the fact that charge sharing is an effect of mutual influence among adjacent pixels, we experimentally studied charge sharing compensation by anti-coincidence circuit (hardware) and a simple polynomial fitting method (software). In the polynomial-fitting compensation method, we extract information of charge sharing from the photon counts of a small detector patch to correct the charge sharing in the central pixel of the patch. We designed a practical experiment with an XCounter detector to compare the performance of the polynomial fitting method and anti-coincidence circuits. In order to focus on the effect on charge sharing only and minimize the interference from other factors, we carefully set some details of the experiments. Beyond our expectation, the polynomial fitting method alone may give competitive performance compared with anti-coincidence circuits. Lower bias for low energy bin and lower noise output in both bins are observed. The results of this study may offer new insights or ideas for the strategies for charge sharing correction or compensation.**

***Index Terms*—photon counting detector, charge sharing, polynomial fitting, spectral CT.**

I. Introduction

DUE to the capability of Photon Counting Detectors (PCDs) to provide photon counts in different energy bins, they have great potential to enhance the imaging quality of spectral computed tomography (CT). In general, the pixel size of PCDs is smaller than that of Energy Integrating Detectors (EIDs), allowing for reconstructing images with higher resolution. When using PCDs to collect data, a low threshold is typically set to exclude counts with very low energies, thereby reducing image noise. However, PCDs also suffer from pulse pile-up and charge sharing that introduce distortions to the collected data and diminish image quality, thereby posing challenges to the widespread application of PCDs. In this work, we only focus on charge sharing problem.

In this work, we only consider charge sharing events happening between the pixel of interest (POI) and its adjacent pixels, just like the charge sharing addition (CSA) methods do. When an X-ray photon impinges on a pixel of a PCD, its adjacent pixels could also collect charges [1], resulting in the generation of several signals. Each of these signals corresponds to an energy lower than the energy of the incident photon. Considering a PCD with one low and one high energy bins, charge sharing events have totally different impacts on the two bins [2]. Therefore, charge sharing leads to the cross-talk between the two energy bins and among adjacent pixels. Consequently, the detector will output distorted data, posing challenges for subsequent reconstruction.

Currently, there are numerous hardware- or software-based methods to compensate for or correct charge sharing or mitigate its impact on the quality of reconstructed images. The hardware-based methods include Charge Sharing Addition (CSA), Charge Sharing Discrimination (CSD), and introducing coincidence counters. For CSA, additional inner-pixel communication circuits are incorporated into the signal processing module of the detector to sum the signals of shared charges and relocate them [3], [4]. For CSD, charge sharing events are distinguished from all events and then rejected [5]. For coincidence counters, Hsieh *et al*. proposed a detector in which each pixel has an extra coincidence counter to count charge sharing events [6]. Based on that, Taguchi *et al*. proposed a detector in which each pixel has several extra coincidence counters to record numbers and directions of charge sharing events [1]. For software-based methods, Cammin *et al*. utilized a Gaussian function and a constant function to fit distorted spectrum [7]. Xu *et al*. modeled different interactions between X-rays and materials, and obtained the energy response function of a PCD that takes charge sharing into account [8].

Our objective is to compensate for charge sharing in the detected data in a calibration manner. Previously, we analyzed the impacts of charge sharing by considering a single pixel and established equations relating the numbers of incident photons of POI and its adjacent pixels [2]. Inspired by it, we utilized a polynomial fitting method to deal with the photon counts of POI and its adjacent pixels, and compensate for charge sharing for POI. In actual experiments, we observed that sometimes the polynomial fitting method might outperform the hardware-based method by anti-coincidence circuits.

This paper was submitted to CT meeting for review on Jan. 23th, 2024. This work was supported in part by National Natural Science Foundation of China under NNSFC 12275151.

All the authors are with the Department of Engineering Physics, Tsinghua University, Beijing, China. They are also with Key Laboratory of Particle & Radiation Imaging (Tsinghua University), Ministry of Education, Beijing, China. (e-mail: zhaosz21@mails.tsinghua.edu.cn, xingyx@mail.tsinghua.edu.cn, shenle.2007@tsinghua.org.cn).

II. METHOD

In this section, we begin with an introduction to data pre-processing, i.e., the formation of inputs and outputs for polynomial fitting. Then, we explain our implementation of the polynomial fitting. After that, we present the specific configurations of the experiment. Finally, some metrics are defined to show and compare results.

A. The inputs and outputs for polynomial fitting

In this work, we only consider a detector with two energy bins. The simplest and most intuitive approach is to utilize the detector counts as input. However, the counts are from hundreds to thousands in both the low and high energy bins. Therefore, in the polynomial method, higher-order terms generated by these counts exhibit significant differences in magnitude, which poses difficulties for the convergence of the iteration. We employ effective attenuation lengths obtained from detected counts as follows as inputs that could also compress the data range and help the fitting converge,

$$Y_{c,i}^{\mathrm{det}} = -\ln \frac{N_{c,i}^{\mathrm{det}}}{N_{c,i}^{\mathrm{det,air}}}, \qquad c \in \{\mathrm{L}, \mathrm{H}\}. \tag{1}$$

Here, $Y_{c,i}^{\mathrm{det}}$ serves as the input to the fitted polynomial. $N_{c,i}^{\mathrm{det}}$ is the detector counts in different energy bin c of pixel $\#i$. $N_{c,i}^{\mathrm{det,air}}$ is the photon counts when X-ray only goes through air of pixel $\#i$.

We would like to analyze the compensation effects separately for the low and high energy bins because charge sharing has totally different impacts on the two bins. Material thicknesses commonly used in spectral CT depend on the data from both the low and high energy bins, so they are not able to achieve our goal. In addition, we hardly obtain the ground truths of incident photons. As an alternative, the ground truths of effective attenuation lengths for a POI $Y_{c,\mathrm{POI}}^{*}$ is utilized as labels to optimize the coefficients of polynomials. We could obtain them as following,

$$Y_{c,\mathrm{POI}}^{*} = -\ln \frac{\int_{\text{energy bin } c} S_{\mathrm{eff}}(E) \exp\left(-\int_{l_{\mathrm{POI}}} \mu(E)\mathrm{d}l\right)\mathrm{d}E}{\int_{\text{energy bin } c} S_{\mathrm{eff}}(E)\mathrm{d}E} \tag{2}$$

Here, $S_{\mathrm{eff}}(E)$ is the effective spectrum. $\mu(E)$ is the attenuation coefficient. $\int_{l_{\mathrm{POI}}} \mu(E)\mathrm{d}l$ is the integration of the attenuation coefficients along the ray path of a POI. After optimizing, the outputs of polynomials are denoted as $\hat{Y}_{c,\mathrm{POI}}$.

B. The polynomial fitting method for charge sharing correction

Charge sharing leads to cross-talk among adjacent pixels, so we focus on more than POI but also its neighboring pixels as shown in [2] . To compensate for charge sharing for a POI, we choose the $Y_{c,i}^{\mathrm{det}}$ of a 5×5 patch centered on the POI as the inputs. We utilize two polynomials to fit the ground truths of effective attenuation lengths in the two energy bins of POI, individually. For each energy bin, the polynomial takes 50 logarithms of photon counts of the two energy bins in the patch and outputs the effective attenuation lengths (denoted by $\hat{Y}_{\mathrm{L,POI}}$ and $\hat{Y}_{\mathrm{H,POI}}$) only for the POI.

Currently, we only choose a quadratic polynomial. It includes a constant term, linear terms, square terms, and quadratic cross-terms. Since each pixel is highly related to only its eight surrounding pixels [2], we remove some cross-terms to simplify the polynomial. Only when two logarithms, Y_1^{det} and Y_2^{det}, come from the same pixel or two adjacent pixels are they included in the polynomial, as shown in Fig. 1.

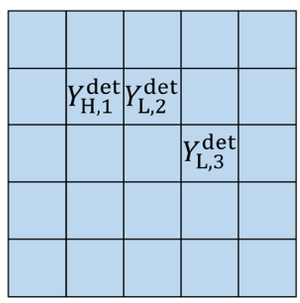


Fig. 1. A 5×5 patch processed by the polynomial fitting method. It also shows which quadratic terms are included. $Y_{\mathrm{H},1}^{\mathrm{det}} Y_{\mathrm{L},2}^{\mathrm{det}}$ and $Y_{\mathrm{L},2}^{\mathrm{det}} Y_{\mathrm{L},3}^{\mathrm{det}}$ are included, and $Y_{\mathrm{H},1}^{\mathrm{det}} Y_{\mathrm{L},3}^{\mathrm{det}}$ is not included.

C. Configurations of the experiment

We tested our method on an imaging system with an XCounter PCD in our laboratory as shown in Fig. 2. PMMA blocks combined with an aluminum block of different thicknesses as shown in Fig. 2(b) were scanned to get dataset for optimizing the coefficients in the polynomials, i.e., the training dataset. The thicknesses of PMMA blocks are from 6 mm to 52 mm with a step of 2 mm and the thicknesses of the aluminum block are from 2 mm to 20 mm with a step of 2 mm. A 50 mm diameter PMMA cylinder combined with twelve 5 mm diameter aluminum cylinders as shown in Fig. 2(c) was scanned to get test data independent of training data.

The XCounter detector has 1,536 × 128 pixels in total and each pixel has an area of 0.1 mm × 0.1 mm. For data collection, the X-ray source is set to 130 kVp and 0.3 mA, the source-to-detector distance is 455 mm and the source-to-isocenter distance is 316 mm. Under each view, the time for scanning one frame is 0.1 seconds. As for the setting of energy bins, we choose a narrow low energy bin [40, 50] keV to minimize the impact of beam hardening as much as possible when calculating effective attenuation lengths. For the high energy bin, we choose a wide energy bin [50, 130] keV because beam hardening is not significant in the high energy bin, and we hope to collect as many photon counts as possible.

We collected two sets of data, with and without anti-coincidence modes from the XCounter PCD and refer them as no anti-coincidence (noAnti) data and anti-coincidence (Anti) data, respectively. For both Anti and noAnti data and for a POI, we obtain effective attenuation lengths $Y_{c,\mathrm{POI}}^{\mathrm{det}}$ refer to (1) as references. In this way, the $Y_{c,\mathrm{POI}}^{\mathrm{det}}$ from detected data directly and $\hat{Y}_{c,\mathrm{POI}}$ obtained by polynomial fitting could be comparable.

We utilized noAnti data and Anti data to optimize coefficients of polynomials and conduct tests, respectively. Both of the two training datasets have 1,564,800 patches, and both of the two test datasets have about 190,000 patches.

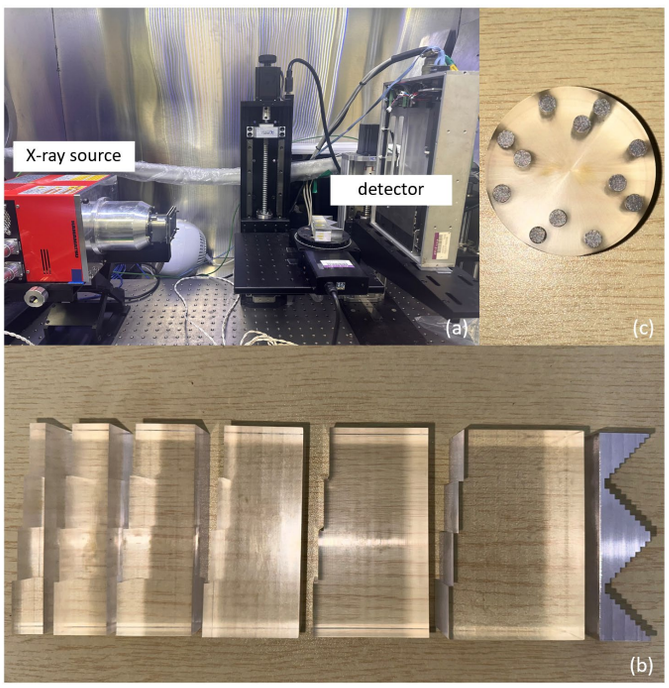


Fig. 2. (a) Photograph of the imaging system. (b) Photograph of the phantoms for training. (c) Photograph of the phantoms for the test.

D. *Quantitative metrics*

When performing polynomial fitting and using the coefficients obtained from the fitting for testing, it is necessary to ensure that the distribution of the training data covers the distribution of the testing data. The number of charge sharing events is significantly related to the numbers of incident photons. Only considering charge sharing, the numbers of incident photons and charge sharing events jointly determine the detected counts. Therefore, we sum the counts of all pixels in a patch to roughly estimate the data distribution as follows,

$$N_{\text{patch}}^{c} = \sum_{i=1}^{N_{\text{pixel}}} N_{c,i}^{\text{det}} \tag{4}$$

Here, N_{pixel} is the number of pixels in a patch and $N_{\text{pixel}} = 5 \times 5$.

Two quantitative metrics, the average of absolute normalized bias (AANBias) and the average of normalized standard deviation (ANSTD) are utilized to evaluate the results,

$$\text{AANBias} = \frac{1}{2N} \sum_{n=1}^{N} \sum_{c\in\{\text{L,H}\}} \frac{\left| \frac{1}{N_{\text{sample}}} \sum_{s=1}^{N_{\text{sample}}} Y_{c,n,s} - Y_{c,n}^{*} \right|}{Y_{c,n}^{*} + \epsilon_c}, \tag{5}$$

$$\text{ANSTD} = \frac{1}{2N} \sum_{n=1}^{N} \sum_{c\in\{\text{L,H}\}} \frac{\sigma\left(\{Y_{c,n,s}\}\right)}{Y_{c,n}^{*} + \epsilon_c}. \tag{6}$$

Here, $Y_{c,n}^{*}$ are the ground truths of effective attenuation lengths refer to (2) for the nth PMMA and aluminum thickness. $Y_{c,n,s}$ are the outputs of the polynomial fitting method for the sth noise realization. N is the number of all the different thicknesses, and N_{sample} is the number of noise realizations for each thickness in the test dataset. $\sigma(\{A_s\})$ represents the standard deviation of the set $\{A_s\}$.

Both of the above indicators are relative values. When the thicknesses are very small, approaching zero, a small absolute error could lead to a significantly relative error, which is not conducive to measuring the performance of different methods. Therefore, we ignored samples with effective attenuation lengths close to zero for now.

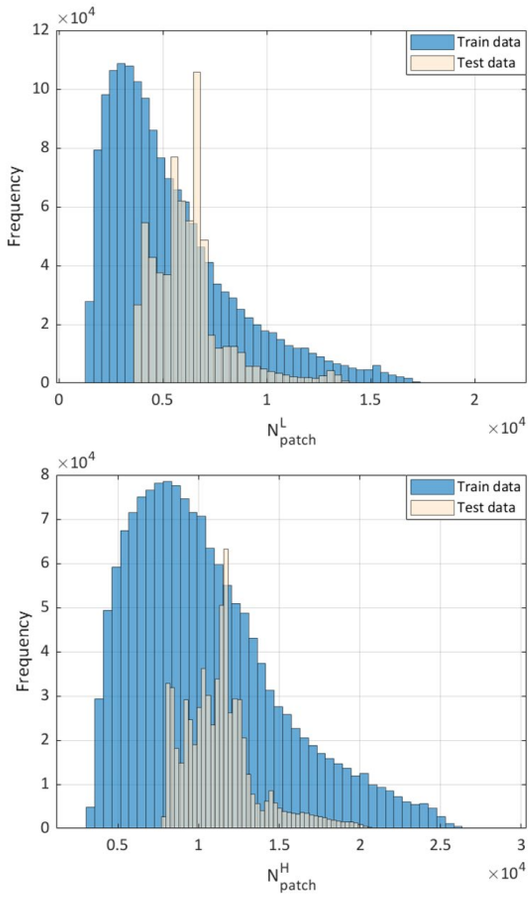


Fig. 3. Distributions of $N_{\text{patch}}^{c}, c \in \{\text{L,H}\}$ for training and test datasets for noAnti data.

III. Experimental Results

A. *The distribution of dataset*

The distributions of noAnti data are shown in Fig. 3. Although the distributions of the training and test data are not exactly the same, we only ensure that the photon count range of the test dataset is covered by the training dataset.

B. *AANBias and ANSTD for polynomial fitting*

Results of AANBias and ANSTD are plotted in Fig. 4. We initially focus on the results obtained by directly processing the detected data refer to (3). For noAnti data, the effective attenuation length Y_c^{det} directly computed from N_c^{det} is of big error in both energy bins because of charge sharing. Higher bias

in the low energy bin implies that it is more severely affected by charge sharing than the high energy bin. After hardware correction, the results obtained from the Anti data show a significant reduction in bias. In addition, hardware correction mitigates the impacts that charge sharing reduces the counts in the high energy bin and increases the counts in the low energy bin. Therefore, there are fewer counts in the low energy bin for Anti data than those for noAnti data, leading to higher standard deviation for Anti than that for noAnti. The standard deviations in the high energy bin could also be explained similarly.

After we utilize the polynomial fitting method to process noAnti data, the bias slightly decreases whereas the standard deviation significantly decreases in the high energy bin. A large bias indicates that there is some mismatch between the second-order polynomial and the charge sharing pattern in the high energy bin. We guess that the polynomial decreases the noise by reducing the total error. For the low energy bin, the effectiveness of polynomial fitting is entirely different. Both the bias and the standard deviation show significant decreases, even lower than the results from hardware correction. These results suggest that the polynomial can better describe the charge sharing pattern in the low energy bin than in the high energy bin. This also tells that charge sharing exhibits different patterns between the low and high energy bins.

Next, we focus on the results of Anti data processed by polynomial fitting. Compared with the results obtained by (3), the bias increases while the standard deviation decreases in the high energy bin. Partially mismatched polynomial fit a lot to the noise under the requirement of reducing the total error. In the low energy bin, the polynomial still compensates for charge sharing significantly, decreasing bias and standard deviation at the same time. Furthermore, polynomial fitting may also mitigate the impacts of some other non-ideal factors, making the fitting results better than hardware correction.

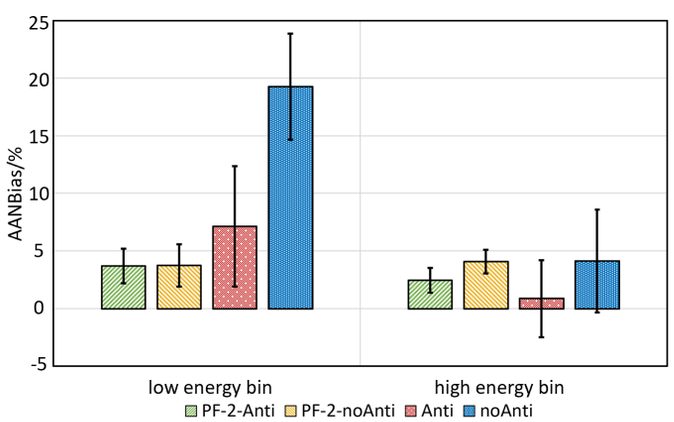


Fig. 4. Results from experiments. The error bars mean ANSTD. “Anti” and “noAnti” mean the results are obtained directly refering to (3). “PF-2” means the results are estimated by the second-order polynomial fitting method.

IV. Discussion

We use a polynomial fitting method to compensate for charge sharing, and compare the results with that of hardware correction. The differences in the results of low and high energy bins suggest that we may need to compensate for charge sharing for low and high energy bins in different ways. Sometimes, using a simple polynomial fitting might be a usable choice except for hardware correction. After utilizing hardware, results could be further improved by using a polynomial. Last but not least, our polynomial method is a simple method and other optimization method may be employed to achieve better results.

However, we still need more extensive experimental verification of the above conclusions. Different conditions may lead to different results.

V. Conclusion

In this work, we employ a simple polynomial fitting to compensate for charge sharing. Beyond our expectation, it performs much better than hardware correction in the low energy bin but only has slight effectiveness in the high energy bin. We believe that this phenomenon is worth further investigation. That’s our future work.

Acknowledgment

This work is supported by Center of High Performance Computing, Tsinghua University.

References

[1] K. Taguchi, “Multi-energy inter-pixel coincidence counters for charge sharing correction and compensation in photon counting detectors,” *Medical Physics*, vol. 47, no. 5, pp. 2085-2098, Jan. 2020, DOI: 10.1002/mp.14047.

[2] SZ Zhao, K. Taguchi, XY Xing, July, 2023, “A new analysis and compensation method for charge sharing in PCD.” Presented at Fully3D. [Online]. Available: https://arxiv.org/abs/2310.16846.

[3] T. Koenig, *et al*., “Charge summing in spectroscopic x-ray detectors with high-Z sensors,” *IEEE Transactions on Nuclear Science*, vol. 60, no. 6, pp. 4713-4718, Dec. 2013, DOI:10.1109/TNS.2013.2286672.

[4] T. Poikela, *et al*, “Timepix3: a 65K channel hybrid pixel readout chip with simultaneous ToA/ToT and sparse readout,” *Journal of instrumentation*, vol. 9, no. 05, pp. 05013, May. 2014, DOI:10.1088/1748-0221/9/05/C05013.

[5] E. Fredenberg, *et al.*, “Energy resolution of a photon-counting silicon strip detector,” *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 613, no.1, pp. 156-162, 2010, DOI:10.1016/j.nima.2009.10.152.

[6] SS. Hsieh, “Coincidence counters for charge sharing compensation in spectroscopic photon counting detectors,” *IEEE Transactions on Medical Imaging*, vol.39, no.3, pp. 678-687, Mar. 2020, DOI:10.1109/TMI.2019.2933986.

[7] J. Cammin, *et al*., “A cascaded model of spectral distortions due to spectral response effects and pulse pile-up effects in a photon-counting x-ray detector for CT,” *Medical Physics*, vol.41, no.4, pp. 041905, Mar. 2014, DOI:10.1118/1.4866890.

[8] X Ji, *et al*., “An analytical model of energy response function for CdTe-based photon counting detectors with inter-pixel communication for anti-coincidence corrections.” Presented at Medical Imaging 2022: Physics of Medical Imaging. SPIE. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/12031/120314N/An-analytical-model-of-energy-response-function-for-CdTe-based/10.1117/12.2611848.full.

Spectral quantification in different lumen diameters for cardiovascular applications

Leening P. Liu[1,2], Martin V. Rybertt[1], Pouyan Pasyar[1,2], Nadav Shapira[1], Harold I. Litt[1], Peter B. Noël[1]

Abstract— **The first clinical dual-source photon-counting CT couples high spatial resolution with spectral imaging that is advantageous to imaging of small vessels, such as the coronary arteries, in cardiovascular disease. While both the high spatial resolution and quantification accuracy have been established in PCCT, the effect of lumen size on spectral quantification has not been evaluated. Phantoms with an internal tube diameter ranging from 4 to 12 mm were printed with calcium-based polylactic acid filament to mimic a coronary artery. These diameter phantoms were filled with solutions with iodine concentrations of 2, 5, and 10 mg/mL and scanned with phantoms of varying sizes on a PCCT. Virtual monoenergetic images (VMI) at 70 keV, iodine density maps, and virtual non-contrast maps were measured to determine the effect of lumen diameter on spectral quantification at different iodine concentrations, radiation doses, and phantom sizes. Each evaluated spectral result exhibited consistent quantification at lumen diameters greater than 6 mm with all phantom sizes. VMI 70 keV were within ±15, ±12, and ±4 of VMI 70 keV at a lumen diameter of 12 mm and the small phantom for iodine concentrations of 2, 5, and 10 mg/mL. At a lumen diameter of 4 mm, significant deviations were present in VMI 70 keV, iodine density maps, and VNC with large phantoms, which averaged 55 HU, 1.4 mg/mL, and 61 HU at an iodine concentration of 5 mg/mL, respectively. The consistent spectral results across lumen diameters demonstrated the synergy between high spatial resolution and quantification that will spur the use of quantitative metrics and development of new applications in diagnostic cardiac imaging.**

Index Terms— **Photon-counting CT, Cardiac CT, Spectral CT, Iodine Quantification**

I. Introduction

CARDIOVASCULAR diseases, including coronary artery disease, have been the leading cause of death worldwide, with deaths and disability continuing to rise the past 30 years [1]. This epidemic of heart disease has established a need for effective non-invasive assessment of heart structures and function. One such tool is cardiac computed tomography (CT) that utilizes iodinated contrast to image internal structures of the heart, such as coronary arteries [2, 3]. The use of contrast enables the isolation of the coronary tree, which is prone to fibrous and calcified plaques in coronary artery disease. Accurate assessment of these plaques for both their composition and size is critical to determining the degree of stenosis in the vessel [4], and thus diagnosis and treatment of coronary artery disease. With conventional CT, this evaluation is limited by the spatial resolution of the scanner, which makes visualization of lumen coronary arteries difficult, and the blooming of calcified plaques that result in overestimation of their size.

The introduction of the first clinical dual-source photon-counting CT (PCCT) overcome these challenges in conventional CT for improved diagnosis of cardiovascular disease. Its photon-counting detector facilitates the acquisition of conventional images in ultra-high resolution mode (0.2 mm slice thickness). Additionally, high resolution (0.4 mm slice thickness) spectral results are available in every scan [5]. Of the spectral results, virtual monoenergetic images (VMI) at high energies minimize calcium blooming of plaques to better estimate the size of plaques [6] while also reducing metal artifacts from stents to evaluate restenosis [7], [8]. Iodine density and virtual non-contrast (VNC) maps, on the other hand, separate iodinated contrast/vessels from other anatomical structures to identify stenoses in the coronary tree. Additionally dual-source PCCT can utilize both sources simultaneously to acquire images at a high temporal resolution with minimal motion artifacts. As a result, PCCT can obtain spectral images at high spatial resolution with minimal motion artifacts for the diagnosis of coronary artery disease [7, 9].

With high resolution spectral images, PCCT promote quantification of spectral results in small structures such as calcified plaques and the lumen of coronary arteries. For the use of quantitative metrics or diagnostic imaging, it necessitates quantitative accuracy to ensure confidence in quantification and thus diagnosis. While quantification accuracy has been established in PCCT [10], larger diameter targets were typically utilized in these assessments. As a result, the effect of object size on spectral quantifications is largely unknown. Thus, the coupling of spatial resolution and spectral quantification may have an impact on diagnostic assessment, particularly in small structures.

II. Methods

A. Diameter phantom

To evaluate the effect of vessel size on quantification of spectral results, 3-D printed diameter phantoms were designed as inserts for a body phantom (Multi-energy CT phantom, Gammex). They were 3-D printed (TAZ Sidekick, LulzBot) with a calcium-based polylactic acid (PLA) filament (StoneFil,

This study was funded by the National Institutes of Health (R01EB030494) and Siemens Healthineers.

L.P. Liu and P. Pasyar are with the Department of Bioengineering and Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. M.V. Rybertt, N. Shapira, H.I. Litt, and P.B. Noël are with the Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. (email: leening@seas.upenn.edu)

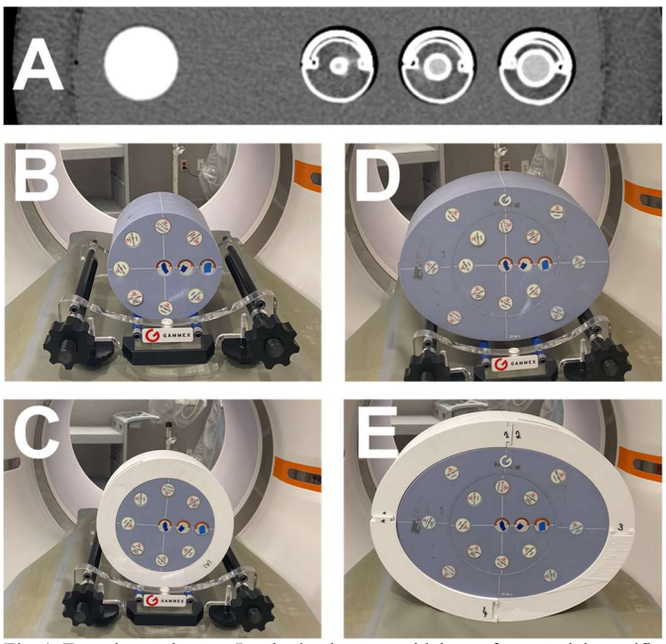


Fig. 1. Experimental setup. In a body phantom with bores for material-specific inserts, 3-D printed phantoms of different diameters (A, 4, 8, 12 mm shown) were placed in phantoms of varying size. The original 20x20 cm phantom (B, S) was placed in a 25x25 (C, M), 30x40 (D, L), and 40x50 cm (E, XL) extension rings.

Formfutura) to mimic a coronary artery with a concentric calcified plaque in background attenuation of water in axial images (**Fig. 1A**). The phantom consisted of two components: 1) the main body that consisted of a u-shaped constant diameter tube to allow filling of the phantom with iodinated contrast of varying concentrations, and 2) the lid that contained the iodinated contrast within the phantom. The main body was printed with different lumen diameters (4, 6, 8, 10, and 12 mm) to represent different vessel diameters.

B. Image acquisition

These phantoms were then scanned on a first-generation dual-source photon-counting CT (Naeotom Alpha, Siemens Healthineers) in standard resolution mode with a tube voltage of 120 kVp. Scans were performed with different phantom sizes, starting with the 20 cm diameter phantom (S, **Fig. 1B**). For the medium phantom size (M), the 20 cm diameter phantom was placed in a 3-D printed PLA extension ring of 25x25 cm (**Fig. 1C**) while the large phantom size (L) corresponded to the use of the elliptical 30x40 cm extension ring that accompanies the original 20 cm phantom (**Fig. 1D**). The phantom and the 30x40 cm extension ring were then placed in a 3-D printed PLA elliptical extension ring of 40x50 cm to constitute the extra-large sized phantom (XL, **Fig. 1E**) like a Matryoshka doll. For each phantom size, the diameter phantoms were filled with different concentrations of iodinated contrast (2, 5, 10 mg/mL) and scanned three times at volumetric CT dose index (CTDIvol) of 5, 10, 15, and 20 mGy.

C. Image analysis

For each scan, spectral results, specifically VMI 70 keV, iodine density maps, and VNC, were reconstructed with a field of view of 250 mm, a slice thickness of 3 mm, a reconstruction filter of Qr40, and a quantum iterative reconstruction (QIR) level of 0. Regions of interest (ROI) 60% of the lumen diameter (2.4, 3.6, 4.8, 6, 7.2 mm) were then placed automatically by first thresholding the images and then determining the center of the tube. The positions of these ROIs were then copied to other spectral results and repetitions. Corresponding mean and standard deviation of measured values were determined across each of 8 consecutive slices containing each diameter phantom in the three repetitions. Scatter plots of mean measured values for each spectral result were utilized to visualize the effect of vessel diameters and phantom size on quantification at 20 mGy. Error bars represented the standard deviation of the mean values across image slices. Additionally, average differences relative to values at a lumen diameter of 12 mm with the S phantom, which were least affected by phantom size and lumen diameter, were calculated to characterize deviation for each lumen diameter, spectral result, and phantom size combination.

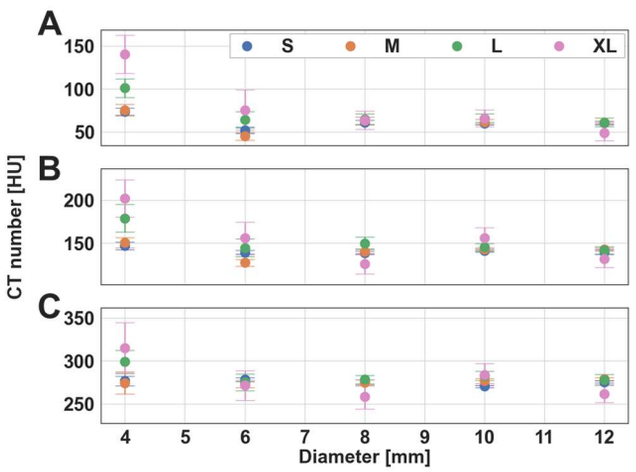


Fig. 2. CT number of VMI 70 keV for different lumen diameters and phantom sizes. VMI 70 keV exhibited consistent values for large lumen diameters at iodine concentrations of 2 (A), 5 (B), and 10 (C) mg/mL. Deviation in CT number at a lumen diameter of 4 mm increased with phantom size but decreased with iodine concentration. Error bars represented the standard deviation of values measured across slices of 3 different scans.

III. Results

VMI 70 keV demonstrated consistent quantification across lumen diameters greater than 6 mm in all sizes while smaller sized phantoms additionally performed well at a lumen diameter of 4 mm (**Fig. 2**). For S and M phantoms, average differences in VMI 70 keV relative to that in 12 mm at 20 mGy ranged from -14.7 to 15, -10.8 to 11.8, and -3.8 to 2.6 HU for 2, 5, and 10 mg/mL of iodine, respectively. Disparities in L and XL phantoms similarly decreased with increasing iodine concentration. In particular, at a lumen diameter of 4 mm, average differences were 53.9 and 85.4 HU at 2 mg/mL iodine and 35.1 and 31.4 HU at 10 mg/mL for L and XL phantoms, respectively. While VMI 70 keV quantification in L and XL phantoms highlighted significant deviation at a lumen diameter of 4 mm, values at lumen diameters greater than 6 mm were similar to those at S with differences averaging 5.5, 4.5, and -3.2 HU for 2, 5, and 10 mg/mL of iodine, respectively.

Iodine density, on the other hand, exhibited similar values

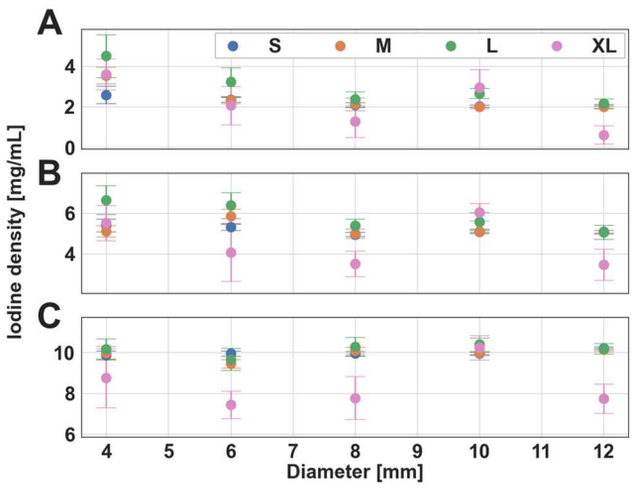


Fig. 3. Effect of lumen diameter, phantom size, and iodine concentration on iodine density maps. Measured iodine density at 2 (A), 5 (B), and 10 (C) mg/mL displayed similarity to iodine density with the S phantom and 12 mm lumen diameter for S, M, and L phantoms at large lumen diameters. Error bars represented the standard deviation of values measured across slices of 3 different scans.

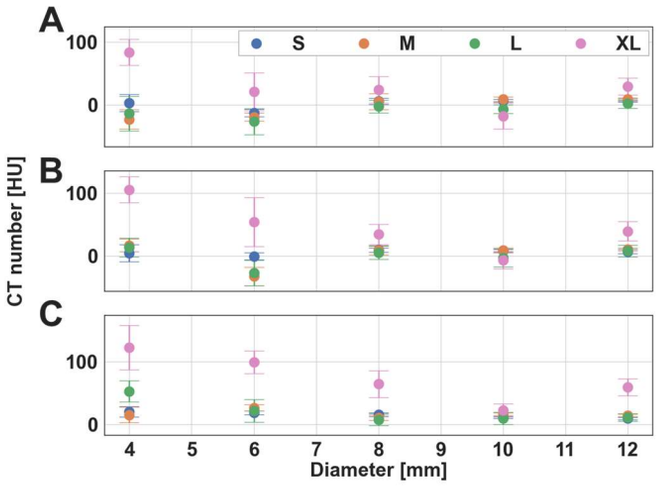


Fig. 4. VNC quantification at different target sizes for varying patient sizes and iodine concentrations. While VNCs highlighted their ability to remove attenuation of iodine almost completely, particularly at larger lumen diameters, deviation from the expected values increased with decreasing lumen diameter for all three iodine concentrations, 2 (A), 5 (B), and 10 (C) mg/mL. Error bars represented the standard deviation of values measured across slices of 3 different scans.

for all lumen diameters in smaller phantoms and with iodine concentrations greater than 2 mg/mL (**Fig. 3**). Iodine density fell within ± 0.7 mg/mL with S and M phantoms and iodine concentrations of 5 and 10 mg/mL, while at 2 mg/mL of iodine with a lumen diameter of 4 mm, average differences measured 0.8 and 1.6 mg/mL for S and M phantoms, respectively. Quantification specifically deteriorated at a lumen diameter of 4 mm for L phantom with average differences of 3.1 and 1.9 mg/mL for 2 and 5 mg/mL iodine, improving at 10 mg/mL with a difference of 0.2 mg/mL. While iodine density values with L phantom at lumen diameters greater than 8 mm were similar to those of phantom sizes S and M and lumen diameters greater than 6 mm, iodine density in the XL phantom differed with an average difference of -2.4 mg/mL across lumen diameters for 10 mg/mL of iodine.

VNC displayed deviation from expected values at large phantom sizes, particularly at small lumen diameters (**Fig. 4**). Because VNC remove attenuation from iodine from image, VNC values at lumen diameters greater than 8 mm, with the exception of with the XL phantom, fell within ± 15 HU of 5 HU. Larger magnitude aberrations were appreciated in other lumen diameters, especially 4 mm where average differences measured -14.8, -30.1, -17.4, and 85.4 HU at 2 mg/mL iodine for S, M, L, and XL, respectively. Conversely, at 10 mg/mL, VNC values at small lumen diameters remained consistent with those at large lumen diameters for S and M phantoms with average differences of 12 and 11.8 HU, respectively. Deviations at small lumen diameters with large phantoms increased with high concentrations to 62.7 and 145.5 HU for L and XL phantoms, respectively.

IV. Conclusion

Spectral results illustrated consistent quantification in lumen diameters greater than 6 mm across different phantom sizes for varying iodine concentration. At lumen diameters less than 6 mm, quantification increased in accuracy with decreasing phantom size. This finding presents a promising avenue for further research, aimed at advancing PCCT hardware and software. Such developments could enable highly sensitive iodine imaging at ultra-high resolutions.

The invariant values across lumen diameters boosts confidence in the use of spectral results for quantitative imaging of small structures, such as for cardiac imaging. The high resolution of PCCT illuminates small structures that may have pathological significance, and spectral imaging improves visualization of calcified structures in addition to enabling the use of quantitative metrics not affected by acquisition parameters, patient habitus, or lumen diameter. This combination offers synergistic diagnostic information for the evaluation of small vessels, including accurate estimation of the size of calcified plaques. Future studies will exploit both spatial resolution and spectral imaging to evaluate new cardiovascular applications, such as determining internal plaque structure and composition.

Acknowledgements

We acknowledge support from National Institutes of Health (R01EB030494) and Siemens Healthineers.

References

[1] G. A. Roth *et al.*, "Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study," *J Am Coll Cardiol*, vol. 76, no. 25, pp. 2982–3021, Dec. 2020, doi: 10.1016/J.JACC.2020.11.010.

[2] M. J. Budoff *et al.*, "Assessment of coronary artery disease by cardiac computed tomography: A scientific statement from the American Heart Association Committee on Cardiovascular Imaging and Intervention, Council on Cardiovascular Radiology and Intervention, and Committee on Cardiac Imaging, Council on Clinical Cardiology," *Circulation*, vol. 114, no. 16, pp. 1761–1791, Oct. 2006, doi: 10.1161/CIRCULATIONAHA.106.178458.

[3] M. J. Son *et al.*, "Current Role of Computed Tomography in the Evaluation of Acute Coronary Syndrome," *Diagnostics*, vol. 11, no. 2, Feb. 2021, doi: 10.3390/DIAGNOSTICS11020266.

[4] A. Knez *et al.*, "Usefulness of multislice spiral computed tomography angiography for determination of coronary artery stenoses," *Am J Cardiol*, vol. 88, no. 10, pp. 1191–1194, Nov. 2001, doi: 10.1016/S0002-9149(01)02060-4.

[5] K. Rajendran *et al.*, "Full field-of-view, high-resolution, photon-counting detector CT: technical assessment and initial patient experience," *Phys Med Biol*, vol. 66, no. 20, p. 205019, Oct. 2021, doi: 10.1088/1361-6560/AC155E.

[6] S. Van Hedent, N. Große Hokamp, R. Kessner, R. Gilkeson, P. R. Ros, and A. Gupta, "Effect of Virtual Monoenergetic Images From Spectral Detector Computed Tomography on Coronary Calcium Blooming," *J Comput Assist Tomogr*, vol. 42, no. 6, pp. 912–918, Nov. 2018, doi: 10.1097/RCT.0000000000000811.

[7] V. Sandfort, M. Persson, A. Pourmorteza, P. B. Noël, D. Fleischmann, and M. J. Willemink, "Spectral photon-counting CT in cardiovascular imaging," *Journal of Cardiovascular Computed Tomography*, vol. 15, no. 3. Elsevier Inc., pp. 218–225, May 2021. doi: 10.1016/j.jcct.2020.12.005.

[8] Q. Liu, Y. Wang, H. Qi, Y. Yu, and Y. Xing, "Exploring the best monochromatic energy level in dual energy spectral imaging for coronary stents after percutaneous coronary intervention," *Sci Rep*, vol. 11, no. 1, pp. 1–7, Sep. 2021, doi: 10.1038/s41598-021-97035-7.

[9] L. P. Liu *et al.*, "Consistency of spectral results in cardiac dual-source photon-counting CT," *Scientific Reports 2023 13:1*, vol. 13, no. 1, pp. 1–10, Sep. 2023, doi: 10.1038/s41598-023-41969-7.

[10] L. P. Liu *et al.*, "First-generation clinical dual-source photon-counting CT: ultra-low-dose quantitative spectral imaging," *Eur Radiol*, vol. 32, no. 12, pp. 8579–8587, Dec. 2022, doi: 10.1007/S00330-022-08933-X.

Photon-counting CT using a Conditional Diffusion Model for Deblurring and Texture-preservation

Christopher Wiedeman[1†], Chuang Niu[1†], Mengzhou Li[1], Bruno De Man[2], Jonathan S Maltz[3*], and Ge Wang[1*]

***Abstract*—Ultra-high resolution images are desirable in photon counting CT (PCCT), but resolution is physically limited by interactions such as charge sharing. Deep learning is a possible tool for image deblurring, but sourcing paired training data that adequately models the target task is difficult. Additionally, deblurring algorithms can distort noise texture, which is an important in many clinical diagnostic scenarios. Here, we train conditional denoising diffusion probabilistic models (DDPMs) for PCCT deblurring, with the objective to retain textural characteristics of local noise. PCCT simulation methods are used to synthesize realistic resolution degradation. To preserve noise texture, we explore decoupling the noise and signal image inputs and outputs via deep denoisers, explicitly mapping to each during the deblurring process. Our experimental results indicate that our DDPM trained on simulated data can improve sharpness in real PCCT images. Additionally, the disentanglement of noise from the original image allows our model more faithfully preserve noise texture.**

***Index Terms*—Photon-Counting CT, Deblurring, Super-Resolution, Diffusion Models, CatSim.**

I. Introduction

X-RAY CT is a preferred modality when high-resolution 3D tomography of internal anatomy is required. Improving CT resolution beyond its current capability is crucial for tasks such as coronary plaque and inner-ear imaging. Unfortunately, x-ray resolution is limited in energy-integrating detectors (EIDs), which convert x-rays into visible light for detection. Optical crosstalk is prevented with borders around each pixel, and reducing pixel size increases the total "dead space" occupied by these septa, reducing dose efficiency.

Photon-counting detectors (PCDs) promise numerous advantages over EIDs for CT imaging. PCDs convert x-ray photons directly into charge clouds, which are registered as discrete events rather than as aggregate energy. This discrete and more direct process improves quantum efficiency, minimizes electronic noise effects, and can produce spectral images using a single x-ray exposure. More importantly, pixels can be densely spaced without compromising dose efficiency, as there are no reflective septa. Photon-counting CT (PCCT) has shown improved resolution and image quality over EID systems [1].

Despite these advantages, PCCT is subject to other resolution-limiting considerations. First, the focal spot of the x-ray source can be a limiting factor, especially when imaging larger anatomies, where higher tube current density typically necessitates the use of larger focal spots. Secondly, as pixel density increases, the likelihood of charge sharing, in which the charge associated with a single incoming photon is spread across multiple neighboring pixels, also increases. Thirdly, quantum interactions within the detector material distort spectral and spatial information. For CdTe/CZT (high Z) detectors, this distortion is caused by K-escape fluorescence depositing secondary ≈30 keV photons at some distance from the point of primary interaction. For edge-on deep silicon detectors (low Z), Compton scatter becomes a concern. Improvement in hardware and fabrication cannot easily correct these drawbacks, as they occur at the atomic level. As such, researchers have turned to algorithmic methods for further resolution improvement, but directly reversing the complex processes listed above with physics models is computationally impractical. Thus, we investigate the deep learning (DL) algorithms – specifically diffusion generative models – for enhancing PCCT resolution in the image domain.

DL image super resolution (SR) and deblurring are closely related problems in computer vision, where the former includes upscaling the pixel count while also sharpening features. Architecturally, most deep SR models are convolutional neural networks, where significant improvements have come from optimizing the modular sequence (EDSR) and residual connections (RCAN) within networks [2], [3]. SR enhancers have also been investigated in a plug-and-play fashion, where a deep SR module is used a prior in place of a denoiser [4]. Generative adversarial networks (GANs) have provided a superior loss objective for producing perceptually realistic high-fidelity images [5]. PULSE has produced better SR results by learning a high-resolution data manifold and searching for images that properly downscale to the input [6].

More recently, denoising diffusion probabilistic models (DDPMs) have demonstrated and largely defined state-of-the-art performance in a multitude of generative and inverse problems [7]–[9]. DDPMs are models trained to gradually reverse a Wiener process performed on target samples over many timesteps, eventually learning to recover new samples from pure Gaussian noise. Compared with GANs, DDPMs produce better diversity as they do not suffer from mode collapse, and have achieved superior image super resolution [10], [11]. However, model convergence and lengthy inference times due to the iterative denoising process restrict the use of DDPMs in high-dimensional problems.

Although desirable, translating deblurring and/or SR to the medical image domain is complicated by several obstacles. 3D medical images can have millions of voxels, and scaling both

[1]: Department of Biomedical Engineering, Center for Biotechnology & Interdisciplinary Studies, Rensselaer Polytechnic Institute, Troy, NY USA
[2]: Global Research Center, GE HealthCare, Niskayuna, NY
[3]: Molecular Imaging and Computed Tomography, GE HealthCare, Waukesha, WI, USA
†: Equal Contribution, *: Corresponding Authors

the training data and model is computationally challenging, particularly for the aforementioned DDPM. Furthermore, these DL methods require either paired or unpaired low resolution (LR) and high resolution (HR) data. To produce LR datasets, downsampling and linear shift-invariant blurring kernels are often applied to degrade existing images. A model learns to recover the original images from LR-HR pairs [12]–[14]. However, models trained with this approach frequently fail to perform satisfactorily on real images, most likely because the degradation operator inadequately models the true mapping between low and high resolution. This is especially true for PCCT, where the complex interactions that influence resolution and noise cannot be easily captured with simple operators [15]. Finally, it is crucial to consider and preserve noise texture while enhancing features in CT, as this noise is a key feature for trained readers. As such, for clinical use, SR models must be robust to noise level and avoid suppressing or distorting noise. For example, denoising may suppress diagnostically-relevant noise textures, such as those that identify interstitial lung disease.

The following work investigates the use of DDPMs for PCCT SR, and is a continuation of [16]. To synthesize paired data, CatSim is used to generate realistic, low-resolution counterparts to real PCCT head and chest scans, accounting for pixel cross-talk and noise. The degradation is modulated by altering the size of the digital phantom, focal spot and detector pitch. During testing, the trained model is applied to the original PCCT images to observe whether the learned features transfer to the target image domain. Interestingly we found that training a conditional DDPM for SR leads to oversmoothing of CT noise in homogeneous regions. As such, we explore two new input/output schemes, where deep denoisers are first applied to disentangle noise from the signal. Feeding noise and signal separately is used to emphasize proper noise mapping during SR.

II. Methods

A. Data and Simulation

Twenty chest and seven head PCCT scans from GE HealthCare's Deep Silicon prototype scanner were obtained as digital phantoms. The CatSim PCCT module was used to simulate two instances of noised, low resolution (LR) and noised, high resolution (HR) counterparts to the phantoms [17], [18]. The two LR noise instances are used to train a LR denoiser using Noise2Sim [19]. The noised HR images paired with the original phantoms are used to train a supervised HR denoiser. One chest and head scan and their corresponding simulated images are withheld from training for validation and testing.

All scans are simulated with 120 kVp tube potential and 1 s rotation period, with geometry and hardware specific to the GE system system. Exact parameters for focal spot, pixel size, and detector pitch are not disclosed for proprietary reasons. Noisy HR images are simulated with a 350 mA source current, as well as a 50%-reduced focal spot size, increased number of views, and no pixel cross-talk, to ensure noise addition without resolution loss. LR images are simulated with cross-talk modeled, and a 220 mA tube current. As simulating with system-specific parameters results in minimal resolution loss, phantoms are digitally shrunk by 50% in all directions, effectively increasing the degradation during scanning. The assumption is that learning to resolve shrunken features under a realistic forward model will transfer to resolving smaller features in real images. All volumes are reconstructed using filtered back projection, with voxel sizes and locations identical to the original phantoms.

B. Conditional DDPM and Noise-Preservation

1) Conditional DDPM: Following [16], we formulate PCCT SR as a conditional generation problem. Given a Gaussian noise image $\mathbf{x}$ and a conditional LR PCCT image $\mathbf{n}$, we aim to predict the HR PCCT image $\mathbf{y}$ through learning the conditional distribution $p(\mathbf{y}|\,\mathbf{x})$. Here $p(\mathbf{y}|\,\mathbf{x})$ is approximated with a reverse Markovian diffusion process, where each iteration step is parameterized with the neural network f_θ. For optimizing f_θ, DDPM curated a forward Markovian process with fixed steps. Specifically, the forward process gradually adds Gaussian noise to an HR image, resulting in an image series $\mathbf{y}_0 \rightarrow \mathbf{y}_1 \rightarrow \cdots \rightarrow \mathbf{y}_T$, where the noise level gradually increases with time step t, $\mathbf{y}_0$ and $\mathbf{y}_T$ are the HR and pure Gaussian noise image respectively, and T is the predefined number of iteration steps. The training loss can be derived as:

$$L = \mathbb{E}_{\mathbf{x},\mathbf{y}}\,\mathbb{E}_{\epsilon,\gamma}\,||f_\theta(\mathbf{x}, \sqrt{\gamma}\,\mathbf{y}_0 + \sqrt{1-\gamma}\,\epsilon, \gamma) - \epsilon||_l^l, \quad (1)$$

where the loss function is to predict the Gaussian noise added in $\mathbf{y}_t$, conditioned on the LR image and the noise level γ, we set $l = 1$, $(\mathbf{x}, \mathbf{y})$ is a paired training sample, $\gamma \sim p(\gamma)$ is defined as in [11], and ϵ is normally-distributed noise. The inference of reverse Markovian diffusion for generating HR PCCT image is:

$$\mathbf{y}_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}}\left[\mathbf{y}_t - \frac{1-\alpha_t}{\sqrt{1-\gamma_t}} f_\theta(\mathbf{x}, \mathbf{y}_t, \gamma_t)\right] + \sqrt{1-\alpha_t}\,\epsilon_t. \quad (2)$$

Thus, the HR image ($t = 0$) can be iteratively generated from a pure Gaussian noise image ($t = T$) conditioned on an LR image.

2) Noise-Preservation Schemes: According to the Noise2Sim theorem [19], neural networks cannot be trained to predict random noise from independent noisy inputs such that noise is suppressed while signals are preserved. In practice, we find that conditional DDPM also suppresses some noise components in HR images when conditioned on LR images with independent noise, resulting in inconsistent noise patterns (Figure 5). To address this problem, we propose disentangling the noise components from the noisy LR input and thus enable the DDPM to generate an HR image more conditioned on LR contents. We explore two alternative schemes for the conditional DDPM: (1) *DDPM-noise-split*: stack each denoised LR image and the disentangle noise as a two-channel conditional input; (2) *DDPM-denoise*: use the denoised LR image only as the conditional input. Both are compared to *DDPM*, which is fed the regular LR input. In this study, we aim to not only develop SR models with simulation data but also apply SR models to real PCCT data. Therefore, we use the Noise2Sim method to denoise both simulated LR

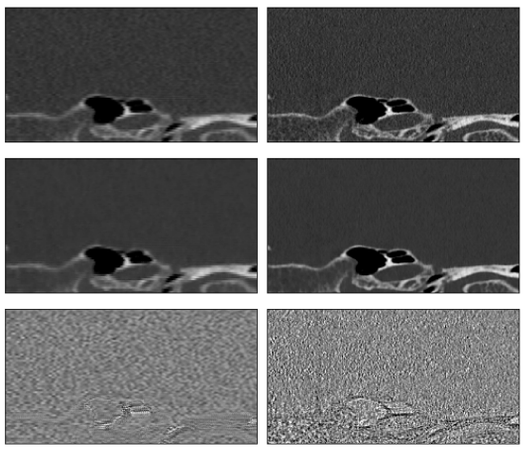


Fig. 1. Denoising example results for a simulated low-resolution (left column) and true PCCT (right column) patch. Top row: original image, [-500, 2000] HU. Middle: Denoised image [-500, 2000] HU. Bottom: Noise map, [-200, 200] HU.

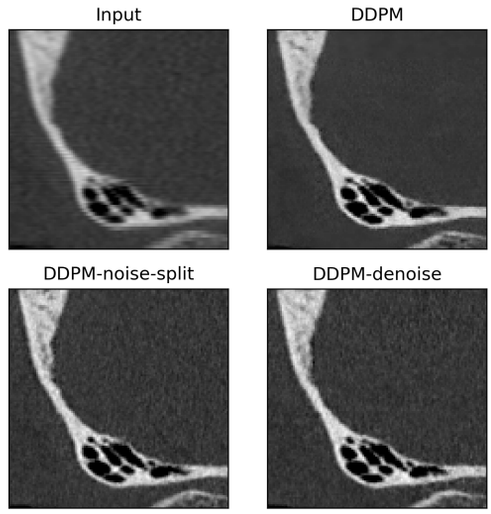


Fig. 2. Qualitative SR output comparison for simulated validation image. Input is the simulated LR image. [-500, 2000] HU.

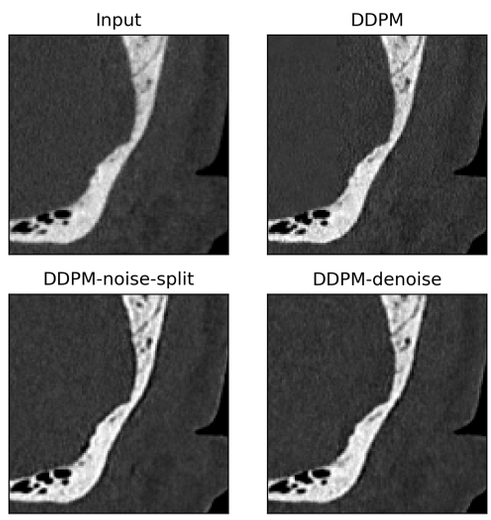


Fig. 3. Qualitative SR output comparison for real PCCT test image. Input is a real PCCT image. [-500, 2000] HU.

and real PCCT images for the proposed schemes. Specifically, we use two simulated LR images with independent noise as inputs and targets for training an LR deep denoiser. To train the real PCCT denoiser, we simulate HR images with higher noise as input, to predict the real PCCT image. The denoising results shown in Figure 1 demonstrate the effectiveness of the trained deep denoisers.

3) Implementation Details: In our implementation, 128 $\times$ 128 patches are randomly cropped for training the networks. We use the same network architecture as in [16], and the attention is applied to the layer with the smallest spatial dimension. The batch size is 4 and the Adam optimizer is used with a learning rate of 10^{-4}. The number of sampling steps for DDPM is set to 2,000, the number of training iterations to 500,000, and all other hyper-parameters are set equal to those employed in [16]. We use the same denoising network architecture (two-channel UNet) as in [19]. We train both denoising networks for 10,000 iterations using the Adam optimizer with a batch size of 8 and a learning rate of 10^{-5}. Our implementation is based on PyTorch. While it is well-known that automatic mixed precision improves training speed, we find that this sometimes results in unstable training in our experiments; this technique is not used in this study.

III. Results and Discussions

We examine temporal bone images in the coronal plane. Figure 1 illustrates an example of both the LR and HR deep denoisers. Overall, it appears that both denoisers successfully separate noise from signal. Importantly, although the HR denoiser was trained to map from noisier simulations to original images, the model successfully denoises the original PCCT images. Both denoisers still inevitably remove some high frequency signal, which can be seen from the residual bone structures in the noise image. However, this fraction of information is retained during the SR enhancement for *DDPM-noise-split*, as it is explicitly fed both signal and noise channels.

Figures 2 and 3 compare the SR results for all models on temporal bone regions in the simulated validation and test scans, respectively. We can see that all DDPMs substantially sharpen bone features relative to the simulated LR image. The learned feature sharpening also appears to transfer to the real PCCT scans, although this enhancement is more modest. In both validation and test cases, *DDPM* undesirably suppresses noise texture, especially in the soft tissue regions. In contrast, *DDPM-noise-split* and *DDPM-denoise* better preserve this texture, indicating that these schemes can more effectively map to the desired noise.

Quantitatively, we apply each model to an image slice a GE-PCCT scan of CatPhan714 and calculate the modulation transfer function (MTF), which is plotted in Figure 4. All three DDPMs enhance the MTF, although *DDPM-noise-split* and *DDPM-denoise* interestingly achieve scores ¡ 1 at some

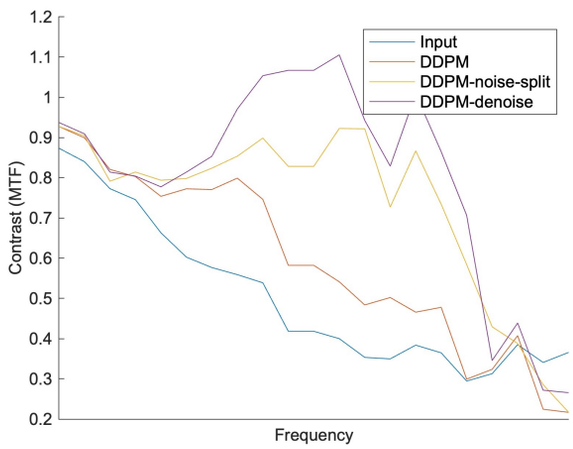


Fig. 4. MTF for the input compared with, SR models calculated from an actual PCCT scan of CatPhan714. Frequency units are redacted for proprietary reasons.

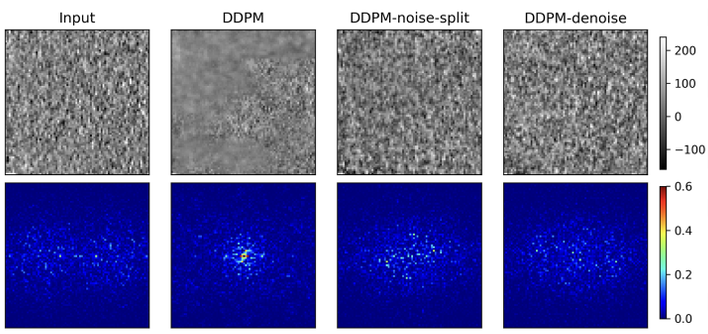


Fig. 5. Top row: Example patch from homogenous soft tissue region in real PCCT image (HU). Bottom row, power spectral density of patch noise (% total power).

frequencies, which can occur when the model boosts contrast above the desired target. It also should be noted that the image slice was in the axial plane while the models have only been trained on coronal plane images. This domain mismatch will be corrected in the future as we train SR models in other planes.

As noted previously, *DDPM* produces oversmooth noise texture. We further analyze this by isolating a patch of homogeneous soft tissue and computing the power spectral density (PSD) of the noise, as seen in Figure 5. One can see that the noise from *DDPM* has an unrealistic concentration of low-frequency components. On the other hand, *DDPM-noise-split* and *DDPM-denoise* have distributed noise characteristics more similar to the input, indicating once again that these schemes are better able to preserve noise texture while still enhancing feature sharpness. Still, these noise spectra are not perfectly representative of the input, and we aim to future improve noise spectral fidelity.

IV. Conclusion

In this work, we demonstrate how DDPMs can be combined with simulation methods for PCCT deblurring while better preserving noise texture. Subsequent work will train axial and sagittal plane SR and denoising models, unifying them in the 3D framework presented in [16]. Furthermore, we will improve the noise preservation and transfer learning to the target PCCT data, as well as evaluate the approach on both head and chest images with reader studies.

Acknowledgment

The research reported in this publication was supported by GE HealthCare and First Imaging as well as the National Institutes of Health (NIH) under Award Numbers R01EB026646, R01CA233888, R01CA237267, R01HL151561, R42GM142394, R21CA264772, R01EB031102, R01EB032716. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

[1] 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.

[2] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *CVPR*, 2017, pp. 136–144.

[3] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *ECCV*, 2018, pp. 286–301.

[4] K. Zhang, W. Zuo, and L. Zhang, "Deep plug-and-play super-resolution for arbitrary blur kernels," in *CVPR*, 2019, pp. 1671–1681.

[5] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *CVPR*, 2017, pp. 4681–4690.

[6] S. Menon, A. Damian, S. Hu, N. Ravi, and C. Rudin, "Pulse: Self-supervised photo upsampling via latent space exploration of generative models," in *CVPR*, 2020, pp. 2437–2445.

[7] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *NeurIPS*, vol. 33, pp. 6840–6851, 2020.

[8] Y. Song, L. Shen, L. Xing, and S. Ermon, "Solving inverse problems in medical imaging with score-based generative models," in *ICLR*, 2022.

[9] W. Xia, W. Cong, and G. Wang, "Patch-based denoising diffusion probabilistic model for sparse-view CT reconstruction," *arXiv preprint arXiv:2211.10388*, 2022.

[10] P. Dhariwal and A. Nichol, "Diffusion models beat GANs on image synthesis," 2021.

[11] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE TPAMI*, 2022.

[12] H. Yu, D. Liu, H. Shi, H. Yu, Z. Wang, X. Wang, B. Cross, M. Bramler, and T. S. Huang, "Computed tomography super-resolution using convolutional neural networks," in *ICIP*, 2017, pp. 3944–3948.

[13] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, S. Ju, Z. Zhao, Z. Zhang, W. Cong *et al.*, "CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE)," *IEEE TMI*, vol. 39, no. 1, pp. 188–203, 2019.

[14] X. Jiang, Y. Xu, P. Wei, and Z. Zhou, "CT image super resolution based on improved srgan," in *ICCCS*. IEEE, 2020, pp. 363–367.

[15] M. Li, D. S. Rundle, and G. Wang, "X-ray photon-counting data correction through deep learning," *arXiv preprint arXiv:2007.03119*, 2020.

[16] C. Niu, C. Wiedeman, M. Li, J. Maltz, and G. Wang, "3D photon counting CT image super-resolution using conditional diffusion model," in *17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, Stony Brook, NY, USA, 2023.

[17] B. D. Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. McOlash, P. Sainath, C. Shaughnessy, B. Tower, and E. Williams, "CatSim: a new computer assisted tomography simulation environment," in *Medical Imaging 2007: Physics of Medical Imaging*, vol. 6510, 2007, p. 65102G.

[18] M. Wu, P. FitzGerald, J. Zhang, W. P. Segars, H. Yu, Y. Xu, and B. D. Man, "XCIST—an open access x-ray/CT simulation toolkit," *Physics in Medicine & Biology*, vol. 67, no. 19, p. 194002, 2022.

[19] 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, 2023.

Novel medical imaging diagnostics: Unleashing Artificial Intelligence Potential through Sinogram-Driven Models

Bingxi He, Yongbo Wang, Wei Mu, Shuo Wang, Mengjie Fang, Di Dong, Jianhua Ma, and Jie Tian, *Fellow, IEEE*

***Abstract*—Medical imaging combined with deep neural networks has emerged as a focal point for aiding clinical diagnoses, therapeutic evaluations, and prognostic forecasts. Nevertheless, prevailing methodologies predominantly rely on reconstructed medical images, neglecting the untapped potential of lossless raw data. Given the nature of AI to treat images as data for analysis, we propose an innovative starting point to focus on the raw data of computed tomography (CT) images. Introducing the Sinogram-Driven Diagnosis Model (SDDM), we transcend image reconstruction, enabling nodule classification with unprocessed signal. SDDM integrates attention mechanisms across voxels, spatial, and angle dimensions, heightening knowledge expression within the nodule spatial domain. A pyramid-structured 3D convolutional neural network (3D-CNN) facilitates feature extraction. Employing multi-task learning to predict nodule attributes, and discern benign from malignant cases. Notably, our study emulates real data acquisition processes by juxtaposing simulated raw data from the LIDC-IDRI dataset with secondary reconstructed CT images. Experiments demonstrate impressing performance of our SDDM on identifying nodule, which proves the feasibility of diagnosis by CT raw data. We believe that the proposed method is promising to change the current medical diagnosis pipeline since it has the potential to free the radiologists.**

***Index Terms*—Artificial Intelligence, CT scan, Diagnosis, Lung, Sinogram.**

I. Introduction

IN the wake of significant advancements in deep neural network (DNN) technologies, the domain of Computer-Aided Diagnosis (CAD) systems employing DNNs has witnessed a notable proliferation in recent times [1]. These CAD systems, tailored to address various clinical problems, exhibit the capacity to quantitatively extract radiomic signatures from medical images. It is noted that most present-day CAD systems are constructed through the reconstructed images, meticulously tailored to suit human visual cognition rather than intelligent computing [2]. Fundamentally, the reconstruction procedure introduces a variant of lossy compression throughout the course from image acquisition to diagnostic interpretation [3]. Currently, prevailing research adheres to the analysis paradigm of "signal-image-knowledge." However, some scholars have proffered an intriguing notion: Artificial intelligence (AI) technology can now uncover high-dimensional features from medical images that are imperceptible to the human eye. So why not use this technology to directly analyze the raw signals with more comprehensive information [3], [4] ?

Hence, with the aim of harnessing the full potential of AI to establish a seamless connection between signal and knowledge, this paper propose a novel sinogram-driven diagnostic model (SDDM). The methodology involves simulating and reconstructing existing computed tomography (CT) data to replicate authentic data acquisition procedures, which comes from [5]. Initially, the process involves the reshape of original data through the introduction of focal location and spatial coordinate system transformation. This step aims to minimize the impact of extraneous data within non-focal regions, thereby enhancing the efficiency of subsequent analysis. Subsequently, the integration of voxel attention, spatial attention, and angle attention mechanisms, each represents within, between and different angles of rays, leads to the creation of a comprehensive spatial feature volume. The synthesized spatial feature volume utilized a model structure similar to Unet [6] for high-dimensional feature aggregation and employed a pyramid-structured 3D convolutional neural network (3D-CNN) architecture [7] for multi-scale feature extraction and the prediction of lung nodule diagnosis.

II. Methodology

A. Datasets

The lung image database consortium (LIDC) and image database resource initiative (IDRI) dataset [8] which we used in this study contains in excess of 1,000 scans, each delineation is complemented by intricate annotations from an array of radiologists, delineating both the spatial coordinates and intrinsic attributes of the lung nodules. In addition, we excluded nodules with a diameter of 3 millimeters or less, as well as nodules with the Median Malignancy Level (MML) of 3 [9]. This refinement yielded 1417 nodules, bifurcated into 978 benign and 439 malignant categories, as detailed in Table 1. Each nodule is characterized by eight discerning attributes: subtlety, internal structure, calcification, sphericity, margin, lobulation, spiculation, and radiographic solidity. CT scans were resampled to obtain isotropic resolutions of 1.0×1.0×1.0 mm^3 using spline interpolation. All gathered data were systematically partitioned into training, validation, and testing cohorts, maintaining a ratio of 70%, 15%, and 15% at the nodule level.

B. He, W. Mu, S. Wang, M. Fang and J. Tian are with the Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Engineering Medicine, Beihang University, Beijing, China. And J. Tian is the corresponding author (e-mail: jie.tian@ia.ac.cn).

D. Dong is with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing, China.

Y.Wang and J. Ma are with the School of Biomedical Engineering, Southern Medical University, Guangzhou, China.

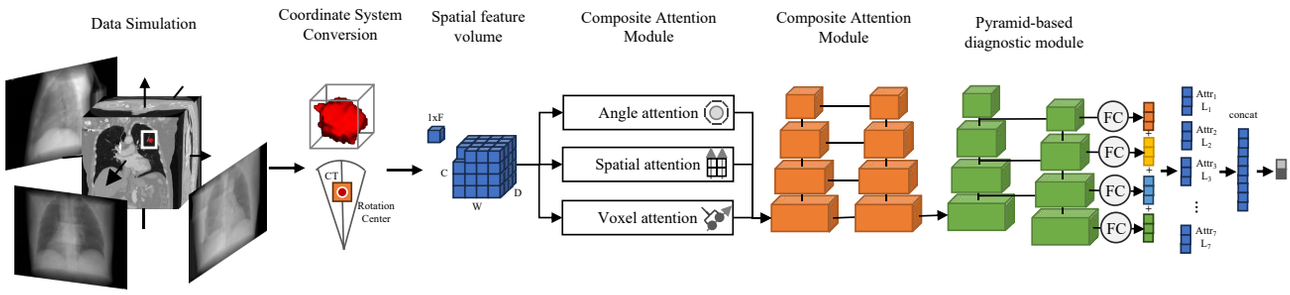


Fig. 1: An overview of this study. It includes two parts: data preprocessing and SDDM, in which SDDM contains three parts: Composite Attention Module, Feature amalgamation module and Pyramid-based diagnostic module.

B. Sinogram data simulation

The cone-beam CT raw data are simulated using Siddon's algorithm [10] which can be used to obtain the intersection length of a thin ray with a voxel. In our experiment, the following geometric parameters are used. The dimension of the arc detector array was assumed to be 888×512. The size of the detector unit is 1.0239×1.0988 mm^2; the source-to-isocenter distance and the source-to-detector distance are 538.52 mm and 946.74 mm, respectively; the number of projections per complete rotation is 360-view. The maximum cone angle was set 33.031°. And the corresponding image is reconstructed by the FDK algorithm with a ramp kernel. It's worth mentioning that we simulated an axial scan protocol. Our main focus is on whether the effective attenuation information in the projection data can be utilized for subsequent AI learning. Therefore, we do not consider physical effects such as scatter and beam hardening.

C. Sinogram data Preprocesing

1) Target nodule area mapping: Initially, we proceeded by extracting the coordinates of the nodule mask c_v within the CT image, subsequently determining the coordinates of its central point $\overline{c_v}$. Concurrently, the dimensions of the nodules s_v in the CT image were duly preserved. Subsequent to this, a pivotal step involved mapping the CT image's coordinate system with the scanning coordinate system, effectively establishing a correspondence where the identified central point of the CT image c_m concurrently served as the pivotal rotation center for our simulation. This strategic alignment allowed us to calculate the spatial disparity $c_d = c_v - c_m$ between the positions of c_v and the c_m. This disparity measurement as a pivotal tool in delineating the spatial orientation of nodules concerning the rotation center, consequently enabling the derivation of relative positional disparities among distinct nodular entitie–a determinant factor for the subsequent extraction and recombination of sinogram data.

2) Extraction and recombination of sinogram: The preceding procedural steps culminate in the determination of the nodules' spatial positioning c_d and dimensions s_v. Consequently, our attention becomes singularly focused on the intrinsic attributes within these nodules during the ensuing extraction and reassembly stages. In this study, a methodical uniform sampling strategy is employed to encompass the nodules' spatial extents, effectively yielding a comprehensive spatial voxel ensemble $\mathbf{q} = [q_1, q_2, q_3, ..., q_{C\times D\times W}] \in \mathbb{R}^{C\times D\times W}$. Notably, the periodic nature (F) of rotational sampling mandates each voxel's $q \in \mathbb{R}^{1\times F}$ integration of projectional information emanating from diverse angles, all centered around the respective voxel's location. Given the relative coordinate attributed to each voxel within the scanning coordinate system, the application of elementary trigonometric relationships facilitates the computation of the point's position across disparate angular projections. In summation, this phase of our methodology expertly refines the voluminous sinogram data, effecting its transformation into an optimized spatial voxel matrix through the delineation of a single voxel's attributes. Subsequently, this honed spatial voxel matrix $Q = [\mathbf{q}_1, \mathbf{q}_1, ..., \mathbf{q}_F] \in \mathbb{R}^{F\times C\times D\times W}$ is harnessed as the input for the SDDM. In this study, C, W, and D represent channels, width, and depth, respectively. They reflect the length, width, and height of spatial sampling for tumor location, all empirically set to 64. Meanwhile, F was set to 360 in this experiment.

D. Overview of SDDM

The structure of SDDM is shown in Fig.1.

1) Composite Attention Module: The module integrates voxel, spatial, and angle attention mechanisms, each optimizing aspects ranging from micro to macro scales, involving distinct levels: voxel-wise within a single ray, ray-wise within a single angle, and angle-wise across different angles. Voxel attention entails initial processing of the spatial voxel matrix, achieved by rotating spatial features across the F dimension to better align with the incoming ray path. Given the tumor's typically modest size, we simplify by neglecting variations in ray incidence angles in the W − D plane. In addition, the C-dimension is treated as an eigenvector of the W − D planar voxel $\mathbf{q}$, giving it information about the high-angle incident deviation. Employing a multi-head attention mechanism

TABLE I
PERFORMANCE COMPARISON

Data sources	Methods	Train cohort				Test cohort			
		ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
CT	MV-KBC	0.896	**0.952**	0.896	0.896	0.868	0.934	**0.848**	0.877
CT	MTMR-Net	0.860	0.948	**0.906**	0.839	0.877	**0.942**	0.803	0.911
CT	MSCS	0.884	0.924	0.814	0.915	0.863	0.924	0.818	0.884
Reconstructed CT	MV-KBC	0.881	0.934	0.831	0.904	0.882	0.913	0.803	0.918
Reconstructed CT	MTMR-Net	0.873	0.907	0.726	0.939	0.858	0.905	0.742	0.911
Reconstructed CT	MSCS	0.833	0.905	0.831	0.834	0.840	0.903	0.697	0.904
Raw data	SSDM	**0.914**	0.929	0.821	**0.956**	**0.887**	0.922	0.758	**0.945**

TABLE II
Ablation experiments results.

Methods	Train cohort				Test cohort			
	ACC	AUC	SEN	SPE	ACC	AUC	SEN	SPE
FAM + PBDM	0.776	0.837	0.775	0.777	0.765	0.848	0.742	0.776
FAM + PBDM + AA	0.790	0.838	0.772	0.799	0.812	0.870	0.833	0.803
FAM + PBDM + AA + SA	0.854	0.883	0.814	0.872	0.845	0.896	0.773	0.878
FAM + PBDM + AA + SA + VA	0.864	0.900	0.801	0.892	0.854	0.910	0.758	0.898
FAM + PBDM + AA + SA + VA + AL	0.914	0.929	0.821	0.956	0.887	0.918	0.803	0.925

facilitates the calculation of multiple voxel points within a single ray. In spatial attention, mean and maximum attention scores are computed for rays within the same rotation angle. For angular attention, we temporarily treat dimension F as a channel and calculate its attention using SEnet's approach [11].

2) Feature amalgamation module: Despite the amalgamation of the three types of attention, further optimization of the complete spatial voxel matrix is imperative. Drawing inspiration from the cost volume concept presented in MVSnet [12], we employ a streamlined approach using a basic 3D U-Net architecture. This architecture efficiently leverages encoding and decoding mechanisms to aggregate contextual information within a broad receptive field, all achieved at a minimal computational cost. The difference, however, is that since we are not predicting depth, we do not use subsequent Softmax normalization operations.

3) Pyramid-based diagnostic module: Employing Densenet as the backbone network, the module seamlessly incorporates a feature pyramid structure into this framework for the extraction of multi-scale features. Subsequently, these features are subjected to a fully connected layer to achieve aggregation across varying scales, followed by their fusion through summation operators. Through an additional layer of multilayer perceptron (MLP), the fused features are concatenate with previously derived cd and sv attributes. Multiple independent MLPs are employed to establish a classifier with multiple attributes. Ultimately, the independent features used to attribute prediction are merged, culminating in the assessment of nodule malignancy.

E. Pulmonary nodule diagnosis task

The early diagnosis of lung cancer mainly depends on the accurate identification of malignant pulmonary nodules [13]. Accurate classification of benign and malignant pulmonary nodules can effectively improve the cure rate of lung cancer. In present clinical practice, the use of CT to diagnose benign and malignant pulmonary nodules is the most widespread [14]. Although radiologists can assess the extent of malignancy based on the appearance of nodules on images, it is highly influenced by subjectivity and experience, and there may be a certain rate of misdiagnosis. Therefore, many studies use CNN to analyze the potential information of pulmonary nodules from CT and use this information to classify them as benign or malignant [15]. Existing studies improve the accuracy of automatic pulmonary nodule analysis from multi-view (MV-KBC) [16], multi-scale (MSCS) [17], and multiple margin information (MTMR-Net) [18] of pulmonary nodules. In this study, we use raw CT data to diagnose benign and malignant nodules and compare them with CT-based prediction models.

III. Results

A. Comparison

We compare the proposed method with three classification models based on CT data: (1) MV-KBC, (2) MTMR-Net and (3) MSCS. The specific comparison results are shown in Table 1. The CT-based classification model has the highest area under the receiver operating characteristic (AUC) and sensitivity. However, SSDM achieves the highest accuracy and specificity in both train and test cohorts, even in comparison to CT-based models. The AUC of the model built based on the reconstructed CT data is slightly lower than that of the CT data model. The AUC of SSDM on the LIDC dataset reaches 0.922. The AUC of our proposed SSDM is 0.9$\sim$1.9 percent points

higher than the AUC of the model based on the reconstructed CT data, and it is close to the model's performance based on the original CT data. In addition, SSDM achieves the highest accuracy and specificity, even in comparison to raw CT models.

B. Ablation study

In this study, we introduced our raw data model incorporating five distinct strategies: FAM+PBDM (Feature amalgamation module and Feature amalgamation module), AA (Angle attention), SA (Sptial attention), VA (Voxel attention), and AL (Attribution learning). To systematically elucidate the contribution of each strategy to the model's performance, we embarked on a series of ablation experiments. We commenced with a model solely informed by strategy FAM+PBDM, subsequently integrating strategies AA through AL in a progressive fashion. By sequentially appending each strategy and assessing the consequent alterations in performance, we aimed to dissect the individual and combinatorial significance of these strategies. These five ablation trials offered a comprehensive perspective, enabling a rigorous assessment of the unique value brought by each strategy within our framework. Detailed results of the are shown in Table 2.

C. Discussion and Conclusion

We present SDDM, a novel model capable of direct diagnostic inference via the raw data of CT (Sinogram), thereby establishing a direct signal-to-knowledge mapping. This model incorporates a novel hybrid attention mechanism leveraging X-ray properties for nuanced feature fusion across scales. Our experiments highlight SDDM's superiority over models based on secondary reconstructed CT, while also demonstrating comparable performance to models developed from original CT. This underscores the significant potential of unprocessed data-centric modeling, enhancing the prospects for more precise future medical interventions. This study centers on the utilization of LIDC-IDRI datasets, and we anticipate forthcoming investigations involving more extensive datasets to substantiate the signal-to-knowledge hypothesis. Concurrently, our current dataset remains within the simulation-reconstruction phase. Simultaneously, we are diligently procuring genuine clinical raw data, despite the inherent challenges. The exigency lies in developing a direct diagnostic model tailored for authentic raw data, essential for advancing our research objectives. SDDM is currently constructed within the axial scan mode of CT, and our future endeavors will encompass the exploration of more intricate spiral scanning modes. Furthermore, it is imperative to develop lesion detection models utilizing raw data, as this will foster the eventual integration of raw data analysis technology into prevalent medical imaging equipment, underscoring the potential for transformative advancements.

REFERENCES

[1] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.

[2] D. Wu, K. Kim, B. Dong, G. E. Fakhri, and Q. Li, "End-to-end lung nodule detection in computed tomography," in *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9*. Springer, 2018, pp. 37–45.

[3] C. Chung, J. Kalpathy-Cramer, M. V. Knopp, and D. A. Jaffray, "In the era of deep learning, why reconstruct an image at all?" *Journal of the American College of Radiology*, vol. 18, no. 1, pp. 170–173, 2021.

[4] G. Wang, J. C. Ye, and B. De Man, "Deep learning for tomographic image reconstruction," *Nature Machine Intelligence*, vol. 2, no. 12, pp. 737–748, 2020.

[5] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, 2018.

[6] 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: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[7] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.

[8] S. G. Armato III, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman *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*, vol. 38, no. 2, pp. 915–931, 2011.

[9] W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," in *Information Processing in Medical Imaging: 24th International Conference, IPMI 2015, Sabhal Mor Ostaig, Isle of Skye, UK, June 28-July 3, 2015, Proceedings 24*. Springer, 2015, pp. 588–599.

[10] R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional ct array," *Medical physics*, vol. 12, no. 2, pp. 252–255, 1985.

[11] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[12] Y. Yao, Z. Luo, S. Li, T. Fang, and L. Quan, "Mvsnet: Depth inference for unstructured multi-view stereo," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 767–783.

[13] A. Del Ciello, P. Franchi, A. Contegiacomo, G. Cicchetti, L. Bonomo, and A. R. Larici, "Missed lung cancer: when, where, and why?" *Diagnostic and interventional radiology*, vol. 23, no. 2, p. 118, 2017.

[14] D. J. Brenner and E. J. Hall, "Computed tomography—an increasing source of radiation exposure," *New England journal of medicine*, vol. 357, no. 22, pp. 2277–2284, 2007.

[15] R. Golan, C. Jacob, and J. Denzinger, "Lung nodule detection in ct images using deep convolutional neural networks," in *2016 international joint conference on neural networks (IJCNN)*. IEEE, 2016, pp. 243–250.

[16] Y. Xie, Y. Xia, J. Zhang, Y. Song, D. Feng, M. Fulham, and W. Cai, "Knowledge-based collaborative deep learning for benign-malignant lung nodule classification on chest ct," *IEEE transactions on medical imaging*, vol. 38, no. 4, pp. 991–1004, 2018.

[17] X. Xu, C. Wang, J. Guo, Y. Gan, J. Wang, H. Bai, L. Zhang, W. Li, and Z. Yi, "Mscs-deepln: Evaluating lung nodule malignancy using multi-scale cost-sensitive neural networks," *Medical Image Analysis*, vol. 65, p. 101772, 2020.

[18] L. Liu, Q. Dou, H. Chen, J. Qin, and P.-A. Heng, "Multi-task deep model with margin ranking loss for lung nodule analysis," *IEEE transactions on medical imaging*, vol. 39, no. 3, pp. 718–728, 2019.

An overview of the open-source CT simulation environment XCIST

Mingye Wu, Jiayong Zhang, Paul Fitzgerald, James Jobin, Nandan Reddy, Anupama Debnath, Eri Haneda, Ying Fan, Jed Pack, Peter Lorraine, Michael Mahony, and Bruno De Man

***Abstract* — CT simulations often offer a valuable alternative to constructing new CT system prototypes or to acquiring real-world patient or phantom measurements. Supported by a recently completed NIH NCI ITCR project XCIST, we have open-sourced the CT simulation package CatSim, which has been developed in the context of GE's CT research over the last two decades. In this work, we summarize the inner workings of the XCIST toolkit, we provide a brief summary of a series of validation studies of XCIST (including spatial resolution, spectrum and noise modeling, and metal artifacts), and we provide a number of application examples of hybrid data simulations. In summary, XCIST offers accurate physics models in the sense that it can closely reproduce spatial resolution and noise properties of real CT scanners. The hybrid data simulation examples show promising uses for deep learning training data generation. We finally highlight some opportunities for further development and some potential next steps.**

***Index Terms* — Computed Tomography, Simulations, X-ray Physics, Virtual Imaging Trials**

I. Introduction

During the past 20 years, our team at the GE HealthCare Technology & Innovation Center (formerly known as GE Research) developed the CT modeling software CatSim [1]. More recently we developed an open-source version of CatSim and combined it with XCAT phantoms [2] contributed by Duke University and reconstruction routines [3] contributed by University of Massachusetts Lowell, resulting in XCIST [4] (funded by a recently completed NIH NCI ITCR grant U01CA231860). We currently see an average of 1200 views and 30 Git clones per 2-week period [5].

Imaging simulations play three crucial roles in researching and developing new CT technologies:

First, simulations enable and accelerate CT research. In the past, new imaging technologies have taken a long time to develop. CT simulations that faithfully represent images that would result from a real imaging system can dramatically reduce the time to evaluate new technologies. A prerequisite for this is the availability of accurate and efficient models for the emerging CT technologies that are the focus of current cancer imaging research.

Second, simulations are the basis of virtual imaging trials. Optimizing imaging efficacy over a population of patients and tumors for optimal task-based performance requires extensive patient trials. The considerable number of user-defined parameters make a systematic investigation in a clinical trial prohibitively unwieldy. Virtual imaging trials (VITs) [6] are an emerging methodological alternative to clinical imaging trials, where a population of human subjects is replaced with a population of virtual digital subjects, the imaging systems with virtual imaging simulators, and the clinical interpretations with virtual interpretations. As such, VITs provide the practical ability to assess and optimize a host of design factors at the necessary scale in the development and use of imaging technologies—a task not possible through clinical trials. These require an end-to-end simulation environment, including a population of virtual phantoms and lesions on the front-end and image analysis tools and observer models on the rear-end.

Third, simulations are increasingly used for deep learning (DL) training. DL has become pervasive in CT imaging, whether for image generation, for image analysis, or for optimizing the CT scan itself. A major bottleneck in developing DL solutions is collecting large amounts of curated training data. Simulated CT data has proven to be a highly effective means of training neural networks, such as vessel segmentation, artifact reduction, or super-resolution. Generating large amounts of representative CT training data requires efficient and highly realistic simulations.

Specific use case examples of CT simulations include:

- Simulate applications of an existing imaging technology: e.g., evaluate the capability of visualizing angiogenesis with a specific imaging system and a specific contrast agent.
- Simulate the performance of a new imaging system: e.g., evaluate the capability of a new photon-counting CT scanner to detect a specific type of cancer.
- Simulate the impact of a new scan protocol: e.g., evaluate tumor conspicuity with a high-mA, low-kVp CT scan protocol.
- Simulate the impact of a new reconstruction algorithm: e.g., evaluate lesion detectability as a function of dose for a new deep learning reconstruction technique.
- Produce voluminous data for a variety of tumors to train a new image analysis algorithm: e.g., train a new DL architecture for discriminating two types of tumors.
- Simulate new imaging techniques for improved treatment

This abstract was submitted on January 23, 2024.

Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Number U01CA231860 (ITCR grant). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

All authors are with the GE HealthCare Technology & Innovation Center, in Niskayuna, NY, USA and in Bangalore, India.

For any correspondence, please contact Mingye Wu (e-mail: mingye.wu@gehealthcare.com) or Bruno De Man (e-mail: deman@gehealthcare.com).

planning: e.g., simulate a dual-energy CT scan technique for more accurate stopping power images and therefore more precise radiation therapy planning.

Finally, simulations can be used as an education and training tool for biomedical engineers, medical physicists, and radiology residents. The ability to interactively visualize and quantify the effects of patient size and composition, scanner parameters and scan protocols on image quality (IQ) and absorbed dose are of great value in these contexts.

XCIST was developed with the above goals in mind. In this paper we first provide a high-level introduction of XCIST. We then summarize some of the validation studies that have been performed to show that the physics models in XCIST closely approximate real-world CT scanners. Finally, we illustrate some specific use cases of XCIST, with a particular focus on hybrid data CT simulations.

II. Methods

A. *XCIST overview*

The toolkit XCIST includes the CatSim CT simulator as well as a series of XCAT phantoms and reconstruction routines. XCIST was designed to be accessible to many levels of users, from advanced researchers in medical imaging technologies to novices entirely new to simulation. Novice users can download executable code and start from sample simulation scripts. Advanced users can clone the code and develop their own enhancements, with the option to contribute enhancements back to the project. The XCIST code is a hybrid of Python and C/C++. The high-level functions that define models and program flow are written in Python; computationally intensive operations are performed in C/C++. All code is maintained as an open-access project on GitHub.com [5].

XCIST defines objects or patients as a combination of three types of digital phantoms. (1) Analytical phantoms can represent moderately complex phantoms comprised of geometric shapes. (2) Voxelized phantoms can be produced from patient images by assigning a material or a mixture of materials to each voxel based on its CT number. (3) Phantoms comprised of polygon meshes and non-uniform rational b-spline (NURBS) are the basis for (and enable use of) the XCAT phantom library.

The image simulation components of XCIST are a new implementation of CatSim [1], which was developed over the past two decades and previously implemented in Matlab. Based on X-ray physics, CatSim models the X-ray beam, its interaction with matter, and the X-ray detection process. The simulation process can be appreciated by following Figure 1 from top to bottom. First, because simulation of X-ray attenuation by all components depends on the materials, a database of material attenuation properties is defined. The X-ray source spectrum is defined by using a pre-simulated energy spectrum, spatially weighted to model a three-dimensional focal spot profile. The attenuation along each path from source to detector is computed as function of energy. A model for the detector physics converts the detected X-rays to an electronic signal. Objects scatter is estimated and added to the primary radiation. Gantry rotation and patient table translation are modeled by performing simulation at a number of time increments. The X-ray beam is modeled as the sum of multiple projection lines, to realistically reflect finite focal spot size, detector cell size, and rotational blur.

Third-generation clinical CT scanners employ a fan-beam or cone-beam geometry with a curved detector, whereas cone-beam CT (CBCT) scanners employ flat-panel detectors. XCIST includes Feldkamp-type cone-beam reconstruction for axial and helical trajectories.

For dose estimation, XCIST uses a fast, accurate, and patient-specific dose estimation algorithm [7].

A detailed description of the XCIST toolkit is found in [4].

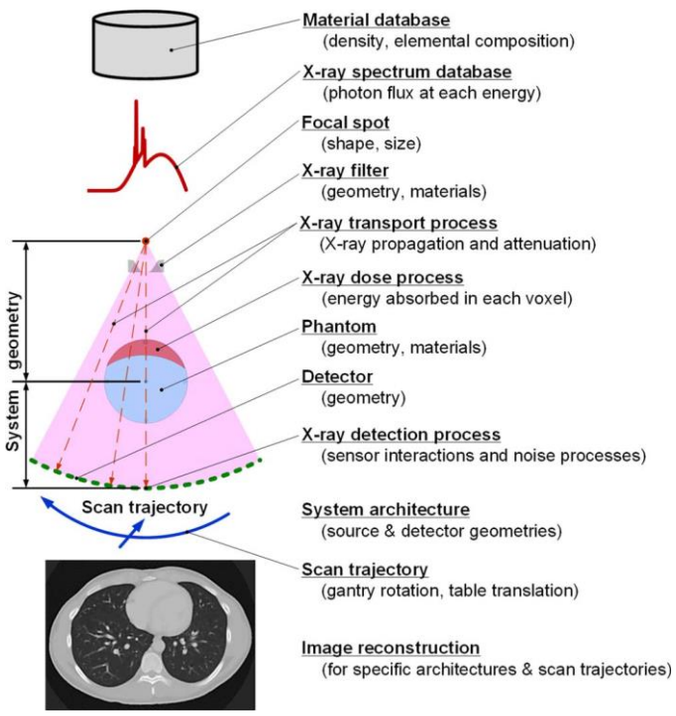


Figure 1: Major X-ray CT processes modeled in CatSim/XCIST are listed, with primary aspects in parentheses.

B. *Spatial resolution modeling*

We compared the spatial resolution obtained empirically with a Lightspeed VCT scanner with the resolution obtained when simulating the same scanner. We evaluated the point-spread function of thin wires in the projection domain and we evaluated both the point spread function and the modulation transfer function of thin wires in the image domain. For the projection domain experiments, we constructed a phantom by inserting two 0.7-mm-diameter stainless steel (SS) wires into a 450-mm-diameter foam cylinder. The wires were located approximately 218 mm from the axis of rotation (AOR); one wire was approximately parallel to the AOR (for in-plane spatial resolution validation) and the other was approximately 45° from the AOR (for subsequent z-resolution validation). For the image domain experiments, we constructed a 450-mm-diameter foam cylinder phantom and we inserted nineteen 0.1-

mm tungsten wires at positions ranging from ~0 to ~225 mm from the phantom center, in ~12.5-mm increments. No attempt was made to exclude the impact of the finite wire thickness from the PSF and MTF metrics since our goal was not to evaluate the absolute system spatial resolution but rather show the close correspondence between simulations and empirical results. A more detailed description of the resolution validation methods is given in [8].

C. Spectrum and noise modeling

Spectra were initialized based on physics models, then refined using empirical measurements on a Lightspeed VCT scanner. Initial spectra at 80, 100, 120, and 140 kVp and at takeoff angles from 5° to 9° were produced using physics-based spectrum estimation tools XSPECT and SpekPy. Empirical experiments were systematically designed with careful selection of attenuator materials and thicknesses, and by reducing measurement contamination from scatter to <1%. Measurements were made on a 64-row CT scanner using the scanner's detector and using multiple layers of polymethylmethacrylate (PMMA), aluminum, titanium, tin, and neodymium. Measurements were made at 80 kVp, 100 kVp, 120 kVp, and 140 kVp and covering the entire 64-row detector (takeoff angles from 5° to 9°); a total of 6,144 unique measurements were made. After accounting for the detector's energy response, parameterized representations of the initial spectra were refined for best agreement with measurements. X-ray transmission errors were computed for measurements versus calculations using the non-optimized and optimized spectra. A more detailed description of the spectrum estimation methods is given in [9].

We further validated the noise behavior by matching three projection-domain metrics between the simulated and the empirical results with a 73.025mm Al slab using a 64-detector-row Lightspeed VCT scanner. Six phantoms designed to evaluate noise modeling in the image domain were also scanned by the same scanner, and further converted to digital phantoms to be simulated with CatSim using the validated spectra, to evaluate the noise performance of CatSim. We scanned phantoms over the standard clinically used range of X-ray techniques (kVp: 80, 100, 120, 140, mA: 25, 50, 100, 300). A more detailed description of the noise estimation methods is given in [10].

D. Hybrid data simulation

A key requirement of many simulation studies is to be both highly realistic (including representing realistic human anatomies) and highly controlled in the sense that the ground-truth of the feature being studied should be precisely known. This can be achieved by combining real CT patient datasets and synthetic features as the objects of the simulation. In doing so, the X-ray physics and reconstruction effects are applied to the combination of real and synthetic components. One challenge of this approach is that the CT patient datasets already have some spatial blur and noise. We compensate for these effects by applying a preprocessing frequency enhancement (to precisely counteract the blur introduced during simulation and reconstruction) and by applying a preprocessing denoising filter (such that the noise in the original image is small or irrelevant in the context of the specific simulation study). For the metal artifact application, scatter was simulated using kernel-based methods and the scatter scaling was tuned to match the Lightspeed VCT scanner. Metal artifact corrections were skipped.

III. Results

A. Spatial resolution validation

Figure 2 shows empirical and simulated wire point spread functions for two wire positions and eliminating any corrections for afterglow or other effects. Qualitatively, the results show remarkable agreement.

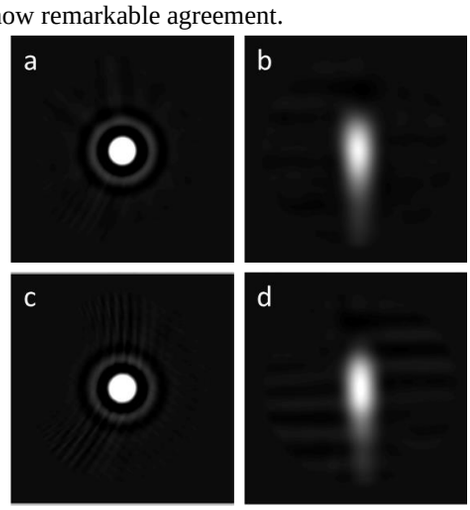


Figure 2: Comparison of wire point spread functions for empirical measurements (top) versus simulations (bottom), for positions close to the AOR (left) and close to the edge of the field-of-view (right).

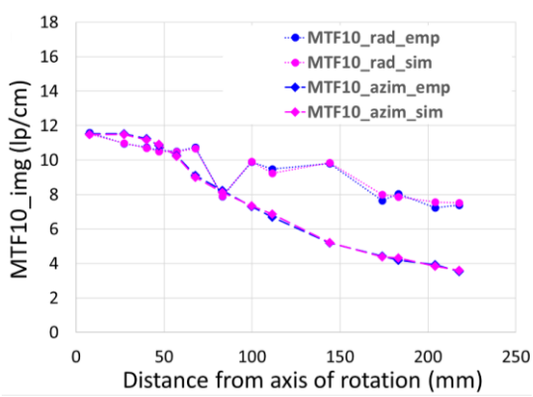


Figure 3: Comparison of the 10% MTF as function of distance to the AOR for empirical measurements (blue) versus simulations (magenta), for radial direction (circles) and azimuthal directions (diamonds).

Figure 3 shows empirical and simulated wire modulation transfer function 10% cutoffs as a function of direction and position. Again, the blue and magenta curves show remarkable agreement (even the dip around 80mm). Quantitatively, all metrics generally agreed well, with most of the average errors substantially less than 5% for all X-ray techniques. Detailed

spatial resolution results are given in [8].

B. Spectrum and noise validation

Figure 4 (left) shows example 120 kVp X-ray tube spectrum estimates, based on XSPECT and iteratively tuned with empirical data. Correspondingly, Figure 4 (right) shows post-log errors for a variety of combinations of materials, which remain within between -0.05 and 0.05. More detailed spectrum validation results are given in [9]. In terms of noise, we found overall good agreement in noise between simulations and measurements. In the projection domain, the differences in noise levels between simulations and empirical measurements were within 3%; while in the image domain, the differences were within 10%. More detailed noise validation results will be given in [10].

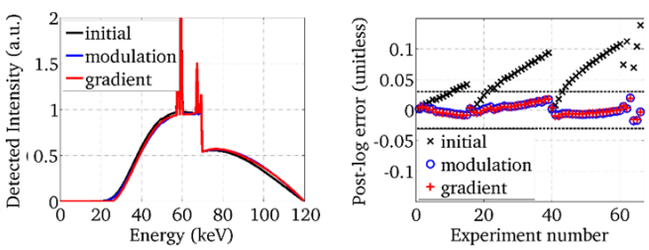


Figure 4: Left: *X-ray spectra initialized with XSPECT for 120 kVp (black line) and optimized using modulation and gradient methods (blue and red lines). Right:* Corresponding p*ost-log errors for the initial spectrum (black) and the modulation- and gradient-based spectra (blue and red).*

C. Hybrid data simulation examples

Several basic simulation experiments and comparisons are shown in [4]. Here we provide three more advanced application examples using a hybrid data simulation approach.

First, we are hosting an AAPM grand challenge on metal artifact reduction. We qualitatively compared metal artifacts in empirical versus simulated images of an anthropomorphic phantom with a series of metal objects (Fig 5 a-b). We then simulated over 10,000 datasets containing real patients and synthetic metal objects [11]. Second, in the context of a liver radiomics study, we created synthetic metastases and embedded these in a hybrid liver CT simulation study (Fig 5 c-d) [12]. Third, we performed a study of calcium blooming artifacts, where we combined synthetic calcifications with real patient CT datasets (Fig 5 e-f) [13].

IV. Discussion & Conclusions

We provided a high-level summary of the XCIST CT simulation environment, which is available in an open-source framework from [5]. We have demonstrated that – with the right parameters – it can accurately represent real-world CT scanners in terms of spatial resolution and noise properties. We have shown important use cases combining human subject CT datasets and synthetic features, such as metal inserts, metastasis and calcifications.

In future work, we intend to augment XCIST with photon-counting CT models, phase-contrast imaging capability, phantom-based IQ and task-based performance evaluation routines, and deep learning interoperability features, and we plan to better validate the dose estimation tool.

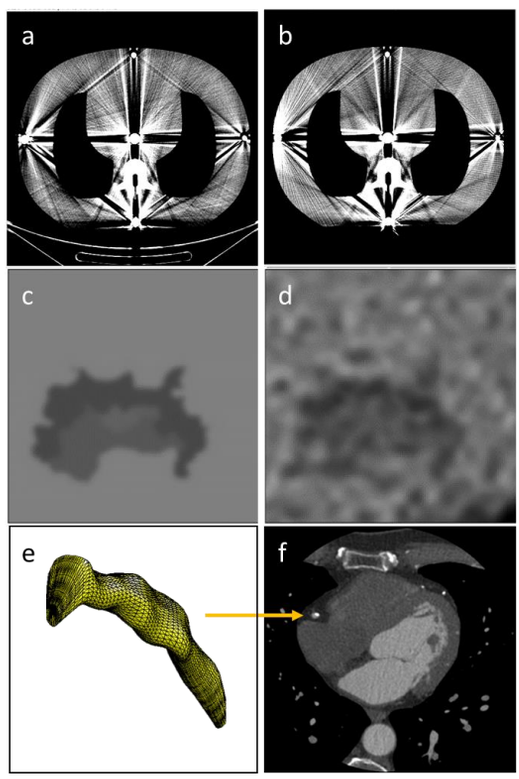


Figure 5: Empirical (a) and simulated (b) anthropomorphic phantom with metal artifacts. Synthetic liver metastasis (c) in a realistic noisy liver background (d). Synthetic calcification (e) simulated inside a patient CT volume.

References

[1] B De Man et al. "CatSim: a new computer assisted tomography simulation environment." SPIE Medical Imaging 2007. Vol. 6510.

[2] P Segars et al. "4D XCAT phantom for multimodality imaging research." Medical physics 37.9 (2010): 4902-4915.

[3] X Tang et al. "A three-dimensional-weighted cone beam filtered backprojection algorithm …", Phys. Med. Biol. 51.4 (2006): 855.

[4] M Wu et al. "XCIST—an open access x-ray/CT simulation toolkit." Physics in Medicine & Biology 67.19 (2022): 194002.

[5] XCIST -- X-ray-based Cancer Imaging Toolkit, GitHub repository. Last updated Jan 2024. URL: github.com/xcist .

[6] E Abadi et al. "Virtual clinical trials in medical imaging: a review." Journal of Medical Imaging 7.4 (2020): 042805-042805.

[7] M Wu et al. "Model-based dose reconstruction for CT dose estimation." Medical Physics 44.9 (2017): e255-e263.

[8] J Zhang et al. "Development and tuning of models for accurate simulation of CT spatial resolution …" Phys. Med. Biol. (accepted).

[9] P FitzGerald et al. "Semiempirical, parameterized spectrum estimation for x-ray computed tomography." Med. Phys. 48.5 (2021): 2199-2213.

[10] J Zhang et al. "Development and application of methods for tuning CatSim for accurate noise simulation" (under review).

[11] AAPM CT Metal Artifact Reduction (CT-MAR) Grand Challenge: URL: www.aapm.org/GrandChallenge/CT-MAR/ .

[12] C Wiedeman et al. "Simulated Deep CT Characterization of Liver Metastases …", (Review Pending).

[13] Y Fan et al. "A virtual imaging trial framework to study cardiac CT blooming artifacts." CT meeting. Vol. 12304. SPIE, 2022.

Accuracy of propagation-based phase-contrast CT under the projection approximation

Giavanna Jadick and Patrick La Rivière, *Member, IEEE*

***Abstract*—X-ray phase-contrast imaging has the potential to improve image contrast with lower dose by probing an object's refractive properties as well as its absorptive properties. To reconstruct a phase-contrast image from a raw dataset, a phase retrieval algorithm must be applied to invert the forward model of the image acquisition scheme. Discrete forward modeling presents unique computational challenges due to high x-ray wave field sampling requirements. At the cost of accuracy, approximations are often applied for the sake of simplicity and experimental convenience. One of the most ubiquitous simplifications is the projection approximation, which neglects refractive effects within an object. The approximation's validity decreases when imaging thicker objects or using a detector with higher spatial resolution. For greater accuracy, one might use the multislice model instead. In this work, we explore the accuracy of the projection approximation in simulated synchrotron micro-CT images. We simulated images using two detector resolutions (0.5 and 2 micron) with either the projection approximation or multislice forward model. Paganin phase retrieval was applied to the resulting datasets, and the final phase-contrast images were reconstructed using filtered back-projection. The 2-micron resolution detector images showed little difference between forward models, indicating the projection approximation is satisfied. With sub-micron resolution, there is a noticeable difference in the phase-contrast images, especially around fine details. An accurate phase retrieval algorithm for such high-resolution data likely requires more detailed forward modeling that avoids making the projection approximation.**

***Index Terms*—micro computed tomography (CT), multislice, phase contrast, phase retrieval, projection approximation, propagation-based imaging (PBI), simulation, x-ray**

I. Introduction

X-RAY phase-contrast imaging (XPCI) has received increasing interest in recent years due to its potential to improve image contrast with lower dose. This is possible by probing an object's refractive properties, rather than absorptive properties alone as in traditional x-ray imaging. Such contrast is especially advantageous in biomedical imaging scenarios that benefit from high contrast between soft tissues with similar attenuation coefficients.

To produce phase-contrast images from a raw detected signal, one must solve the so-called "phase retrieval problem." A variety of techniques have been developed to encode phase information for this purpose, including edge-illumination, grating-based, speckle-tracking, and propagation-based imaging setups [1]. The latter, propagation-based imaging (PBI), offers the unique advantage of not requiring any special optical equipment [2]. As one shifts the detector further from the object being imaged, phase differences accumulated by the modulated x-ray wave field are amplified. If the radiation source is sufficiently coherent and the spatial resolution is high enough, as in synchrotron micro-CT imaging, Fresnel fringes begin to show along object edges in the detected intensity. The experimental convenience and single-shot potential of PBI lends itself to tomographic imaging [3].

However, propagation-based XPCI faces several challenges before clinical realization and optimal implementation. Discrete forward modeling presents unique computational challenges; depending on the energy of the x-ray beam, sub-micron sampling is likely necessary for modeling free-space propagation, even for detectors with relatively low resolution. Further, to solve the phase retrieval problem, one must invert the chosen forward model, which may not have an analytical solution and thus can magnify the computational challenge. For this reason, several approximations are commonly implemented in XPCI forward models.

One of the most ubiquitous simplifications is the projection approximation, which neglects refractive effects within an object. This is typically valid when the object's thickness is within the near-field regime, generally at higher energies and lower resolutions [1]. More accurately, one can utilize a "multislice" forward model, which consists of dividing an object into sufficiently thin slices and accounting for refraction between subsequent layers. Under the projection approximation, the x-ray wave field exiting an object is characterized by the line integrals through the real (phase) and imaginary (absorptive) components of the object's complex index of refraction $n_E(\vec{r}) = 1 - \delta_E(\vec{r}) + i\beta_E(\vec{r})$, where E specifies energy-dependence. If one can recover the complex exit wave from the raw data, under the projection approximation, it is straightforward to use the real and imaginary line integral components for tomographic reconstruction.

In the case that the projection approximation is not valid, it is not immediately obvious how or to what degree the resulting tomographic reconstruction will differ from expectation. With rapidly advancing x-ray imaging technology, achieving improved spatial resolutions for larger fields-of-view, this topic is increasingly relevant [4], [5]. The purpose of this work is to explore the impact of the projection approximation in micro-CT images.

This work was supported by the United States National Science Foundation Graduate Research Fellowship under Grant No. 2140001.

G. Jadick (e-mail: giavanna@uchicago.edu) and P. La Rivière (e-mail: pjlarivi@uchicago.edu) are with the Department of Radiology, University of Chicago, Chicago, IL 60637, USA.

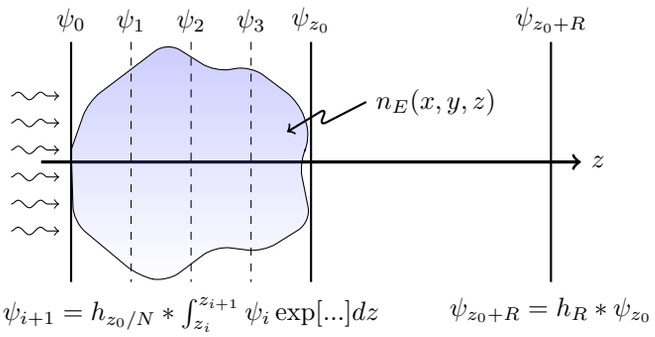


Fig. 1: Schematic of the multislice forward model using N slices, where $\exp[...]$ indicates the line integrals through the object's complex index of refraction as given in Eq. 1.

II. Methodology

Propagation-based XPCI micro-CT acquisitions were simulated using the projection approximation and multislice forward models. Simulation parameters were varied to probe conditions where the projection approximation may or may not be valid. The approximated phase-contrast images were compared to the multislice images to assess how the difference in forward model affected the simulation accuracy.

A. Forward models

We modeled a coherent, monoenergetic plane wave ($n_E \equiv n$, wavenumber $\equiv k$). Under the projection approximation, the wave field exiting an an object can be written in terms of the accrued amplitude (β) and phase (δ) shifts,

$$\psi_{z_0} = \psi_0 \exp\left[-k \int_0^{z_0} dz\ \beta(\vec{r}) + i\delta(\vec{r})\right], \qquad (1)$$

where z is the optical axis and z_0 is the location of the object exit plane [1]. One metric of projection approximation validity is the condition $N_F \gg 1$, with N_F as the Fresnel number

$$N_F = \Delta^2(\lambda z_0)^{-1}, \qquad (2)$$

where Δ is the spatial resolution, λ is the x-ray wavelength, and z_0 is the object thickness [1]. If this condition is not satisfied, it is more accurate to use the multislice approach (Fig. 1). The object is divided into N slices, each of which should be sufficiently thin to satisfy the condition $N_F \gg 1$. The object exit wave is then computed iteratively. The wave field at each successive slice ψ_{i+1} is found by multiplying the incident wave ψ_i by the line integrals through the object slice (as in Eq. 1) and then applying free-space propagation over the slice width z_0/N.

We modeled the free space propagation of a wave as a convolution with the Fresnel operator,

$$\psi_{z_0+R} = \psi_{z_0} *\ h_R = \mathcal{F}^{-1}\left\{\mathcal{F}\left\{\psi_{z_0}\right\} \cdot H_R\right\}, \qquad (3)$$

where h_R is the spatial-domain and H_R is the Fourier-domain Fresnel operator for propagation distance R,

$$H_R(\nu_x, \nu_y) = \exp\left[-\frac{i\pi hc}{E} R(\nu_x^2 + \nu_y^2)\right], \qquad (4)$$

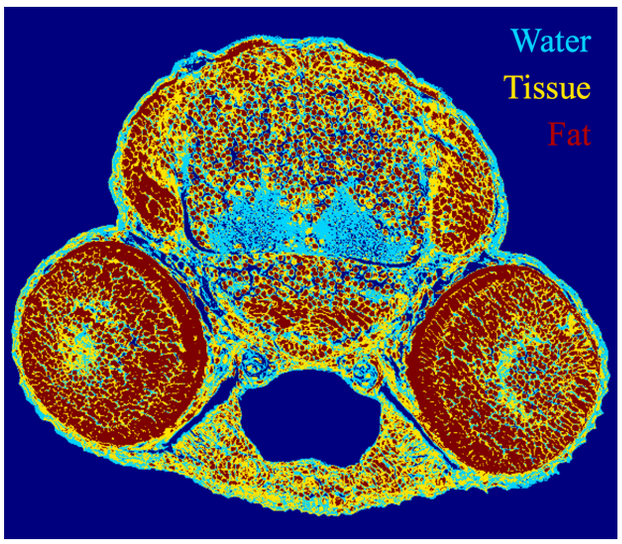


Fig. 2: The 5-mm zebrafish phantom used for the micro-CT simulations.

h is Planck's constant, and c is the speed of light in a vacuum.

After modeling the incident x-ray wave field's modulation through the object, free-space propagation was applied to compute the wave arriving at a detector placed distance R from the object exit plane, ψ_{z_0+R}. The final detected signal is equal to the magnitude of this complex wave field.

B. Simulations

Parallel-beam, 20-keV micro-CT images of a 5-mm zebrafish phantom (Fig. 2) were simulated using either the projection approximation or multislice modulation with 10 slices. The computational phantom was created from a micro-CT image of a metal-stained zebrafish. The two metals, osmium and uranyl acetate, approximately targeted cell lipids and nuclei, respectively. This facilitated segmentation of the CT image into four materials (air, water, fat, tissue).

Each simulation used a 50-mm propagation distance to the detector and 7-mm beam width. Two detector resolutions were modeled (2.0 and 0.5 μm), representing tasks that straddle the projection approximation criterion ($N_F = 0.8$ and 12.9, respectively). The wave field was upsampled to a resolution of 0.25 μm during each simulation. Projections were acquired at 5000 view angles over a 180-degree rotation. Paganin's method for phase retrieval was applied to each projection assuming a single-material object of water [6]. This technique utilizes the projection approximation, couples the chosen material's δ and β values, and yields an estimate of the object thickness through each ray. The estimated thickness sinograms were reconstructed into a 5-mm, 4096×4096 array (1.22-μm pixel size) using filtered back-projection.

III. Results

Figures 3 and 4 show the reconstructed phase-contrast images simulated with and without the projection approximation

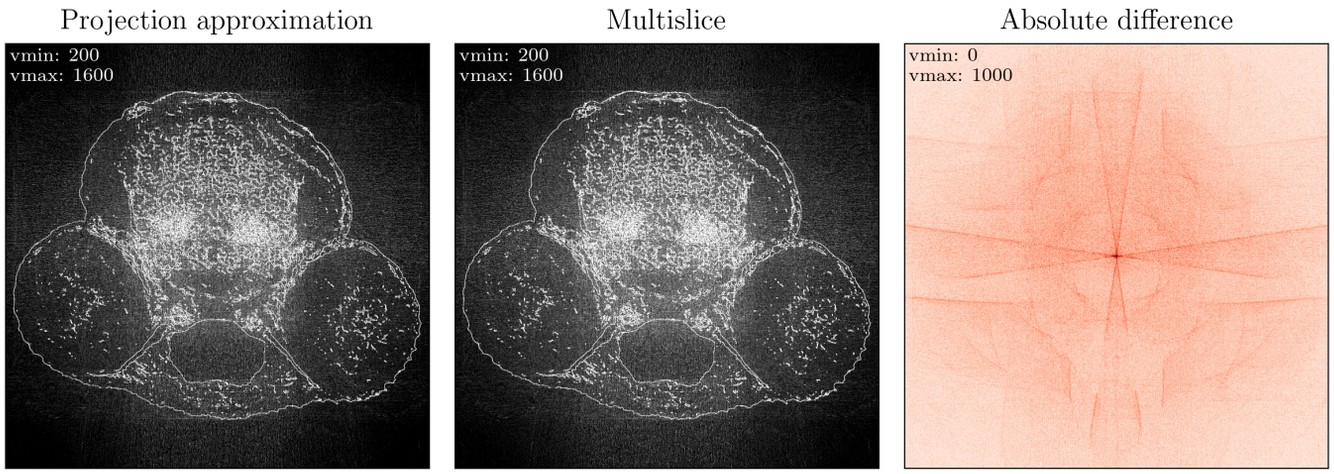


Fig. 3: For the lower resolution, 2-μm detector pixel data, reconstructed images including Paganin phase retrieval and the absolute difference between the images using different forward models.

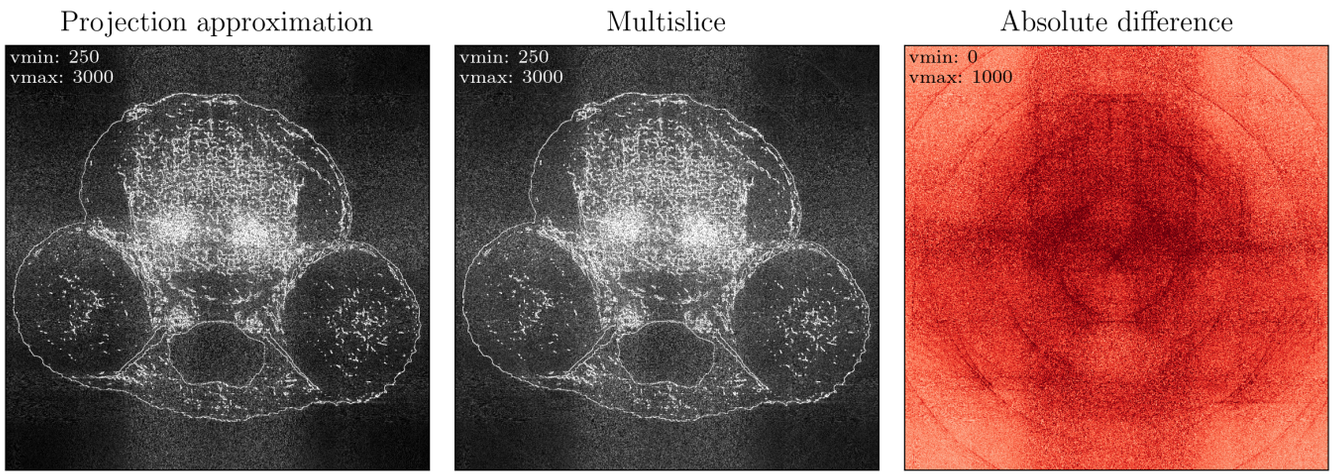


Fig. 4: For the higher resolution, 0.5-μm detector pixel data, reconstructed images including Paganin phase retrieval and the absolute difference between the images using different forward models.

at the two detector resolutions. The images using the 2-μm detector resolution show little difference whether or not applying the projection approximation. This indicates that the conditions for the projection approximation are likely satisfied, as predicted by $N_F = 12.9 \gg 1$.

The 0.5-μm images show a greater difference between the two forward models, especially in regions of the phantom with fine details. This is expected, as the smaller Fresnel number ($N_F = 0.8$) for this detector indicates that refractive effects within the 5-mm phantom thickness are not insignificant, and propagation-based phase effects manifest as Fresnel fringes along the edges of structures. This increased intensity at structure edges appears to cause bright streaks in the reconstructed images along lines passing through many fine structures. This is slightly more noticeable in the multislice image but apparent with the projection approximation as well.

In all reconstructions, intricate fatty regions (membranes) display good contrast, but it is challenging to resolve tissue regions (cellular nuclei) on the water background. There is a hazy increase in intensity towards the center of the images, particular with the multislice simulations. This may be due to the images not satisfying the conditions for Paganin phase retrieval. In addition to the projection approximation, Paganin's method assumes the object is made of a single material object and the detector is in the near-field regime.

Figure 5 shows the exit wave intensity difference between the two forward models. The 0.5-μm sampling is sufficiently high to resolve striations in the sinogram that appear along the edges of fine structures, while the 2-μm sampling shows aliasing artifacts due to block averaging over the high-frequency fringes around these fine structures.

IV. Discussion

The projection approximation is widely implemented in propagation-based XPCI forward models due to its simplicity, convenience for phase retrieval, and applicability to tomo-

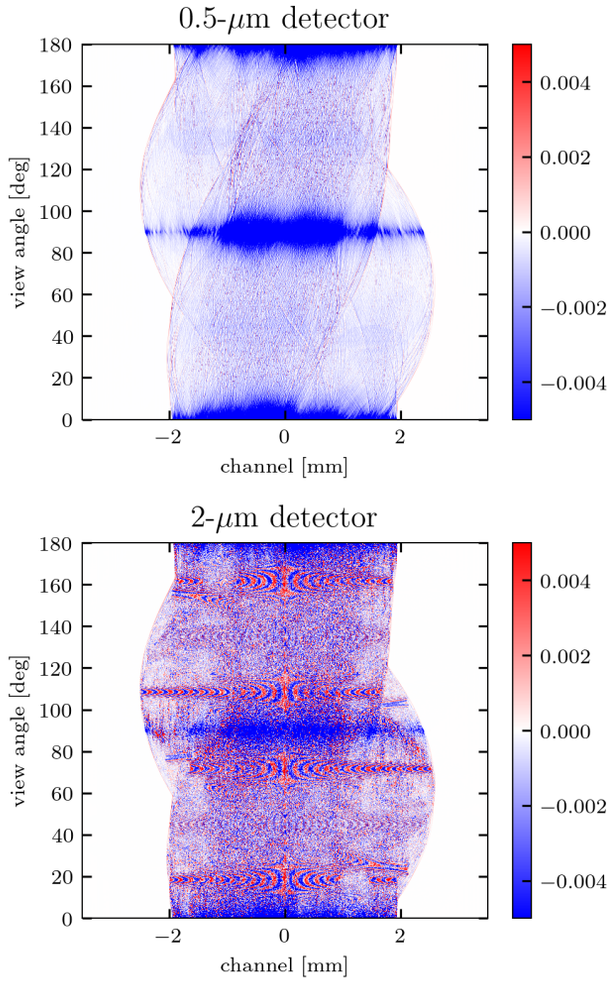


Fig. 5: Difference in exit wave intensity (projection approximation – multislice) for the two resolutions.

graphic reconstruction. The approximation is valid when the refractive effects within an object are negligible at the target resolution. As modern imaging tasks utilize larger objects and smaller detector pixels, accurate modeling may require a more complex approach such as the multislice forward model. In this work, we assess how different forward models affect simulated micro-CT images. We find that with a sub-micron resolution detector, images simulated with the projection approximation noticeably differ from those using the multislice approach. The simulated image pairs offer qualitative insight into the consequences of violating this forward model assumption.

We applied Paganin phase retrieval to the simulated CT acquisitions, as it is widely utilized in experimental settings due to its convenience of only requiring a single input image [6]. More generally, since a phase retrieval algorithm aims to recover two unknowns (phase and amplitude components of an object), two independent measurements are required. These are typically acquired at different propagation distances, which is experimentally inconvenient and error-prone [3]. To perform phase retrieval from just one image, Paganin's algorithm applies simplifications that reduce its quantitative accuracy, including the projection approximation. Our results match the expectation that when the projection approximation is not satisfied (0.5-µm detector), there is a discrepancy in Paganin phase-contrast images simulated with the multislice forward model. Paganin's algorithm also assumes that an object comprises a single material and that the propagation distance to the detector is in the near-field regime. While the zebrafish phantom has a composition similar to that of water, it also included fat and soft tissue. Although the 2.0-µm detector images show only small differences between the projection approximation and multislice forward models, there are imperfections in the phase retrieval that are apparent in both images regardless of resolution. This may be due to not satisfying the additional assumptions of Paganin's algorithm.

Tomographic reconstruction algorithms such as filtered back projection require a set of line integrals through an object as their input. For a phase retrieval algorithm that rightfully applies the projection approximation, one expects to recover the line integrals through an object's δ and β components. Since Paganin's phase retrieval assumes a single material, it recovers line integrals through the object thickness. If the assumptions for the phase retrieval algorithm are not satisfied, propagation-based artifacts such as residual Fresnel fringes will contaminate the recovered line integrals. We see this effect especially in the 0.5-µm detector images, where the assumptions of Paganin's algorithm are less valid. Consequently, the reconstructed phase-contrast image appears to have increased intensity along lines passing through highly detailed structures.

V. Conclusion

For accurate propagation-based XPCI forward modeling, the multislice approach may be used for high-resolution imaging of relatively thick objects. To reconstruct phase-contrast images for such datasets, phase retrieval methods should account for the refractive effects occurring within the object and minimize their use of other simplifying assumptions. At lower resolutions, the projection approximation is often sufficient.

Acknowledgment

The authors would like to thank Phillip Vargas for creating the zebrafish phantom used in the micro-CT simulations.

References

[1] D. M. Paganin and D. Pelliccia, "X-ray phase-contrast imaging," in *Advances in Imaging and Electron Physics*, vol. 218, pp. 63–158, Elsevier.
[2] T. E. Gureyev, S. C. Mayo, D. E. Myers, Y. Nesterets, D. M. Paganin, A. Pogany, A. W. Stevenson, and S. W. Wilkins, "Refracting Röntgen's rays," *J. Appl. Phys.*, vol. 105, no. 10, p. 102005.
[3] A. Burvall, U. Lundström, P. A. C. Takman, D. H. Larsson, and H. M. Hertz, "Phase retrieval in x-ray phase-contrast imaging suitable for tomography," *Opt. Express*, vol. 19, no. 11, pp. 10359–10376.
[4] C. Sundberg, M. U. Persson, J. J. Wikner, and M. Danielsson, "1-μm spatial resolution in silicon photon-counting CT detectors," *J. Med. Imaging*, vol. 8, no. 6, p. 063501.
[5] M. A. Yakovlev, D. J. Vanselow, M. S. Ngu, C. R. Zaino, S. R. Katz, Y. Ding, D. Parkinson, S. Y. Wang, K. C. Ang, P. La Rivière, and others, "A wide-field micro-computed tomography detector: micron resolution at half-centimetre scale," *J. Synchrotron Radiat.*, vol. 29, no. 2, pp. 505–514.
[6] D. Paganin, S. C. Mayo, T. E. Gureyev, P. R. Miller, and S. W. Wilkins, "Simultaneous phase and amplitude extraction from a single defocused image of a homogeneous object," *J. Microsc.*, vol. 206, no. 1, pp. 33–40.

Image Reconstruction With Local Directional Total Variation Regularization Using Tomographic Incompleteness Maps

Matthieu Laurendeau, Frédéric Jolivet, Laurent Desbat, Guillaume Bernard, Sébastien Gorges and Simon Rit

***Abstract*—Limited angle acquisition is a well-known challenge in computed tomography reconstruction because the lack of data generates severe geometric artifacts in reconstructed images. Recently, directional total variation (DTV) regularization has shown promising results for this kind of problem but it requires fine-tuning of a global directional hyperparameter in addition to the regularization weight. In this work, we propose a new regularization, called local directional total variation (LDTV), which is a DTV based on local directional weights determined from tomographic incompleteness maps. We evaluated LDTV and compared it to state-of-the-art algorithms on simulated two-dimensional acquisitions of the Forbild head phantom with a source trajectory made of two orthogonal arcs of 60° each. The reconstructed images show that the LDTV regularization performs better in this geometry for both noiseless and noisy data.**

***Index Terms*—limited angle problem, tomographic incompleteness, directional total variation (DTV), local DTV (LDTV)**

I. Introduction

In X-ray computed tomography (CT), some applications require limited space which implies a reduced acquisition angle. This situation could arise both in the medical field, with tomosynthesis for example, and in the industrial field where the aim is to minimize the size of the system and reduce its scanning time.

Several image reconstruction algorithms have been developed to address the limited angle problem. Iterative methods have demonstrated improved results compared to filtered-backprojection (FBP) algorithms [1]. They minimize a cost function made of a data fidelity term and a regularization term. The latter implements an image prior compensating the lack of data to obtain more plausible results.

M. Laurendeau is with Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Étienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69373, Lyon, France, with Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France and with Thales AVS, Moirans, France (email: matthieu.laurendeau@insa-lyon.fr).

F. Jolivet is with Thales AVS, Moirans, France (email: frederic.jolivet@thalesgroup.com).

L. Desbat is with Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France (email: laurent.desbat@univ-grenoble-alpes.fr).

G. Bernard is with Thales AVS, Moirans, France (email: guillaume.bernard@thalesgroup.com).

S. Gorges is with Thales AVS, Moirans, France (email: sebastien.gorges@thalesgroup.com).

S. Rit is with Univ Lyon, INSA-Lyon, Université Claude Bernard Lyon 1, UJM-Saint Étienne, CNRS, Inserm, CREATIS UMR 5220, U1294, F-69373, Lyon, France (email: simon.rit@creatis.insa-lyon.fr).

Recently, a new regularization, directional total variation (DTV), was developed [2] and showed promising results for the limited angle problem. This regularization strengthens the image prior in the direction where data is missing. In [3], the authors studied orthogonal arcs with limited angular range for dual-energy CT application and showed similarly good results using DTV.

Although DTV could improve visual image quality, it requires fine-tuning of the directional parameter, which could be time-consuming and sensitive to the acquisition geometry and the scanned object. Moreover, the regularization applies the same directional weight for all pixels of the image, which may not accurately describe spatial variability of data incompleteness due to limited angle acquisitions.

In this work, we present the local directional total variation (LDTV) regularization based on the tomographic incompleteness map previously developed in [4] which uses the incompleteness metric of [5]. This regularization locally adjusts both the strength and the directional weight parameters by taking into account the geometry of the scanner. We simulate noiseless and noisy projections of the Forbild head phantom [6] with a limited angle geometry made of double orthogonal arcs (DOA), compute the corresponding tomographic incompleteness map and compare several regularizations, including state-of-the-art DTV [1] and our LDTV method.

II. Materials and methods

A. Geometry and tomographic incompleteness

Similarly to [3], we used a two-dimensional (2D) fan-beam geometry with limited angle DOA. An illustration of the geometry is provided in Fig. 1. We used 120 source positions, evenly distributed every degree on two orthogonal arcs of 60° each. The source-to-object distance (SOD) and the source-to-detector distance (SDD) were 510 mm and 1020 mm, respectively. In addition to having an acquisition limited to an angular range of 120°, this geometry was selected because the direction of missing data is spatially variable with both horizontal and vertical directions.

Monochromatic fan-beam projections were simulated from a reference image of the Forbild head phantom [6] in the region $\Omega \subset \mathbb{R}^2$ with 256×256 pixels and 1 mm spacing (Fig. 1). The linear detector was large enough (1200 mm, 1 mm spacing) to avoid lateral truncation of the projections.

We computed the tomographic incompleteness map of the scanner geometry, as described in [4], in the source trajectory

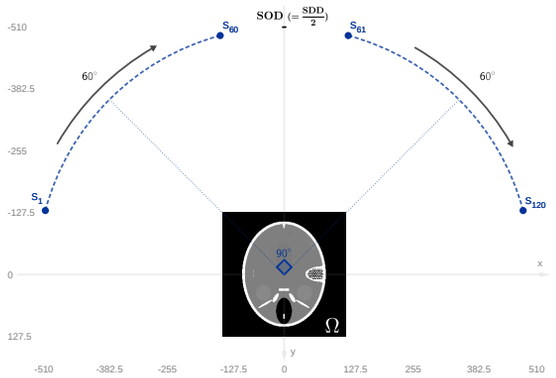


Fig. 1. Double orthogonal arcs (DOA) geometry. The two 60° blue dashed arcs represents the 120 sources distributed every degree. The image of the Forbild head phantom is placed at the center of the arcs.

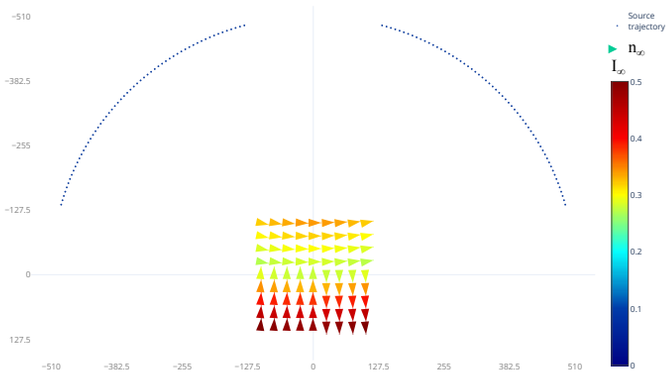


Fig. 2. Tomographic incompleteness map of the DOA geometry. The source points are represented in blue. The arrows show the most incomplete co-direction $\mathbf{n}_\infty$ and their color the corresponding tomographic incompleteness value I_∞.

plane. The tomographic incompleteness criterion [5], noted $\mathrm{I}(\mathbf{p},\mathbf{n}) \in \mathbb{R}^+$, measures how much a 3D point fulfills Tuy's condition, i.e., *for a continuous source trajectory* Γ *and without truncations, a point* $\mathbf{p}$ *can be reconstructed if all planes passing through* $\mathbf{p}$ *intersect* Γ. One advantage of the incompleteness criterion I is that it allows evaluation of the tomographic incompleteness for a discrete source trajectory. The value of $\mathrm{I}(\mathbf{p},\mathbf{n})$ is defined as the tangent of the minimal angle between the plane $\Pi_{\mathbf{p},\mathbf{n}}$ passing through $\mathbf{p}$ with co-direction $\mathbf{n}$ and all the X-ray lines defined by the source points and $\mathbf{p}$ [5]. When $\mathrm{I}(\mathbf{p},\mathbf{n})$ reaches 0, $\Pi_{\mathbf{p},\mathbf{n}}$ intersects the source trajectory and larger values of $\mathrm{I}(\mathbf{p},\mathbf{n})$ indicate a larger angle between $\Pi_{\mathbf{p},\mathbf{n}}$ and the closest source position. As in [4], the tomographic incompleteness map calculates for all pixels $\mathbf{p}$ in the image the most incomplete co-direction:

$$\mathbf{n}_\infty(\mathbf{p}) = \arg\max_{\mathbf{n}} \mathrm{I}(\mathbf{p},\mathbf{n}) \quad \forall \mathbf{n} \in S^2_{1/2}, \tag{1}$$

with $S^2_{1/2}$ the unit hemisphere, and the corresponding incompleteness value:

$$\mathrm{I}_\infty(\mathbf{p}) = \mathrm{I}(\mathbf{p},\mathbf{n}_\infty) = \max_{\mathbf{n}} \mathrm{I}(\mathbf{p},\mathbf{n}) \quad \forall \mathbf{n} \in S^2_{1/2}. \tag{2}$$

Numerically, for this 2D geometry, we only evaluated 720 homogeneously distributed co-directions of $S^2_{1/2}$ in the source trajectory plane, i.e. on the unit half circle. Fig. 2 illustrates the 2D incompleteness map for our DOA geometry with colored arrows, each giving two pieces of information: the direction encodes the direction $\mathbf{n}_\infty$ normal to the most incomplete line and the color encodes the incompleteness value I_∞. In Fig. 2, the map is roughly separated in two regions: the upper half with horizontal co-directions and tomographic incompleteness values $\mathrm{I}_\infty(\mathbf{p})$ in the range $[0.27, 0.34]$, and the lower half with vertical co-directions and tomographic incompleteness values of about 0.40. This is expected as the hole between the two arcs of the source trajectory creates an incomplete direction defined by the midpoint of the arcs $(0,-510)$ for pixels in the upper half, while for pixels in the lower half, the incompleteness is larger in the horizontal direction. In summary, the studied geometry implies a significant tomographic incompleteness $\mathrm{I}_\infty(\mathbf{p}) > 0.26\ \forall \mathbf{p} \in \Omega$, the upper half is mainly incomplete in the vertical direction while the lower half is even more incomplete but in the horizontal direction.

The aim of the proposed algorithm is to account for this incompleteness map in the regularization of an iterative reconstruction algorithm adapted to the limited angle problem.

B. Local Directional Total Variation

Data collected by the X-ray tomography system using our DOA geometry was modelled by the following discrete linear system:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \epsilon \tag{3}$$

where $\mathbf{x} \in \mathbb{R}^{N=N_i \times N_j}$ is the sought attenuation map, $\mathbf{A} \in \mathbb{R}^{M \times N}$ the discrete X-ray transform of the DOA geometry, $\mathbf{y} \in \mathbb{R}^M$ the projection data and $\epsilon \in \mathbb{R}^M$ the associated noise. In a limited angle framework as the DOA geometry, reconstructing $\mathbf{x}$ with FBP leads to poor results due to the lack of data. In recent years, iterative reconstruction algorithms with an image prior have given the best results by constraining the problem to a space of plausible solutions [1]. To estimate $\mathbf{x}$, one can solve the following optimization problem:

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \frac{1}{2}||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \mathrm{R}(\mathbf{x}) \tag{4}$$

where the first term, the data fidelity term, ensures that the final solution $\mathbf{x}^*$ matches the projection data $\mathbf{y}$ and R is the regularization term embedding the image prior.

Total variation (TV) regularization, which promotes piecewise constant solutions, has proven efficient for CT reconstruction, in particular for the limited angle problem [1]. Recently, Zhang *et al.* [2] developed DTV regularization based on the anisotropic version of TV, which weights differently the regularization on the horizontal and the vertical axes of the image. It can be written as:

$$\begin{aligned}\mathrm{R}_{\mathrm{DTV}}(\mathbf{x}) = \lambda \sum_{i=0}^{N_i-1} \sum_{j=0}^{N_j-1} &|\beta\,(\mathbf{x}_{i,j} - \mathbf{x}_{i+1,j})\,| \\ &+|\sqrt{1-\beta^2}\,(\mathbf{x}_{i,j} - \mathbf{x}_{i,j+1})\,|\end{aligned} \tag{5}$$

where $\lambda \in \mathbb{R}^+$ controls the strength of the regularization with respect to the data fidelity term and $\beta \in [0,1]$ is the directional weight. DTV is well adapted to the limited-angle problem where the lack of data is in the same direction for

every pixel of the reconstructed image. The parameter β is tuned to regularize more this specific direction. However, it is not obvious to guess the best β because it depends on the geometry and the scanned object. Moreover, DTV defines only one direction parameter while in some geometries, e.g. the DOA geometry, the incompleteness direction is not the same for every pixel in the image, as explained above.

We propose to use the tomographic incompleteness map to locally adjust the regularization. We formulate our LDTV regularization as follows:

$$\begin{aligned}\mathrm{R}_{\mathrm{LDTV}}(\mathbf{x}) = \\ \sum_{i=0}^{N_i-1}\sum_{j=0}^{N_j-1} \tilde{\mathrm{I}}_\infty(\mathbf{x}_{i,j})(|\,\mathbf{n}_{\infty 2}(\mathbf{x}_{i,j})\,(\mathbf{x}_{i,j}-\mathbf{x}_{i+1,j})\,| \\ +|\,\mathbf{n}_{\infty 1}(\mathbf{x}_{i,j})\,(\mathbf{x}_{i,j}-\mathbf{x}_{i,j+1})\,|)\end{aligned} \quad (6)$$

with $\mathbf{n}_{\infty 1}$ and $\mathbf{n}_{\infty 2}$ the first and second components of the co-direction $\mathbf{n}_\infty$ and

$$\tilde{\mathrm{I}}_\infty(\mathbf{x}) = \lambda_{\min} + \frac{\lambda_{\max}-\lambda_{\min}}{\max \mathrm{I}_\infty}\,\mathrm{I}_\infty(\mathbf{x}) \quad (7)$$

the tomographic incompleteness value linearly adjusted between the hyperparameters $\lambda_{\min} \in \mathbb{R}^+$ and $\lambda_{\max} \in \mathbb{R}^+$. The regularization is stronger when the incompleteness level is higher and the horizontal and vertical directional weights are defined by $\mathbf{n}_{\infty 2}$ and $\mathbf{n}_{\infty 1}$ respectively as they design the direction normal to the most incomplete line.

C. Numerical experiments

The cost function of Eq. 4 is convex but non-smooth when dealing with TV-like regularizations. Aiming at a fair comparison between several regularization terms, we have used the same FISTA-TV optimization algorithm with non-negative constraints for all iterative methods [7].

We generated noiseless projections of the Forbild head phantom and corrupted them with pre-log Poisson noise with 10^6 photons per pixel in air. We reconstructed images using the noiseless data with four algorithms: conventional FBP with a Hamming window and three iterative reconstruction algorithms with anisotropic TV, DTV and the proposed LDTV regularization terms. The two last have also been used to reconstruct images from noisy data. For the TV, DTV and LDTV, we performed 1200 iterations of FISTA with 60 denoising iterations which was enough to achieve the convergence according to the relative error with the ground truth (GT).

Image quality was quantified with the peak signal-to-noise ratio (PSNR). For the regularization terms, the hyperparameters were automatically tuned by minimizing the normalized root mean squared error (NRMSE) with the GT image using 600 iterations of the Nelder-Mead downhill simplex algorithm.

III. Results

The reconstructed images with noiseless projections are shown in Fig. 3. FBP reconstruction was severely impacted by the lack of data, and neither the inner left ear nor the profile were close to the GT. The high frequency noise in this result demonstrated that limiting the solution to be more piece-wise constant was relevant. TV reconstruction, with $\lambda = 2.9\times10^{-4}$, really increased the visual quality of the image with a PSNR score of 28 dB. The image was more piece-wise constant, as expected, but there were residual artifacts at the border of the head, and it had difficulty to reconstruct high frequency information in the two ears. The best hyperparameters for DTV regularization were $\lambda = 2.8 \times 10^{-4}, \beta = 0.97$ and the final reconstruction had a PSNR score of 30 dB, notably better than the reconstruction with TV regularization. One can see the effect of the horizontal directional regularization on the left ear compared to the TV result. Also, the small disks in the right ear were closer to the GT. Finally, the reconstruction with LDTV regularization outperformed the other reconstruction algorithms with a high PSNR value of 36 dB confirmed by the visual impression of a higher image quality. The best hyperparameters for LDTV were $\lambda_{\min} = 3.3 \times 10^{-5}$ and $\lambda_{\max} = 2.6 \times 10^{-4}$. The borders of the head were better reconstructed, as one can notice above the left and right inner ears or at the bottom of the phantom. Moreover, the left inner ear showed 7 squares as the GT and the right ear was more accurate. Lastly, the profiles show that LDTV was able to reconstruct the two low contrast spheres at the top of the phantom unlike the other reconstruction algorithms.

DTV and LDTV images reconstructed from noisy data are shown in Fig. 4. DTV had a PSNR of 27 dB with $\lambda = 6.4\times10^{-4}$ and $\beta = 0.99$. Visually, noise clearly degraded the result. The reconstructed image with LDTV regularization had a PSNR of 30 dB. The optimal hyperparameters were $\lambda_{\min} = 2.7 \times 10^{-4}$ and $\lambda_{\max} = 9.3 \times 10^{-4}$ and the visual image quality was improved compared to DTV. The right ear and the border of the head were closer to the GT. However, the left ear and the low contrast profile were visually degraded compared to the result without noise.

IV. Discussion

The LDTV method uses the incompleteness map (Fig. 2) and therefore applies a stronger regularization in the lower half of the image than the upper half, in the horizontal directions and vertical directions respectively. DTV uses a single parameter and cannot locally adjust the regularization direction. The optimal directional hyperparameter of DTV was horizontal, which is consistent with the most incomplete direction of the incompleteness map and supports its use in LDTV.

For this work, we did not study the dependence of the regularization hyperparameters to the object. We have used an optimization algorithm to find the hyperparameters that minimize the NRMSE with the GT. The optimal hyperparameters will not only depend on the scanner geometry but also on the scanned object. The LDTV regularization adjusts the directional weights using the tomographic incompleteness, but the optimal hyperparameters $\lambda_{\min}$ and $\lambda_{\max}$ will probably depend on the characteristics of the scanned object.

The LDTV method gave satisfactory results in this 2D framework. The tomographic incompleteness maps deliver three-dimensional (3D) knowledge on the CT geometry. Computing LDTV regularization for a 3D images was beyond the scope of this work.

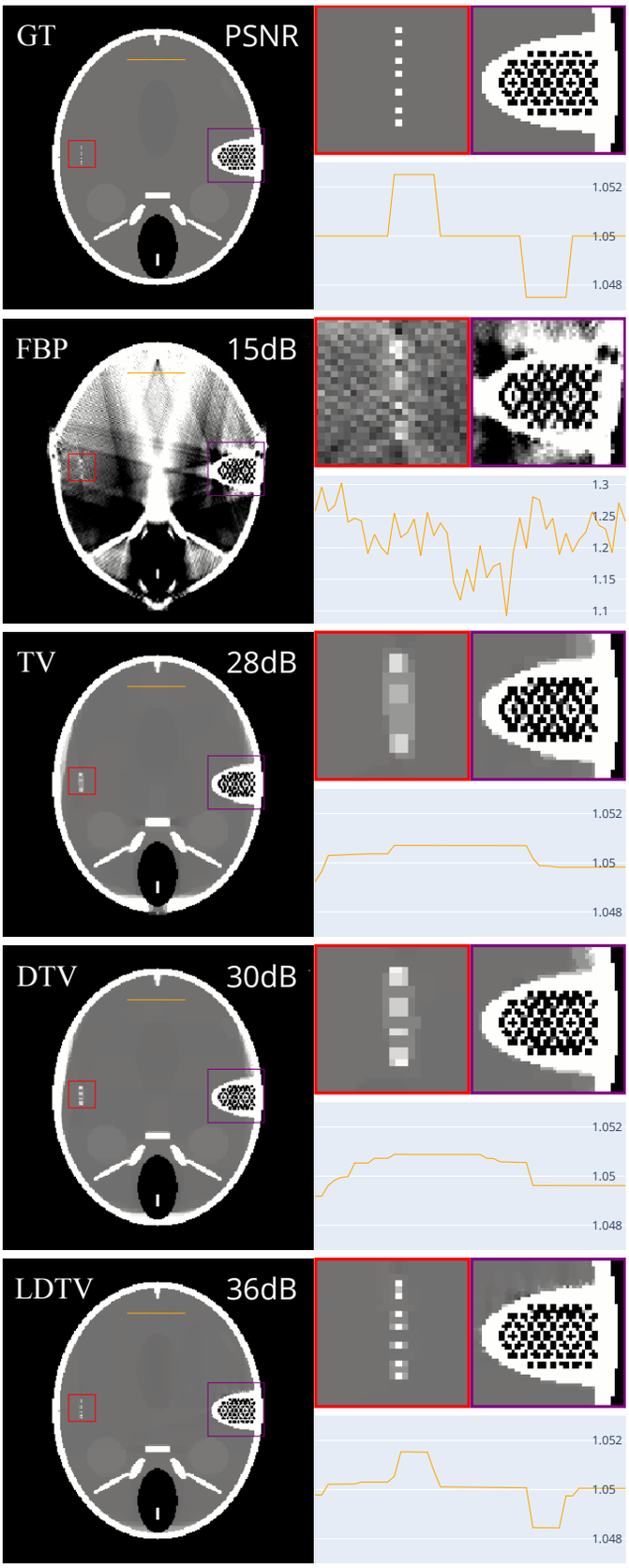


Fig. 3. GT and results of the four reconstruction algorithms using noiseless data with a gray window of $[0.9, 1.2]$. The left column shows the full reconstructed image with the corresponding PSNR value at the top-right. In the right column, the two zoom images represent the inner left and right ears, and the orange profile is taken along the line passing through the two low contrast spheres at the top of the phantom.

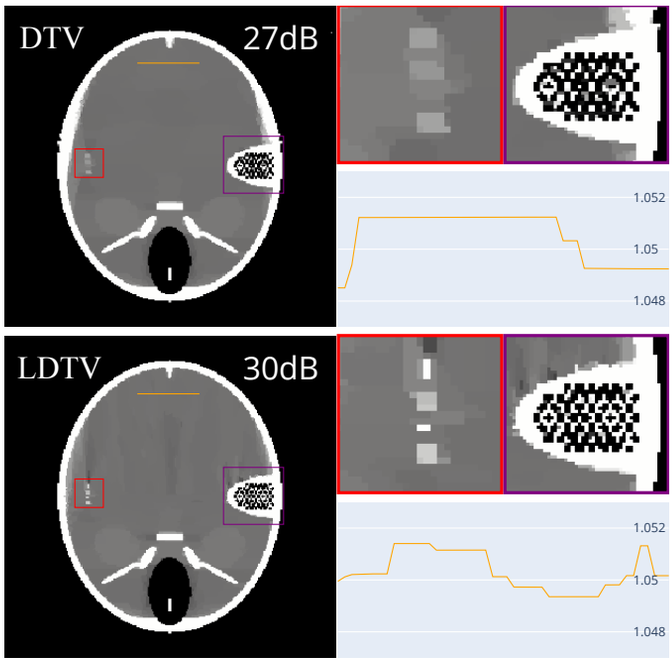


Fig. 4. Reconstructed images using DTV and LDTV regularization from noisy projections. The gray window is $[0.9, 1.2]$. As in Fig. 3, the full reconstructed images with their corresponding PSNR values are in the left column beside the zooms on the ears and the plot through the two low contrast spheres.

V. Conclusion

This article presented the LDTV, a new regularization using local tomographic incompleteness direction and level. To illustrate its effect, we have simulated and reconstructed noiseless and noisy data from the Forbild head phantom in a limited angle geometry. We have compared our method to state-of-the-art reconstruction algorithms. The results showed that LDTV regularization is a good solution to account for the acquisition geometry in the regularization term. LDTV may therefore be used to improve image quality of CT images reconstructed from projections acquired with a limited angle geometry.

References

[1] E. Sidky, C.-M. Kao, and X. Pan, "Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT," *Department of Radiology, University of Chicago, 5841 S. Maryland Ave., Chicago, IL, 60637, USA*, vol. 14, pp. 119–139, Apr. 2006.

[2] 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*, vol. 70, p. 102030, may 2021.

[3] B. Chen, Z. Zhang, D. Xia, E. Y. Sidky, and X. Pan, "Dual-energy CT imaging over non-overlapping, orthogonal arcs of limited-angular ranges," *Journal of X-Ray Science and Technology*, vol. 29, no. 6, pp. 975–985, 10 2021.

[4] M. Laurendeau, L. Desbat, G. Bernard, F. Jolivet, S. Gorges, F. Morin, and S. Rit, "Three-dimensional maps of the tomographic incompleteness of cone-beam ct scanner geometries," in *Fully 3D Reconstruction In Radiology and Nuclear Medicine*, 2023.

[5] R. Clackdoyle and F. Noo, "Quantification of tomographic incompleteness in cone-beam reconstruction," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 4, no. 1, pp. 63–80, Jan. 2020.

[6] Forbild, 2003. [Online]. Available: www.dkfz.de/en/roentgenbildgebung/ct/CT_Phantoms/Head_Phantom/Head_Phantom.html

[7] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–2434, nov 2009.

Deep Learning Based Scatter Estimation Using PYRO-NN: Application to Real-World Objects

Markus Michen, Rajib Kumar Chanda

***Abstract*—This paper presents an advanced CT image processing method that uses deep learning and the PYRO-NN framework to reduce artifacts caused by scattered radiation. The goal is to improve defect detection precision during manufacturing processes, which will be evaluated in future studies. This work contributes to the development of an end-to-end trainable reconstruction pipeline based on previous work, featuring a U-Net model for scatter estimation. Refined label reconstructions, regularization losses, and modified U-Net architectures (MBConvBlocks and EfficientNetV2 encoder) were implemented to further enhance the model's capabilities. Preliminary tests on real-world objects show promising results, with reduced impacts of scattered radiation artifacts compared to uncorrected scans. The approach has the potential to greatly improve the effectiveness and efficiency of CT-based image processing in industrial applications.**

***Index Terms*—Industrial Computed Tomography, Artifact Correction, Scatter Estimation, Deep Learning, U-Net, Nondestructive testing.**

I. Introduction

THE significance of X-ray computed tomography (CT) as a non-destructive testing methodology continues to expand across numerous sectors, with applications ranging from inspection and analysis up to measurement technology. Within the realm of modern industrial manufacturing processes, CT systems have become indispensable tools for real-time quality assurance and control. These systems are typically integrated within the manufacturing line, allowing for immediate scanning and inspection of each produced component. By integrating CT systems into the production process often referred to as Inline-CT, manufacturers can maintain high product quality standards while minimizing rejects and operating costs. These systems must meet strict requirements for scan times and algorithm execution in order to maintain the desired speed of the equipment.

For all CT-based image processing scenarios, artifacts substantially impair image quality and may obstruct or hinder subsequent algorithms acting on the reconstructed data. Therefore, an effective artifact reduction approach is essential to guarantee optimal results.

A key contributor to image artifacts in CT-based tasks is scattered radiation. Due to scattered photons interfering with the unscattered primary radiation, non-linearities are introduced in the imaging process. Factors such as energy, scanning geometry, and the properties of the measured object influence the degree of scattering, which can lead to streaks, decreased contrast, and a cupping effect in the reconstructed volumes (see Figure 1). Simultaneously occurring with scatter, beam hardening arises when the average energy of the X-rays increases with object thickness, causing the beam to become "harder". This phenomenon introduces beam hardening artifacts, characterized by dark shadows and the recurring cupping effect. Distinguishing between scattering and beam hardening effects can be challenging, as they both originate from non-linearities in the imaging process and yield comparable artifacts in the reconstructed data.

Scatter artifact reduction methods can be physical techniques like anti-scatter grids and collimators or algorithmic approaches such as Monte Carlo simulations, kernel-based techniques, and deep learning.

A prominent deep learning-based method was introduced by J. Maier et al., employing a convolutional network to mimic MC simulation for scatter estimation [1], [2]. However, generating training samples with Monte Carlo simulations can be computationally intensive. Additionally, errors are calculated in the projection domain, while undesired artifacts from computed tomography appear in the reconstructed domain.

Our research builds upon the work of Michen et al. [3]. Their method incorporates an end-to-end trainable reconstruction pipeline that performs error calculations in the volume domain using PYRO-NN [4]. It embeds a DSE-like [1], [2] U-Net [5] into the pipeline and updates it by utilizing the reconstruction layers. The U-Net estimates scattered photon intensity, which is subtracted from the input projection to minimize artifacts in the reconstructed volume.

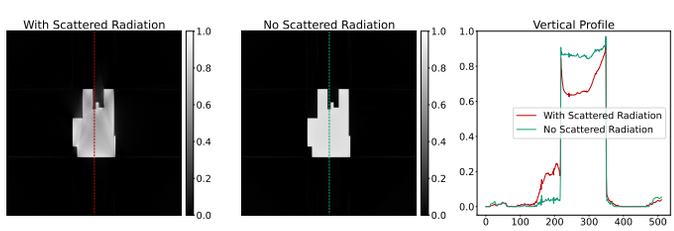


Fig. 1. Simulation of a volume with and without scattered radiation to visualize artifacts of scattered radiation. On the right the corresponding line profile illustrates the cupping effect.

We introduce additional enhancements to their method, such as a refined label reconstruction, regularization loss integration, and adapting the U-Net architecture to incorporate MBConvBlocks [6] and a encoder derived from EfficientNetV2 [7]. These modifications aim to boost the efficiency and effectiveness of the scatter estimation process.

We trained and evaluated our developed method on real

Markus Michen, Rajib Kumar Chanda are with the Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS, Division Development Center X-Ray Technology (email: markus.michen@iis.fraunhofer.de)

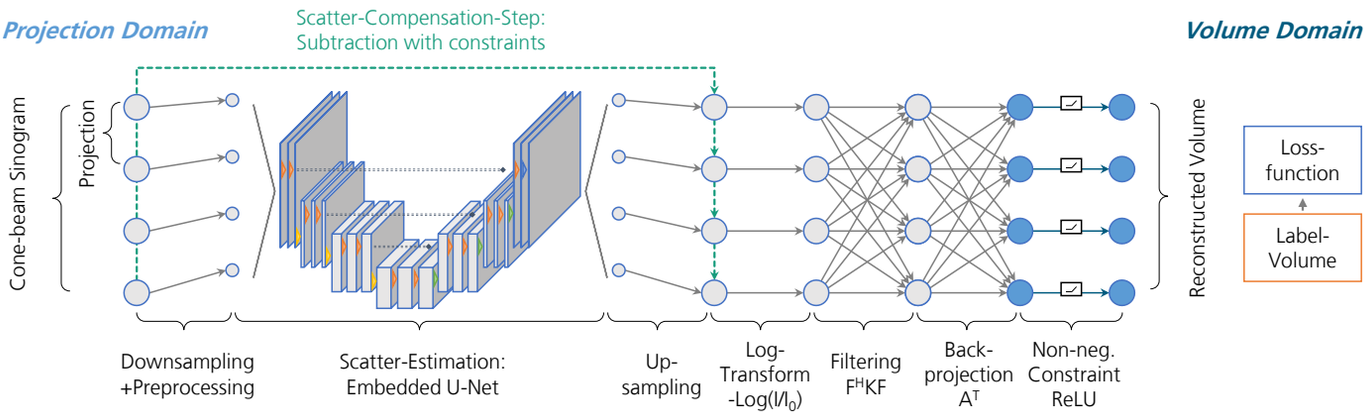


Fig. 2. The end-to-end CT reconstruction pipeline with the embedded U-Net. The light gray dots denote operations in projection domain and the darker blue filled dots operations in volume domain. On top the equations of pipeline can be seen. In green the dashed line highlights the scatter compensation step after the U-Net. Figure adopted and updated from [3].

world objects. We present early findings from our ongoing research and development. In the near future, we aspire to integrate our approach seamlessly into inline CT operations to boost reconstruction accuracy in high scatter scenarios, simplifying subsequent defect detection. We also intend to analyze the impact of enhanced scatter estimation on the defect segmentation.

II. Methods

A. Pipeline

Our approach utilizes an end-to-end reconstruction pipeline. The primary concept behind our developed scatter correction strategy is to calculate the loss function in the volume domain which is affected by artifacts. A U-Net, which is embedded within the pipeline, estimates the scatter distribution for the input projections. The PYRO-NN framework offers the necessary CT operations to transition between the projection and volume domains. This enables backpropagation of errors through the reconstruction layers, including the U-Net operating in the projection domain, resulting in updates to the network's weights. The pipeline consists of several stages and can be seen in Figure 2:

1) **Normalize** input projections by their respective $\boldsymbol{I}_0$ values.
2) **Preprocessing** the projections via log transformation and the 'pep' ($\boldsymbol{I} \cdot -\log(\boldsymbol{I})$) mapping, as these steps have demonstrated a higher correlation with scatter distributions [1].
3) **Scatter estimation**. Feeding the preprocessed projections into the U-Net, whose weights are trained such that the output represents a scatter distribution that can be subtracted from the input projections. To conserve GPU memory, the input projections are downsampled prior to being fed into the U-Net. After the U-Net processing, the projections are upsampled back to their original resolution without significantly sacrificing accuracy. This is achievable due to the low frequent nature of scatter distributions.
4) **Scatter compensation**. To maintain the quality of the projections during scatter compensation, we employ one of three methods: clipping the U-Net output at 95% of the projection value, applying relaxation to restrict the maximum subtracted value, or rescaling the projections to a reasonable range. At present, we utilize clipping.
5) **Filtered backprojection**. Log transforming the compensated projections, followed by filtering backprojection using Fourier transform, multiplication, and inverse Fourier transform. The backprojection is provided by the PYRO-NN package in a custom TensorFlow operation with highly optimized CUDA implementation.
6) **Postprocessing** the output volume with a ReLU activation function.

Finally, the loss between the reconstruction and a label reconstruction is calculated and then backpropagated. We minimize both the L_1- and L_2-losses between the current reconstruction ($\boldsymbol{x}$) and the label reconstruction ($\boldsymbol{y}$), while incorporating total variation regularization $L_{TV}(\boldsymbol{x})$ on the current reconstruction. In the following the pipeline is derived. The current reconstruction ($\boldsymbol{x}$) is defined as:

$$\boldsymbol{x} = \boldsymbol{A}^T \boldsymbol{F}^H \boldsymbol{K} \boldsymbol{F} \cdot -\log(\boldsymbol{I} - \boldsymbol{s}) \tag{1}$$

where $\boldsymbol{I}$ are the normalized intensities by $\boldsymbol{I}_{measured}/\boldsymbol{I}_0$, $\boldsymbol{A}^T$ the backprojection operator supplied by PYRO-NN, $\boldsymbol{F}$ the Fourier transform, $\boldsymbol{F}^H$ its inverse and $\boldsymbol{K}$ the Ram-Lak filter. The current scatter estimate is given by $\boldsymbol{s}$ which is calculated by:

$$\boldsymbol{s} = \text{U-Net}(\boldsymbol{w}, \boldsymbol{I}) \tag{2}$$

So finally the optimized loss reads:

$$L(\boldsymbol{I}, \boldsymbol{y}) = \lambda_1 \cdot L_1(\boldsymbol{x}, \boldsymbol{y}) + \lambda_2 \cdot L_2(\boldsymbol{x}, \boldsymbol{y}) + \lambda_{TV} \cdot L_{TV}(\boldsymbol{x}) \tag{3}$$

with λ_1 and λ_2 equal to 1.0 and $\lambda_{TV} = 0.001$.

B. Training

Our pipeline is exclusively trained on real data using five scans of a typical part from aluminum casting, which features long transmission lengths and shows high amounts of scattered radiation. It was sourced from a CT system that is integrated

into a manufacturing line. Constraints on scan time, geometry, and algorithm runtimes were stringent, reflecting the factory environment. For testing purposes, we used a comparable object scanned under marginally varied parameters. The specific CT settings are detailed in Table I.

In the study mentioned in [3], the original label reconstructions were determined through basic global thresholding. To enhance accuracy, we modified our approach using a local thresholding technique from the skimage library [9]. By employing this improved method, we are better able to differentiate between artifacts and objects, avoiding misclassifications that could occur when using simple global thresholds. Throughout the training, the label reconstruction $\boldsymbol{y}$ can undergo periodic updates, synchronizing its progression with the continuous enhancement of the (initial) reconstruction $\boldsymbol{x}$.

For training we use the AdamW [8] optimizer with a learning rate of $3e^{-4}$ with cosine decay and linear warmup. Weight decay is set to $1e^{-5}$.

After training the U-Net within our end-to-end CT reconstruction pipeline, users may opt to use their own reconstruction tool with just the fast U-Net for scatter estimation. Additionally, in scenarios requiring real-time adaptation, such as when matching a machine's cycle time, scatter distributions can be precomputed based on the scanned object and fixed CT parameters, enabling enhanced efficiency.

TABLE I
CT SCANNING PARAMETERS OF THE TRAINING AND TESTING DATA.

Scan parameters	**Train**	**Test**
Objects	1 object 2 orientations, 5 scans	1 object
Detector pixel size	1460 x 730	960 x 960
Detector size	81.1 x 40.5 [cm]	38.4 x 38.4 [cm]
Focus object distance	194 cm	200 cm
Focus detector distance	234 cm	237 cm
Energy	450 kV	225 kV
Trajectory	circle 360°	circle 360°
Number of projections	800	800
Overall set size	5 scans · 800 proj.	1 scan · 800 proj

III. RESULTS

Our findings involve a metal casting object, characterized by long transmission lengths and heightened scattered radiation. To show generalization across object-, scanning geometry and energy we evaluate the trained network using a scan with different CT parameters than in the training data. And although the scanning parameters differ compared to the training dataset, the results demonstrate noticeable improvements.

We depict a line profile traversing the center of the object and present the corresponding intensity values. Moreover, we compute the signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) within the object at the rectangular regions shown in Figure 3. Our analysis shows the central slice of the testing object being examined (see Figure 3).

The initial reconstruction exhibits significant cupping and smearing artifacts. While the proposed technique alleviates the cupping issue, some smearing persists. Both the line profile analysis and the visual comparison support this observation (see Figure 3). The green line profile of the corrected slice in Figure 3 exhibits a diminished cupping effect relative to the red line profile, which represents the initial scattered reconstruction, this is also supported by the difference profile. Moreover, the significant improvement in signal-to-noise ratio (SNR) from 8.3731 to 16.6385 demonstrates some enhanced quality of the reconstructed images. Although the contrast-to-noise ratio (CNR) experienced only a minor increase from approximately 20 to 22, this indicates a more discernible distinction between different regions. These results collectively highlight the advantageous impact of our proposed technique on CT-based image processing in industrial applications.

Although our pipeline was primarily developed for the correction of scattered radiation, it also takes into account beam hardening artifacts, as these are caused by nonlinearities in the imaging process and cause similar artifacts in the volume domain. Optimizing the pipeline to consider both processes separately could improve the quality of the correction.

Our scattering estimation leads to a better quality reconstruction, but its accuracy in representing the "true scattering distribution" is uncertain due to factors such as detector or system scattering. Conventional methods (beam stops, collimation, ...) for scatter measurement are expensive, complex, hardly applicable to past scans and only provide an approximation of the scatter distribution. MC simulation, which

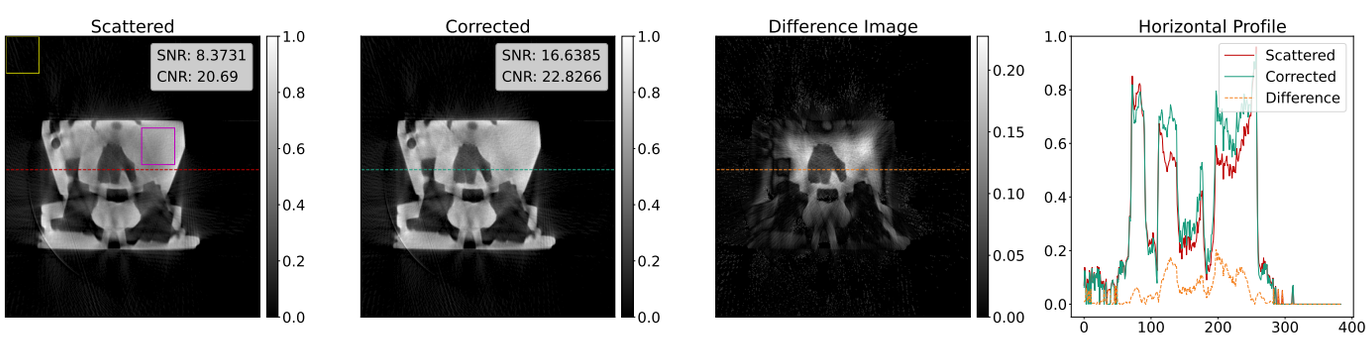


Fig. 3. Preliminary results of our method applied to a cylinder head. From left to right: The uncorrected reconstruction (scatterd), the corrected reconsturction, their absolute difference and the horizontal line profile plot. The yellow box in the upper left is used for the CNR computation. The magenta colored box in the center of the object is used for the SNR calculation. In the line profile plot also a difference of the profiles is given.

could theoretically provide a ground truth, is computationally impractical with large amounts of data. Therefore, any scatter estimate that yields a sufficiently good reconstruction can be considered a good scatter estimate since we do not have access to the ground truth.

IV. Conclusion

Our research introduces an advanced CT image processing method using deep learning and the PYRO-NN framework to reduce artifacts caused by scattered radiation. The goal is to enhance defect detection precision during manufacturing processes, which will be evaluated in subsequent investigations. Additionally, we aim to seamlessly integrate this technology into existing production inline CT systems to optimize reconstruction accuracy under high scattering conditions.

This work contributes to the development of an end-to-end trainable reconstruction pipeline based on [3], featuring a U-Net model for scatter estimation. To further advance its capabilities, we have implemented refined label reconstructions, integrated regularization losses, and modified the U-Net architecture by incorporating MBConvBlocks and EfficientNetV2 encoders.

The results of this study testing on real-world objects demonstrate improved performance compared to uncorrected scans, minimizing the impact of scattered radiation artifacts.

In conclusion, our approach is promising and could significantly improve the effectiveness and efficiency of CT-based image processing in industrial applications.

Author Contributions Statement

Markus Michen - conceptualized the research, carried out the study, authored the complete report and implemented the pipeline.

Rajib Kumar Chanda - focused on evaluating the pipeline, testing various hyperparameters and improving label estimation.

All authors reviewed the manuscript.

References

[1] 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, 2018.

[2] 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," *Medical physics*, vol. 46, no. 1, pp. 238–249, 2019.

[3] M. Michen, U. Haßler, and C. Syben, "Deep learning based scatter estimation," 2022.

[4] C. Syben, M. Michen, B. Stimpel, S. Seitz, S. Ploner, and A. K. Maier, "Pyro-nn: Python reconstruction operators in neural networks," *arXiv preprint arXiv:1904.13342*, 2019.

[5] 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.

[6] S. Gupta and M. Tan, "Efficientnet-edgetpu: Creating accelerator-optimized neural networks with automl," *Google AI Blog*, vol. 2, no. 1, 2019.

[7] M. Tan and Q. Le, "Efficientnetv2: Smaller models and faster training," in *International conference on machine learning*. PMLR, 2021, pp. 10 096–10 106.

[8] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in adam," 2018.

[9] S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu, "scikit-image: image processing in python," *PeerJ*, vol. 2, p. e453, 2014.

CT Gantry Rotation Time and Reconstructed Frame Rate Impact the Accuracy of Flow Field Estimation Using a Physics-informed Neural Network

J. Guo, G. Khurana, L. Severance, B. Maidu, A. Gonzalo, J.C. del Alamo, F. Contijoch

***Abstract*— Contrast-enhanced computed tomography (CT) is a fast and robust imaging modality which acquires 3D high-resolution spatial acquisitions over time. Currently, estimation of 4D flow from CT relies on using anatomical segmentations to prescribe patient-specific boundary conditions for computational fluid dynamics (CFD) simulations. Unfortunately, CFD simulations can be highly sensitive to modeling choices and boundary conditions. Physics-Informed Neural Networks (PINNs) can recover 4D flow fields from movies visualizing the dynamics of a contrast agent, are less sensitive than CFD to model parameters, and do not require boundary conditions. This could enable estimation of blood flow directly from 4D CT contrast dynamics, but it remains unknown how CT data acquisition and reconstruction parameters affect PINN flow estimation. In this study, we used CFD to simulate contrast passing through a 2D Y-shaped vessel with pulsatile flow. Then, we simulated CT scanning with different gantry rotation speeds (2-20 Hz) and used filtered backprojection with different number of frames for image reconstruction. The resulting time series of 2D images were used as inputs to train a PINN constrained by the equations of fluid motion and contrast transport. By comparing the flow fields recovered by the PINN to the ground truth CFD data, we evaluated how artifacts in the simulated image acquisition process and changes in frame rate impacted the accuracy of flow reconstruction. We observed that a PINN's ability to estimate flow velocity increases with CT gantry rotation speed and with frame rate used to reconstruct contrast CT images.**

***Index Terms*—Computed Tomography, Neural Network, Flow Estimation**

I. Introduction

X- ray computed tomography (CT) imaging is widely used clinically to evaluate cardiovascular anatomy and function. CT can provide time-resolved volumetric assessment of the patient's heart and blood vessels. The combination of fast, volumetric high-resolution acquisition and the ability to evaluate patients with implanted metal devices makes X-ray imaging a promising imaging technique to evaluate blood flow irregularities [1].

However, a CT-based quantitative approach for blood flow measurement (similar to Doppler echocardiography or phase contrast encoding of MRI signals) is not available [2],[3]. Instead, flow estimation relies on building computational fluid dynamics (CFD) models in patient-specific anatomies. However, the results of CFD simulations depend on user-defined parameters such as boundary conditions and model assumptions. For example, a Y-shaped vessel will have a flow distribution which depends on the downstream vascular resistance; a parameter set by the user in the CFD model which cannot be directly measured from CT [4],[5]. Therefore, the potential for X-ray based methods to quantitatively evaluate blood flow fields over space and time remains underutilized.

A recent study by Raissi et al. introduced a Physics-Informed Neural Networks (PINNs) framework to estimate fluid flow and pressure fields from images of spatiotemporally evolving contrast [6]. It does so by penalizing deviations from the observed contrast movie ($\mathscr{L}_{\text{Data}}$) as well as adherence to the Navier-Stoke equations ($\mathscr{L}_{\text{Flow}}$) via loss functions shown below:

$$\mathcal{L} = \mathcal{L}_{Data} + \beta \mathcal{L}_{Flow}$$

$$\mathcal{L}_{Data} = \sum_{n} (c(n) - \hat{c}(n))^2$$

$$\mathcal{L}_{Flow} = \iiint [\left(\frac{\partial \hat{u}}{\partial t} + (\nabla \cdot \hat{u})\hat{u} + \nabla \hat{p} - \frac{1}{Re}\nabla^2 \hat{u}^2\right) + (\nabla \cdot \hat{u})^2 + (\frac{d\hat{c}}{dt} + (\nabla \cdot \hat{u})\hat{c} + \frac{1}{Pe}\nabla^2 \hat{c})^2] \, dxdydt$$

where n is the spatiotemporal coordinate *(x, y, z, t),* Re is the Reynolds number, Pe is the Peclet number and β is a parameter that specifies the level of regularization. $\hat{c}$, $\hat{u}$ and $\hat{p}$ are estimates of contrast concentration, flow velocity and pressure and c are the ground truth contrast concentration. $\mathscr{L}_{\text{Data}}$ penalizes the discrepancy between measured and predicted contrast

This work was supported by the Shu Chien – Gene Lay Department of Bioengineering, University of California, San Diego.

J.G., L.S., and F.C., are with the Shu Chien – Gene Lay Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: jig006@ucsd.edu, lseveran@ucsd.edu, fcontijoch@ucsd.edu).

G.K. is with the Department of Computer Science Engineering, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: gukhurana@ucsd.edu).

B.M., A.G., and J.C.A, are with the Mechanical Engineering Department, University of Washington, Seattle, WA 98195 USA (e-mail: bmaidu@uw.edu, algongra@uw.edu, juancar@uw.edu)

J.C.A. is also with the Division of Cardiology, University of Washington, Seattle WA, 98195 USA.

F.C. is also with the Department of Radiology, University of California, San Diego, La Jolla, CA 92093 USA.

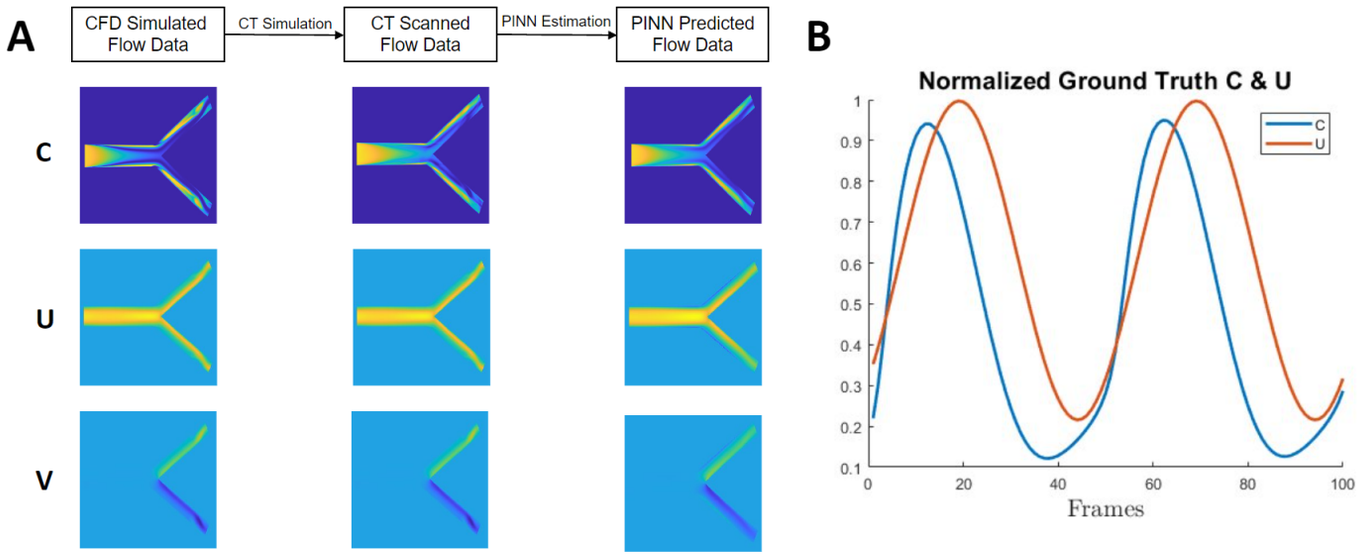


Fig. 1. In-silico pipeline and sample data. Panel A shows the three data processing steps and corresponding sample contrast/velocity fields at time frame 50. Panel B shows the normalized (from 0 to 1) pulsatile flow velocity and contrast concentration value as a function of frame numbers at the vessel inlet.

evolution, $c\ (x, y, z, t)$ and $\hat{c}\ (x, y, z, t)$. $\mathscr{L}_{\text{Flow}}$ is a combination of Navier-Stokes and Advection-diffusion Transport equations. The total loss is the sum of the two equations weighted by β and the equations are weighted equally in this study (β=1). Automated differentiation and gradient descent enable minimization of this loss function.

However, the extent to which this PINN framework can be used to estimate flow from CT images is unknown. Specifically, CT images suffer artifacts due to having a finite temporal resolution (determined by the CT gantry rotation speed) and a limited frame rate. The impact of these limitations on the accuracy of flow estimated using this PINN has not been assessed [7], [8].

II. METHODS

An in-silico pipeline was developed to assess the performance of the Neural Network-based flow estimation in CT shown in Figure 1.

A. CFD Simulation

A 2D Y-shaped vessel of 1 cm in diameter and 7 cm in length was designed and used for pulsatile flow simulation (1 cycle per second, mimicking a heart rate of 60 beats per minute with average forward flow of 20 cm/s) using CFD. Contrast injections were applied at the inlet and the concentration (normalized from 0 – 100 %) across the whole vessel is shown in Figure 1A. Laminar flow was simulated with pulsatile contrast injection and flow velocity shown in Figure 1B. This results in movies of contrast evolving over space and time $c_{CFD}(x, y, z, t)$ and velocity fields on x and y direction, $[u_{CFD}(x, y, z, t), v_{CFD}(x, y, z, t)]$.

B. CT Scanning

The CT scanning process was simulated in Python. The

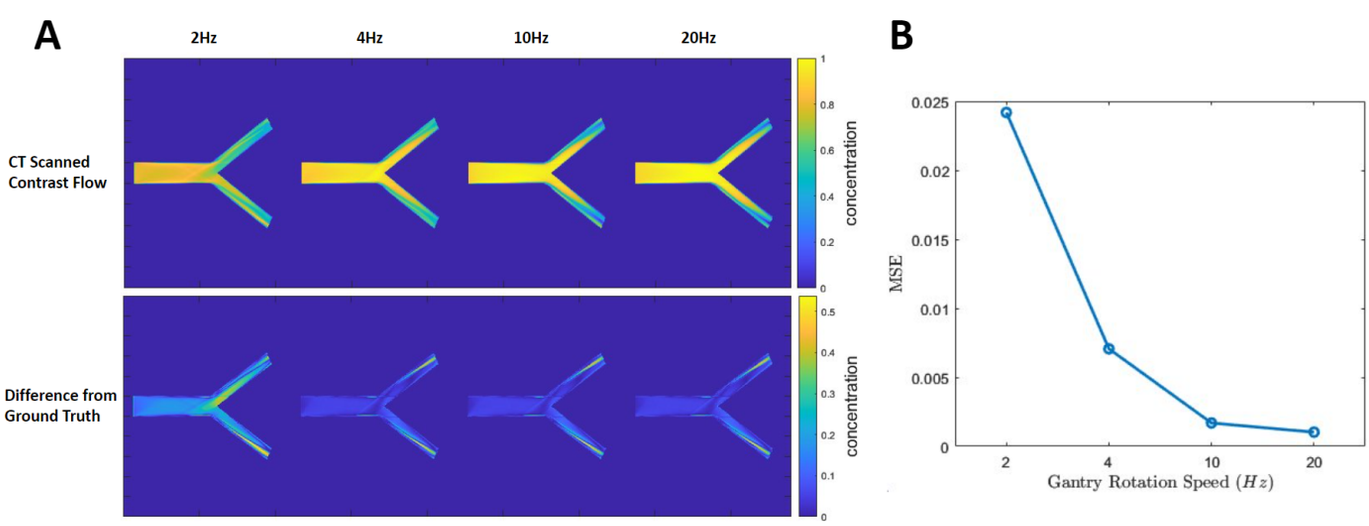


Fig. 2. The effect of CT scanning process on ground truth CFD data. Panel A shows 2, 4, 10 and 20 Hz CT scanned contrast flow and their absolute difference from ground truth. Panel B shows the mean square error calculated from CFD simulated and CT scanned flow data at different gantry rotation speed.

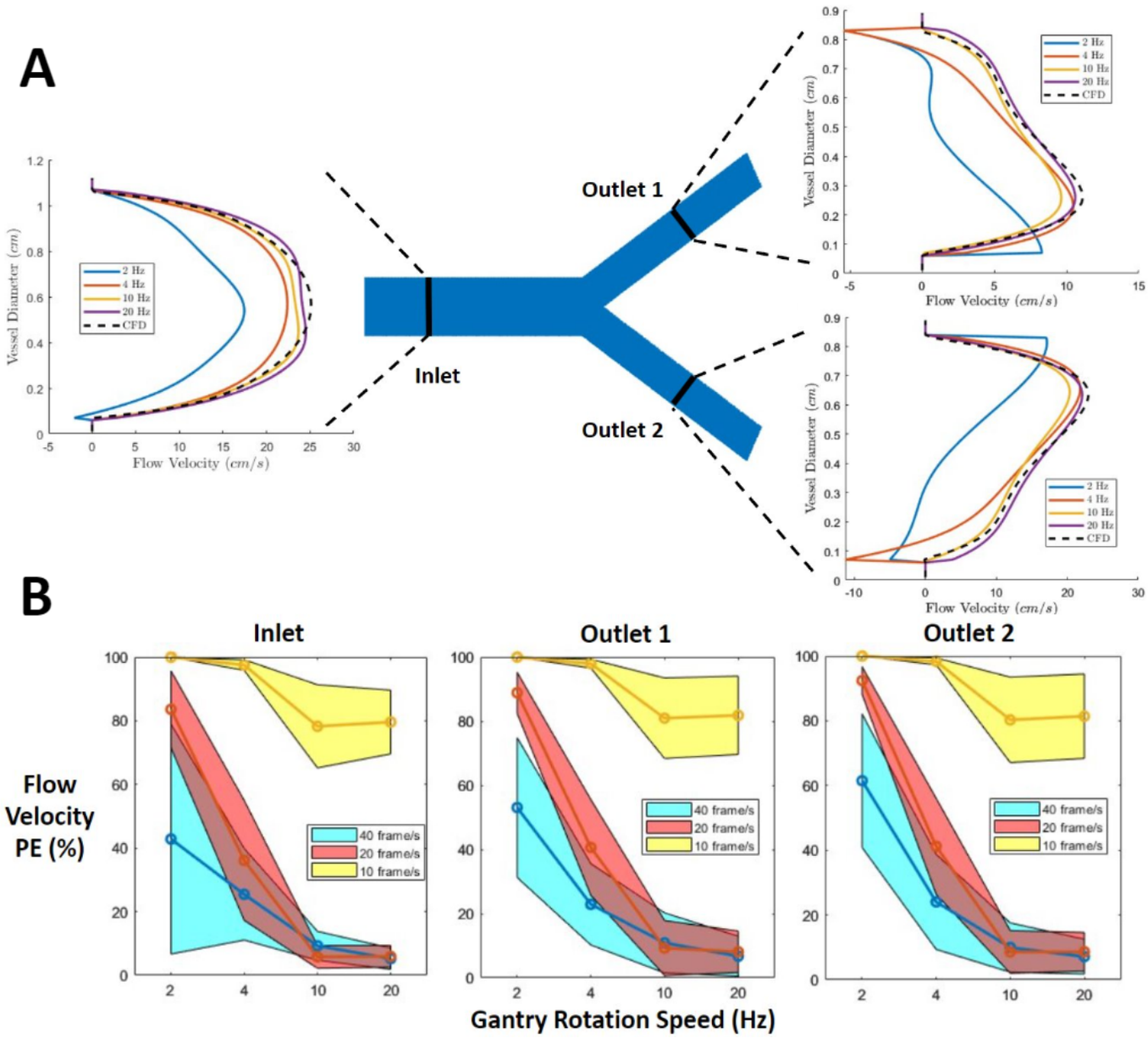


Fig. 3. PINN predicted flow velocity field and percentage error from ground truth. Panel A shows sample flow velocity profile predicted by PINN at the three checkpoints of the Y-shaped vessel. Panel B shows PINN predicted velocity percentage error under different CT gantry rotation speed and reconstruction frame rates.

movie of contrast evolution was scanned with a user-defined gantry rotation speed. We simulated imaging of $c_{CFD}(x, y, z, t)$ at 2, 4, 10 and 20 Hz gantry rotation speed. These values were chosen to mimic a slow rotation of a current scanner CT, high speed CT, dual-source CT scanning speed and evaluate the value of a next generation CT. This process generates CT scanned movies of the contrast $c_{CT}(x, y, z, t)$ at different gantry rotation speed shown in *Figure 2*. Each data point is the mean percent error averaged over all time frames and the shaded areas include one standard deviation up and below the means. As the gantry rotation speed decreases, there are more motion artifacts in the images. The images obtained with gantry speeds of 2 and 4 Hz exhibit noticeable motion artifacts, significantly compromising data fidelity and potentially transmitting these distortions to the final flow field estimation.

Reconstruction was performed via view-sharing with frame rates of 40, 20 and 10 frames per second to investigate the impact of frame rate on PINN-derived flow accuracy.

C. *Physics Informed Neural Network Flow Estimation*

A neural network was used to estimate flow from CT scanned flow data $c_{CT}(x, y, z, t)$ and to generate predicted flow velocity fields $[\hat{u}(x, y, z, t), \hat{v}(x, y, z, t)]$. CFD simulated and PINN predicted U, V velocity fields are shown in Figure 1.

This pipeline facilitates a direct comparison between the flow velocity fields produced by CFD and those estimated by the neural network. CFD flow simulation conducted in a Y-shaped vessel serves as the ground truth data in both Experiment 1 and 2. PINN prediction using CT scanned movies generate flow velocity fields $[\hat{u}(x, y, z, t), \hat{v}(x, y, z, t)]$. The flow velocity was assessed at three distinct checkpoints: one at the inlet and two

at each of the outlets. The location of the three checkpoints and sample flow velocity profiles predicted by PINN are shown in *Figure 3*. Flow accuracy was assesses using the mean value of the predicted velocity at the checkpoints and the percentage error (PE) with respect the ground truth velocity:

$$PE = (\hat{V} - V)/V$$

where $\hat{V}$ is PINN predicted mean velocity at a checkpoint, calculated from [$\hat{u}(x, y, z, t)$, $\hat{v}(x, y, z, t)$] and V is ground truth mean velocity from CFD at the same checkpoint calculated from [$u_{CFD}(x, y, z, t)$, $v_{CFD}(x, y, z, t)$]. Percentage errors exceeding the threshold of 10% were regarded as inaccurate predictions by PINN.

In Experiment 2, the effect of varying the frame rate of the contrast evolution movie was evaluated. This corresponds to the information richness and is determined by the reconstruction rate of the CT-scanned flow data. The experiment conducted in Experiment 1 is replicated at varying CT reconstruction frame rates, including 40, 20, and 10 frames per second (FPS), and the PINN predicted velocity PE at three checkpoints under different frame rates is assessed. Calculated percentage errors exceeding the threshold of 10% are regarded as inaccurate predictions by PINN.

III. RESULTS

CT scanning and reconstruction result in images c_{CT} that differ from the ground trurth contrast movies, c_{CFD}. *Fig 2* shows the impact of 2, 4, 10 and 20 Hz gantry rotations speed on the CT scanned data c_{CT} and the mean squared error (MSE) between c_{CT} and c_{CFD}. The MSE is low for CT scanning at 10 and 20 Hz. However, when the gantry rotation speed drops below 10 Hz, the MSE increases much faster.

The errors in c_{CT} due to the CT imaging process lead to errors in PINN flow field prediction. The percentage error at three checkpoints is reported in *Fig. 3A*. Percentage error below 10% was achieved by the PINN at all three checkpoints when the CT gantry rotation was 10 or 20 Hz. However, accuracy declined at slower gantry rotation speed. For 4 Hz, the percentage errors exceeded 20% (PE in the inlet, outlet 1, and outlet 2: 25.5%, 22.9%, and 24.1%). For 2Hz, values surpassed 40% (PE in the inlet, outlet 1, and outlet 2: 42.8%, 53.1%, and 61.5%).

The perfect error seen in flow estimates agrees with the error in c_{CT} as measured by MSE. Notably, the error in PINN's predictions escalates as the CT gantry rotation speed decreases, resulting in flow velocity percentage errors beyond the acceptable 10% threshold for 4 and 2 Hz CT scans.

The accuracy of PINN flow prediction also varied based on the reconstruction frame rate. The velocity percent error at three distinct checkpoints are presented in *Fig. 3B*. For 10 and 20 Hz gantry rotation, imaging with 20 or 40 FPS achieved <10% error. However, if the frame rate dropped to 10 FPS, then the error increased significantly (~80%). For 2 and 4 Hz gantry rotation, reduction in the reconstructed frame rate led to stepwise losses in performance.

In addition to GRS and frame rate, other factors likely influence the accuracy of PINN flow that were not accounted for in this work. For example, the size of the vessel, the complexity of the flow pattern, and the shape and duration of the contrast bolus could all be optimized to improve PINN-based flow estimation. This is left for future work.

IV. CONCLUSIONS

PINNs offer a powerful alternative method for estimating blood flow velocity fields from 4D CT scans. However, they demand a relatively high GRS and frame rate. Our findings indicate that utilizing low GRS CT scan results in corrupted flow movies, thereby lowering the accuracy of PINN flow reconstructions. Specifically, with the current CT scanner operating at 2-4 Hz GRS, the PINN fails to achieve ideal accuracy in flow field estimation.

Enhancements to the reconstruction pipeline as well as the PINN framework are essential to address these limitations and improve accuracy under current CT scanning conditions.

V. REFERENCES

[1] L. I. Heller, K. H. Silver, B. J. Villegas, S. J. Balcom, and B. H. Weiner, “Blood flow velocity in the right coronary artery: Assessment before and after angioplasty,” *J. Am. Coll. Cardiol.*, vol. 24, no. 4, pp. 1012–1017, Oct. 1994, doi: 10.1016/0735-1097(94)90863-X.

[2] E. E. Apostolakis and N. G. Baikoussis, “Methods of estimation of mitral valve regurgitation for the cardiac surgeon,” *J. Cardiothorac. Surg.*, vol. 4, no. 1, p. 34, Dec. 2009, doi: 10.1186/1749-8090-4-34.

[3] P. Dyverfeldt *et al.*, “4D flow cardiovascular magnetic resonance consensus statement,” *J. Cardiovasc. Magn. Reson.*, vol. 17, no. 1, p. 72, Dec. 2015, doi: 10.1186/s12968-015-0174-5.

[4] E. Soudah, E. Y. K. Ng, T. H. Loong, M. Bordone, U. Pua, and S. Narayanan, “CFD Modelling of Abdominal Aortic Aneurysm on Hemodynamic Loads Using a Realistic Geometry with CT,” *Comput. Math. Methods Med.*, vol. 2013, pp. 1–9, 2013, doi: 10.1155/2013/472564.

[5] D. A. Steinman *et al.*, “Variability of Computational Fluid Dynamics Solutions for Pressure and Flow in a Giant Aneurysm: The ASME 2012 Summer Bioengineering Conference CFD Challenge,” *J. Biomech. Eng.*, vol. 135, no. 2, p. 021016, Feb. 2013, doi: 10.1115/1.4023382.

[6] M. Raissi, A. Yazdani, and G. E. Karniadakis, “Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations,” *Science*, vol. 367, no. 6481, pp. 1026–1030, Feb. 2020, doi: 10.1126/science.aaw4741.

[7] 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, doi: 10.1016/j.compmedimag.2019.06.001.

[8] R. Shah *et al.*, “Effect of gantry rotation speed and scan mode on peristalsis motion artifact frequency and severity at abdominal CT,” *Abdom. Radiol.*, vol. 43, no. 9, pp. 2239–2245, Sep. 2018, doi: 10.1007/s00261-018-1497-0.

Estimation of Blood Flow Parameters in the Left Atrial Appendage from 4DCT Dynamic Contrast Enhancement

Lauren M. Severance, Andrew M. Kahn, Juan C. del Alamo, Elliot R. McVeigh

***Abstract*—Cardiac CT is often used clinically in electrophysiology to evaluate cardiac morphology. One such case is to evaluate patients with Atrial Fibrillation (AF). AF can cause regions of slow blood flow and blood stasis through the left atrial appendage (LAA), and therefore, it may be preferable to evaluate blood flow through the LAA in addition to morphology. Although CT cannot measure flow directly, CT data has been used to estimate flow using modeling approaches such as Computational Fluid Dynamics, which take into account the cardiac geometry to simulate flow.**

Advances in CT technology now enable high-resolution imaging of the whole heart with low radiation doses. With multi-heartbeat imaging during a contrast injection, we can obtain 4-dimentional CT (4DCT) images to measure dynamic contrast enhancement directly. In this study, we use high-resolution 4DCT to acquire images of contrast enhancement across the LAA over multiple heartbeats. The CT contrast signal at each voxel over time is used to create dynamic contrast enhancement maps of parameters derived from a gamma-variate fit. These contrast enhancement maps enable quantification and visualization of spatial-temporal characteristics of flow parameters across the LAA.

***Index Terms*—Cardiovascular CT, Dynamic Contrast Enhancement, Atrial Fibrillation, Left Atrial Appendage**

I. Introduction

CARDIAC CT has been used for many years in electrophysiology to plan cardiac ablation and evaluate morphology. Modern fast CT techniques give us the opportunity to also evaluate blood flow and blood stasis as a risk marker for thrombus [1]. These studies often use Computational Fluid Dynamics to estimate a blood flow field from the CT data, and one important clinical application is to evaluate patients with Atrial Fibrillation (AF) [2]. AF occurs when the heart's upper chambers beat irregularly, creating regions of slow blood flow and blood stasis through the left atrium (LA) and left atrial appendage (LAA), which is a sac that protrudes from the LA. Blood stasis may promote formation of a thrombus in the LAA, which can travel to the brain and cause a stroke. As a result, stroke risk is five times higher in patients with AF than in the general population, and the vast majority of atrial thrombi in patients with AF are located in the LAA [3].

While CT is not able to directly measure flow like transesophageal echocardiography or MRI [4], [5], CT can obtain high-resolution images of the whole heart at low radiation doses. Using multi-heartbeat imaging during a contrast injection, CT can provide high spatial resolution estimates of dynamic contrast agent concentration to directly evaluate blood flow and blood stasis. The purpose of this project was to create accurate local estimates of contrast agent concentration over time, as well as estimates of blood flow parameters, at high spatial resolution over the entire LAA.

II. Methods

A. Image Acquisition

4-dimensional CT (4DCT) images were acquired using a custom CT imaging protocol similar to baseline cardiac CT perfusion [6]. The contrast injection consisted of 80 ml of Iopamido, Isovue 370, Bracco, at 5 ml/sec with a 50 ml saline flush. Acquisitions were made with a cardiac gated CT "half-scan", 280 ms gantry rotation on the GE Revolution CT 256 slice scanner using a low-dose variant of the "dynamic perfusion protocol" acquisition. 26 128-slice volumes were collected every 1-3 heartbeats over 40-60 seconds, depending on patient heart rate, and a 27th volume was collected approximately 100 seconds after the first volume. Volumes were collected at 100 kVp and 100 mA. All timeframes were acquired using prospective gating during a 170 ms window around end-systole (35% R-R).

B. Image Reconstruction and Postprocessing

Images were reconstructed with a 512x512x128 matrix size. Voxel size was ~0.5 mm in the xy-plane and 1.25 mm in z. Images were reconstructed using the vendor provided deep learning based "DLIR-high" method. After reconstruction, all timeframes were aligned using nonrigid registration for motion

Research reported in this publication was supported by the NIH/NHLBI grant R01HL160024.

L.M.S. and E.R.M. are with the Shu Chien – Gene Lay Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093 USA (email: lseveran@ucsd.edu, emcveigh@ucsd.edu).

A.M.K. and E.R.M. are with the Division of Cardiovascular Medicine, University of California San Diego, La Jolla, CA 92093 USA (akahn@health.ucsd.edu).

E.R.M. is also with the Department of Radiology, University of California, San Diego, La Jolla, CA 92093 USA.

J.C.A. is with the Mechanical Engineering Department and the Department of Cardiology, University of Washington, Seattle, WA 98195 USA (email: juancar@uw.edu).

correction [7]. Multiplanar reconstruction was used to cut a slice through the LAA, producing a series of 27 2-dimensional images over a region of interest surrounding the LAA. Prior to fitting with the gamma-variate model, each timeframe was smoothed using a gaussian filter with a standard deviation of 1 mm, and voxels with a cumulative sum across time <100 Hounsfield units (HU) were excluded from fitting. A summary of the pipeline for image postprocessing and the generation of dynamic contrast enhancement maps is provided in Fig. 1.

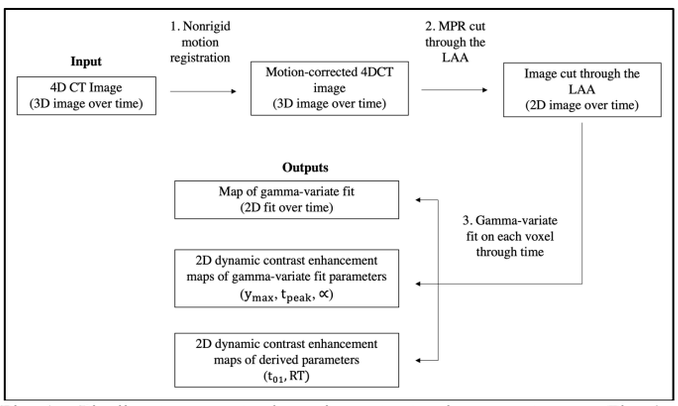


Fig. 1. Pipeline to generate dynamic contrast enhancement maps. The 4-dimensional CT (4DCT) image consists of 27 3D volumes, each collected during a 170 ms window around end-systole. In Step 1, the 27 timeframes are aligned using a nonrigid registration motion correction to produce registered 4DCT images. In Step 2, multiplanar reconstruction is used to create a 2-dimensional (2D) slice through the LAA, creating a set of 27 2D images of the LAA. In Step 3, the 1-dimensional signal of each pixel over time is fit using the gamma-variate model. After each pixel is fit, the parameters produced by the model can be used to create a map of the gamma-variate fit intensity over time (2D fit over time), dynamic contrast enhancement maps of parameters of the fit (y_{max}, t_{peak}, $\propto$), and dynamic contrast enhancement maps of derived parameters (t_{01}, RT).

C. Blood Transit Model Overview

Blood flow through the LAA was quantified using the gamma-variate model, which has previously been used to describe a bolus as it passes through a series of compartments [8], [9] and is widely accepted for fitting CT contrast concentration as a function of time [10], [11]. To apply the model to our 4DCT data, we used the gamma-variate formulation of Bindschadler et al [11] and Madsen et al [9]:

$$y(t) = y_{max}\left(\frac{t}{t_{peak}}\right)^{\propto} e^{\propto\left(1-\frac{t}{t_{peak}}\right)} - y_b \qquad (1)$$

where $y(t)$ is the CT image intensity as a function of time, y_{max}, t_{peak}, and $\propto$ are parameters of the fit, y_b is the measured CT signal baseline, and t is time.

D. Gamma-variate Model Fit and Calculation of Flow Parameters

For each image voxel, we fit the 1-dimensional (1D) signal across time using the gamma-variate equation to obtain a voxel-specific model of the 3 gamma-variate parameters (y_{max}, t_{peak}, and $\propto$). Then, using this voxel-specific model, we derived additional parameters to describe the blood flow, t_{01} (see Fig. 2) and residence time (RT) (see Eq. 2).

The CT image contrast $y(t)$ for each voxel was derived as the 1D voxel signal over time minus a baseline signal (y_b). The baseline signal was calculated as the mean of timepoints 1-3 in the descending aorta.

The 1D signal of CT image contrast over time was fit to the gamma-variate model using the MATLAB "fit" function (MathWorks Inc., Natick, Massachusetts, R2022a). Table 1 shows the lower and upper limits defined for each fit parameter.

TABLE I: Gamma-variate fit parameters limits

Parameter	Lower limit	Upper limit
$\propto$	0	Infinity
y_{max}	Maximum data value of CT image contrast - 40 HU	Maximum data value of CT image contrast + 40 HU
t_{peak}	Time of maximum data contrast minus 5 seconds	Time of maximum data contrast plus 5 seconds

After obtaining a voxel-specific model of the gamma-variate fit, t_{01} and RT were derived. t_{01} was calculated as the time at which the gamma-variate fit reaches 1% of its maximum value. RT was calculated according to the following equation:

$$RT = \frac{\int_0^\infty t * y(t)dt}{\int_0^\infty y(t)dt} \qquad (2)$$

where $y(t)$ is the analytical model for CT image contrast as a function of time, resulting from the gamma variate fit, and t is time. Fig. 2 shows the signal contrast over time, the corresponding gamma-variate model fit, and several model parameters for an example image voxel.

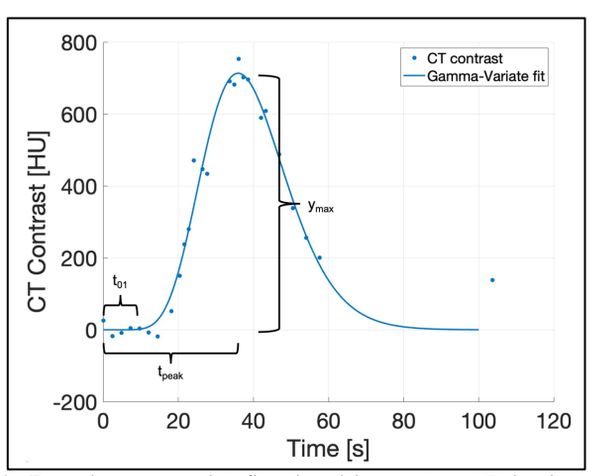


Fig. 2. Example gamma-variate fit and model parameters. CT signal contrast as a function of time (blue dots) is fit using the gamma-variate model (blue line). CT contrast is calculated as the CT image intensity at a given pixel minus a baseline signal. The baseline signal is calculated as the average CT signal at the first 3 timepoints in a region of interest in the descending aorta. The gamma-variate fit parameters t_{peak} and y_{max} are direct outputs of the model. t_{01} is a derived parameter and is defined as the time at which the signal reaches 1% of its peak value.

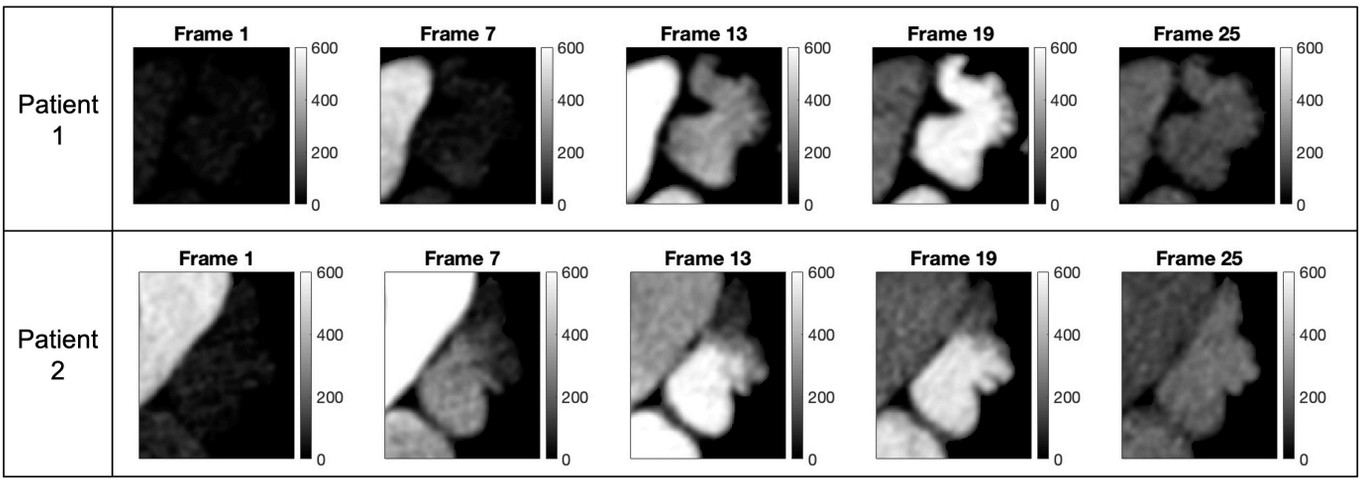


Fig 3. CT images of contrast enhancement over time for 2 patients. The post-processed images enable detailed visualization of contrast enhancement over time, and the trajectory of contrast enhancement can be evaluated across the LAA. For each patient, 5 example timeframes are shown from a dataset containing 27 timeframes.

III. Results

A. Contrast Enhancement Images

High-resolution CT images were acquired at 27 timeframes. After motion registration and multiplanar reconstruction, the images were used to visualize contrast enhancement across the LAA over time. These images were then used as input to the gamma-variate model to calculate parameters of the flow at each image voxel. Fig. 3 shows the contrast enhancement images for 2 patients at 5 example timeframes. In frame 1, the LAA appears dark due to the lack of contrast. Over the course of imaging, the signal within much of the LAA increases up to a maximum intensity, as the contrast enters the LAA. After reaching a maximum, the signal decreases as contrast flows out of the LAA. Across 17 patients in the clinic who underwent imaging with this unique protocol, the mean dose length product (DLP) was 8.28±1.05 mSv.

B. Dynamic Contrast Enhancement Maps

Parameters derived from the gamma-variate fit were used to create 2D dynamic contrast enhancement maps. Each dynamic contrast enhancement map provides a spatially smooth map of a model parameter and can be used to visualize blood flow patterns across the LAA. Fig 4. shows the dynamic contrast enhancement map of RT for two clinical cases. In Patient 1, the RT calculated from the gamma-variate fit is similar for pixels in the proximal, middle, and distal regions of the LAA, and the map of RT appears uniform. In Patient 2, the contrast enhancement of pixels in the middle and distal regions is delayed and reaches a lower peak than pixels in the proximal region, resulting in increased RT in the middle and distal regions. The map of RT across the LAA for Patient 2 displays a delayed filling pattern.

IV. Discussion

We have evaluated parameters of the blood flow across the LAA using dynamic CT contrast enhancement. With fast, high-resolution CT, we can create detailed enhancement maps across the LAA to characterize parameters of the flow. In this study, we found it was sufficient to use the data collected at 27 timepoints at a relatively low dose (approximately 100 mA x-ray fluence) to create smooth enhancement maps. Although the signal at a given voxel may be noisy over time, fitting with the gamma-variate model provides a smooth estimate of the contrast enhancement over time.

Among patients with AF who underwent imaging with this protocol in our clinic, we have observed a large dynamic range of RT and y_{max} values across the LAA with highly variable spatial patterns. It is clear that a significant number of patients who are in sustained AF have virtually no difference in flow parameters between the LA and the distal LAA; this indicates these patients are likely not at higher risk for thrombus generation in the LAA.

In a small cohort of patients, the mean DLP was comparable to other cardiac exams, including cardiac CT angiography. To further limit radiation exposure, future studies are warranted to evaluate the feasibility of using fewer timeframes to create dynamic contrast enhancement maps. The exact number and timing of images needed to obtain reasonable estimates of fit parameters remains an open question. We are currently developing a machine learning network that will allow us to compute dynamic contrast enhancement of maps the full 3D volume of the LA and LAA in a reasonable amount of time. Future optimizations will include the number of images required and the sampling frequency, minimum z-coverage required to capture the LAA during breathing, optimal injection profile for contrast agent for this application, and optimal spatial resolution/patient dose trade-off.

Conclusion

Using a new dynamic contrast CT protocol, we have created a technique to look at blood flow dynamics in the LAA in patients with AF. These data provide a potential method to evaluate the risk of thrombus formation in the LAA and probability of stroke.

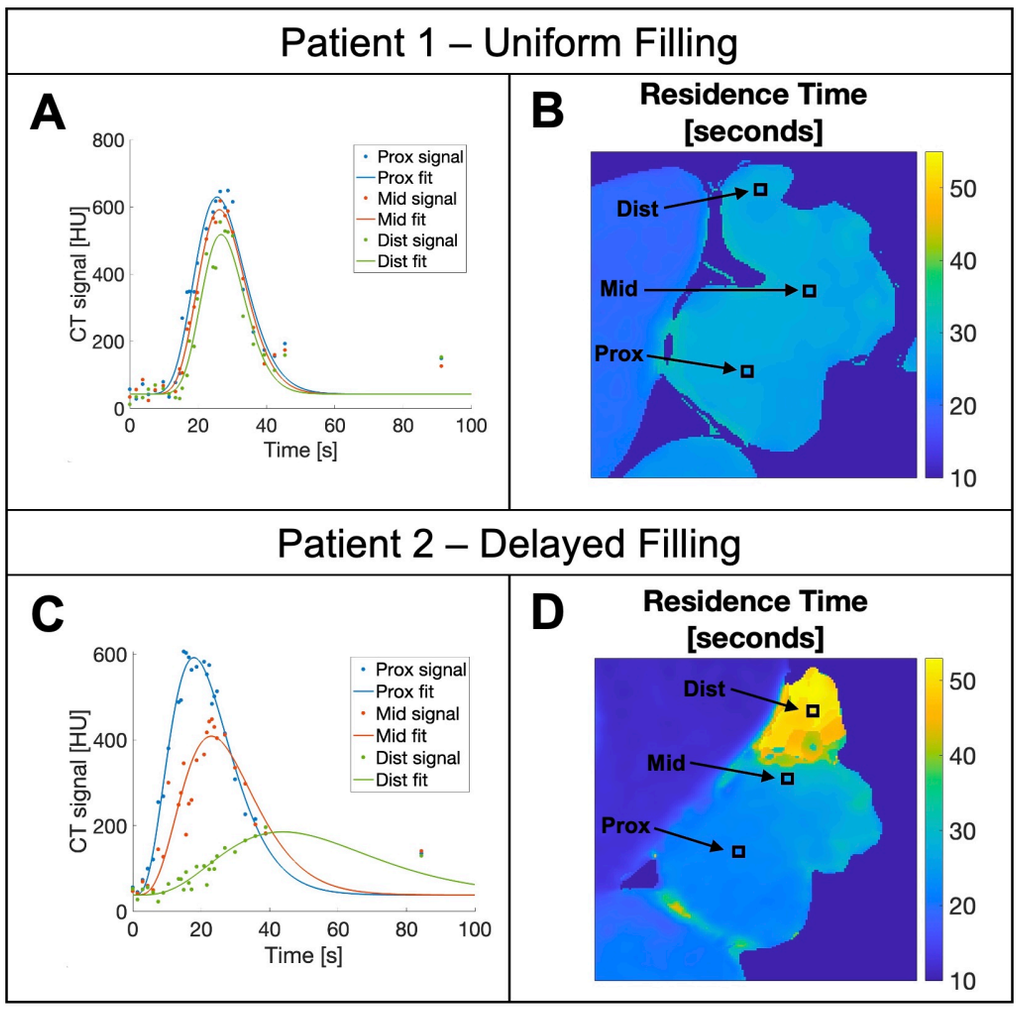


Fig. 4. Maps of residence time (RT) generated from the gamma-variate fit for a slice through the left atrial appendage (LAA). (A and C) CT signal intensity over time and corresponding gamma-variate fit for a single voxel in the proximal region (Prox signal, Prox fit), middle region (Mid signal, Mid fit), and distal region (Dist signal, Dist fit) of the left atrial appendage (LAA). (B and D) The gamma-variate fit from each pixel can be used to create a smooth map, or dynamic contrast enhancement map, of each flow parameter, including blood residence time (RT). In some patients, (Patient 1), the gamma-variate fits of voxels in the proximal, middle, and distal LAA overlap, and the dynamic contrast enhancement map of RT appears uniform across the LAA (B). In other patients, (Patient 2), the gamma-variate fit in voxels of the proximal, middle, and distal LAA does not overlap (C), and the dynamic contrast enhancement map of RT indicates delayed filling in the middle and distal regions of the LAA (D).

References

[1] G. M. Bosi, A. Cook, R. Rai, L. J. Menezes, S. Schievano, R. Tori, G. Burriesci, "Computational Fluid Dynamic Analysis of the Left Atrial Appendage to Predict Thrombosis Risk," *Front. Cardiovasc. Med.*, vol. 5, Apr. 2018, DOI: https://doi.org/10.3389/fcvm.2018.00034.

[2] M. Garcia-Villalba, L. Rossini, A. Gonzalo, D. Vigneault, P. Martinez-Legazpi, E. Duran, O. Flores, J. Bermejo, E. McVeigh, A.M. Kahn, J.C. del Alamo, "Demonstration of Patient-Specific Simulations to Assess Left Atrial Appendage Thrombogenesis Risk," *Front. Physiol.*, vol.12, Feb. 2021, DOI: https://doi.org/10.3389/fphys.2021.596596.

[3] J.L. Blackshear and J.A. Odell, "Appendage Obliteration to Reduce Stroke in Cardiac Surgical Patients with Atrial Fibrillation," *Ann. Thorac. Surg.*, vol. 61, no. 2, pp. 755-759, Feb. 1996. DOI: 10.1016/0003-4975(95)00887-X.

[4] P.M. Verhorst, O. Kamp, C.A. Visser, and F.W. Verheugt, "Left atrial appendage flow velocity assessment using transesophageal echocardiography in nonrheumatic atrial fibrillation and systemic embolism," *Am. J. Cardiol.*, vol. 71, no. 2, pp. 192-196, Jan. 1993. DOI: 10.1016/0002-9149(93)90737-w.

[5] M. Markl, D.C. Lee, N. Furiasse, M. Carr, C. Foucar, J. Ng, J. Carr, and J.J. Goldberger, "Left Atrial and Left Atrial Appendage 4D Blood Flow Dynamics in Atrial Fibrillation," *Circ. Cardiovasc. Imaging*, vol. 9, no. 9, Sep. 2016. DOI: 10.1161/CIRCIMAGING.116.004984.

[6] K. Nieman, S. Balla, "Dynamic CT myocardial perfusion imaging," *J. Cardiovasc. Comput. Tomogr.*, vol. 14, no. 4, pp. 303-306, Jul.-Aug. 2020. DOI: 10.1016/j.jcct.2019.09.003.

[7] S. Thiruvenkadam, K.S. Shriram, B. Patil, G. Nicholas, M. Teisseire, C. Cardon, J. Knoplioch, N. Subramanian, S. Kaushik, R. Mullick, "GPNLPerf: Robust 4d Non-rigid Motion Correction for Myocardial Perfusion Analysis," *Proc. of MICCAI 2016,* pp. 255-263, Oct. 2016, DOI: https://doi.org/10.1007/978-3-319-46726-9_30.

[8] H.K. Thompson, C.F. Starmer, R.E. Whalen, H.D. McIntosh, "Indicator Transit Time Considered as a Gamma Variate," *Circ. Res.*, vol. 14, no. 6, Jun. 1964. DOI: 10.1161/01.res.14.6.502.

[9] M.T. Madsen, "A simplified formulation of the gamma variate function," *Phys. Med. Biol.*, vol. 37, no. 7, pp. 1597-1600, Feb. 1992. DOI: 10.1088/0031-9155/37/7/010.

[10] Z.M. Lin S. Pohlman, A. Cook, S. Chandra, "CT perfusion: comparison of gamma-variate fit and deconvolution," *Proc. SPIE 4683*, pp. 102-109, April 2002, DOI: 10.1117/12.463572

[11] M. Bindschadler, D. Modgil, K. R. Branch, P.J. La Riviere, A.M. Alessio, "Comparison of Blood Flow Models and Acquisitions for Quantitative Myocardial Perfusion Estimation from Dynamic CT," *Phys. Med. Biol.*, vol. 59, no. 7, pp. 1533-1556. DOI: 10.1088/0031-9155/59/7/1533.

INFACT: Generalizable INR with geometry-aware Feature Aggregation for CT reconstruction

Muge Du, Yinong Liu, and Yuxiang Xing*

***Abstract*—Implicit neural representation stands as a novel data representation approach, demonstrating remarkable proficiency in several CT reconstruction tasks. However, existing CT INR methods face challenges: long training time, lack of generalizability, and detail loss. The core issue could lie in the difficulty of INR to learn all details from scratch. We address the above three challenges by integrating data-driven detailed features extracted from geometry knowledge, projection signals and prior image, reforming the objective to only low-frequency content filling and artifact correction. The proposed INR method can directly generalize to unseen cases and allow efficient fine-tuning. We validated the proposed method on a CL (computed laminography) scan, showing generalizing ability to unseen cases and superiority over trained-from-scratch CT INR. With improved training efficiency and reconstruction quality, our method is expected to facilitate the application of INRs in CT.**

***Index Terms*—Computed Tomography, Implicit Neural Representation, Deep Learning, Image Reconstruction, Computed Laminography.**

I. Introduction

IMPLICIT Neural Representation (INR) [1] is a novel data representation method that utilizes coordinate encoding and neural networks to map coordinates to values. INRs can fit highly complex data and performed well in single-case reconstruction in visible light [1], CT [2] and other imaging modalities [3]. For CT reconstruction, several INR approaches have been proposed for ill-posed reconstruction tasks like sparse-view [2], limited-angle [4], metal artifact reduction [5], and dynamic 4D-CT reconstruction [6]. However, most existing CT INR methods exhibit three primary characteristics: Firstly, they must be re-trained for each new case rather than generalized across scenes. Secondly, they are trained from scratch to learn all the details and unable to incorporate existing prior knowledge. Thirdly, for 3D imaging tasks, they cannot be trained on all rays in each iteration due to GPU limitations, but leverage batches of random rays instead during training. Theoretically, INRs must be trained at least once on each training projection ray to fully utilize the projection data, which leads to time-consuming training. Even with state-of-the-art speed-up method like learnable hashgrid encoding [7], training CT INR still requires minutes to hours for high-resolution 3D CT, hence limits their practical usage. As the computation time per iteration is hard to be shorten given the same model capability, the key to speed-up CT INR training is reducing the total iterations, meaning only a subset of the total projection rays are used during training. Unfortunately, we observed that CT INR methods learn from the rays in a coarse-to-fine ways, as they quickly capture rough structures during early iterations but the speed to add details from later training is slow. In practice, even CT INRs trained on full projections may lose certain details, not to mention INRs trained on a projection subset. In summary, existing CT INRs trained from scratch are inefficient, not generalizable, and prone to detail loss.

The core issue could lie in the difficulty of INR to reconstruct all details from scratch. To address this problem, we propose a solution by incorporating data-driven features containing detailed prior information into INR. This transforms the goal of INR from reconstructing all information to easier tasks like low-frequency content filling and artifact correction by referring to prior details. The proposed feature-integrated INR can directly generalize to unseen cases, and only a few more steps of fine-tuning are needed to further improve the reconstruction. We evaluated the method on a CL (computed laminography) scan system. Its performance in ill-posed reconstruction for unseen cases is demonstrated.

Submitted date: Jan 23, 2024. This work was supported by the National Key R&D Program of China (No. 2022YFF0607802).

Muge Du, Yinong Liu and Yuxiang Xing are with the Department of Engineering Physics, Tsinghua University, Beijing, China (e-mail: xingyx@mail.tsinghua.edu.cn).

II. Methods

A. Overall framework

In this work, we propose a novel CT INR framework called INFACT (INR with geometry-aware Feature Aggregation method for CT reconstruction), which takes advantage of both flexible per-scene image reconstruction by INR and generalizable feature extraction by data-driven pre-trained networks, resulting in a framework combining network pre-training (in Fig. 1) and INR-finetuning (in Fig. 2).

The network part consists of six parts, illustrated in Fig. 1: 1. Projection Feature Extraction block to extract projection features from each view's 2D detector signals. 2. Feature Back-projection block to complete projection-to-volume transformation. 3. Feature Aggregation block to extract the common projection features from all the rays crossing the same

voxel. 4. Prior Feature Extraction block to extract detailed information from a prior image. 5. Feature Fusion block to combine and concentrate all obtained features and geometrical embeddings within neighboring voxels. 6. Value Prediction block to compute voxel value by an MLP.

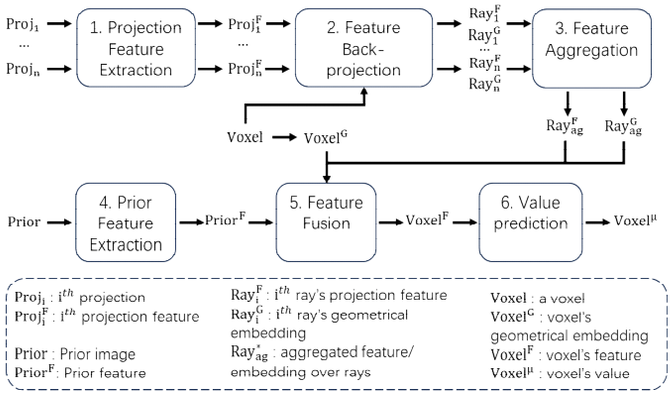


Fig. 1. The network part of the proposed INFACT framework.

We describe the design of key blocks as in the following:

1. Projection Feature Extraction: Before feature extraction, we divide each ray's projection value by the ray's intersected length within the volume space, thus the projection value becomes the average attenuation over the ray and being more geometry-invariant. For feature extraction from each view's projection, we use a 2D CNN with multi-scale feature output, allowing more adaptation to varying resolutions.

2. Feature Back-projection: For a single voxel intersected by a set of rays, each ray's projection feature is interpolated from corresponding view's feature, analogizing back-projection. Besides, we compute geometrical embedding for each ray and each voxel through two learnable hashgrid encoders. Each ray's polar vector is embedded, so is each voxel's position.

3. Feature Aggregation: For a single voxel and its intersected rays, an 1D CNN is used to aggregate along the view dimension. Each ray's projection feature and geometrical embedding are concatenated in channel dimension. Besides, we sort the rays according to their polar azimuth angles before concatenation, thus introducing some geometry-invariance.

4. Prior Feature Extraction: We use a 3D U-Net to extract features from the prior image. Instead of inputting the original image, we use its edge features extracted from gradient operators, thus partially reducing the impact of artifacts in ill-posed geometries.

5. Feature Fusion: We use a 3D U-Net to fuse each voxel's projection features, geometrical embeddings and prior features. All the features are concatenated along channels as CNN input.

B. Fine-tuning INR with pre-trained INFACT network

After pre-training INFACT network, the method is already generalizable to unseen cases. To further refine the result, we introduce an INR fine-tuning method. The comparison between a common trained-from-scratch INR and the proposed INR fine-tuning are illustrated in Fig. 2-a and Fig. 2-b. In Fig. 2-a, the commonly used trained-from-scratch INR consists of a hashgrid encoder and an MLP that successively maps a voxel's coordinates into features and attenuation values. It has to be re-trained for each new scene. In Fig. 2-b, before fine-tuning, we first use the pre-trained INFACT's blocks 1~5 to inference the voxel features on the new scene, then freeze blocks 1~5 and the inferenced features. During fine-tuning, INR's linear layer is designed to project INR's hashgrid encoder into the same channel dimension as the first MLP Layer in Block 6, and being added together. In this way, INR's linear layer and the Block 6's first MLP Layer forms a larger layer. The INR and the whole Block 6 are fine-tuned together.

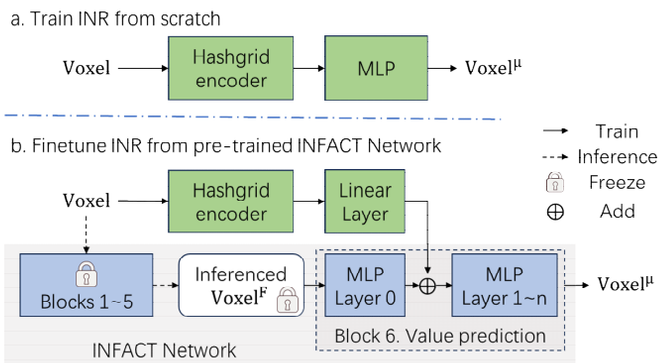


Fig. 2. Illustration of (a) common trained-from-scratch INR, and (b) the proposed INR fine-tuning from pre-trained INFACT network.

C. Network training

When pre-training INFACT, we use patch-based training. For each iteration, we randomly select eight 32*32*32 voxel patches from a volume sample for training. Due to the GPU memory limits, it is not feasible to directly aggregate too many views at once during training nor inference. Therefore, we fix the number of input views to 16 for INFACT. In training, we sample 16 sparse views each time. During inference, we divide the total views into several groups, each containing 16 sparse views. We process each group individually using Block 1~3, and then average the results after Block 3 to form the input to Block 4~6. To notice, the prior image is FDK reconstruction from all views, instead of from distinct groups. All the experiments are conducted on a single RTX 3090 GPU.

III. Experimental Study Results

A. Scanning geometry and dataset

As a preliminary validation of the method, we explored a rotary computed laminography (CL) scanning geometry, commonly employed for imaging plate-like objects such as printed circuit boards (PCBs). Rotary CL offers good internal structures but faces challenges of severe superimposed artifacts. In this experiment, we evaluate the method's effectiveness in addressing superimposed artifacts. The rotary CL geometry is illustrated in Fig. 3. The projection data contains 128 views, each 640×640, and the reconstructed volume is 256×256×52. More results demonstrating adaptation to various scanning geometries will be shown in the future presentation.

For the imaging task of detecting soldering defects on PCBs, we build a simple low-fidelity 'solders and blocks' simulated dataset for training and obtain a high-fidelity simulated PCB data for testing.

Low-fidelity 'solders and blocks' simulated dataset: We use solders cropped from existing PCB volume data, and random

blocks mimicking the planer structures to compose a simple approximation to PCB. For each training sample, we apply random transformation including rotation, resizing, flipping and grayscale perturbation to multiple solders and blocks, and randomly put them into an empty volume, creating multi-layer PCB-like data. A test data is also simulated to evaluate the methods' in-domain performance.

High-fidelity simulated PCB data: We utilize a CT-reconstructed PCB volume data, down-sample and crop it to be 256×256×52 for testing. This data is to evaluate the methods' out-of-domain performance.

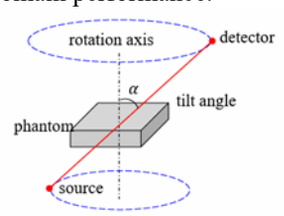


Fig. 3. Rotary CL geometry used for preliminary validation.

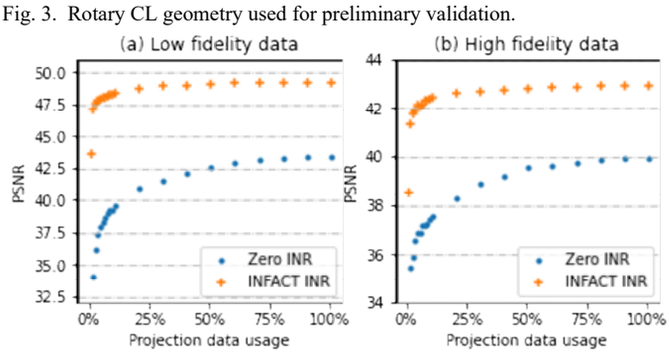


Fig. 4. Comparison of convergence between INFACT INR and Zero INR.

B. Experimental results

We conducted a comparative analysis between INR trained from scratch (Zero INR) and the proposed INR fine-tuned from INFACT (INFACT INR). Additionally, we provide the reconstruction results by FDK and pre-trained INFACT network part without any fine-tuning (INFACT Net).

Fig. 4 compares Zero INR and INFACT INR's performance growing curve during training. The x axis means the average usage of the projection data during training. The first point of INFACT INR's curve means method before fine-tuning, which is INFACT Net. From Fig. 4-a, on the low fidelity data, INFACT Net is comparable to Zero INR trained on 100% projection data, demonstrating the pre-trained INFACT's in-domain generalization capabilities. After fine-tuning on only 1% projection data, INFACT INR outperformed INFACT Net and Zero INR by a large margin, and converged at about 20%, showing the efficiency of the proposed fine-tuning method. For comparison, Zero INR still not converges after using 100% data. The convergence curve in Fig. 4-b is similar to Fig. 4-a, except for the relatively lower INFACT Net's performance, showing it facing out-of-domain challenges. However, after fine-tuning on only 1% projection data, INFACT INR clearly improvement the performance, showing the effectiveness of the proposed fine-tuning method on out-of-domain generalization.

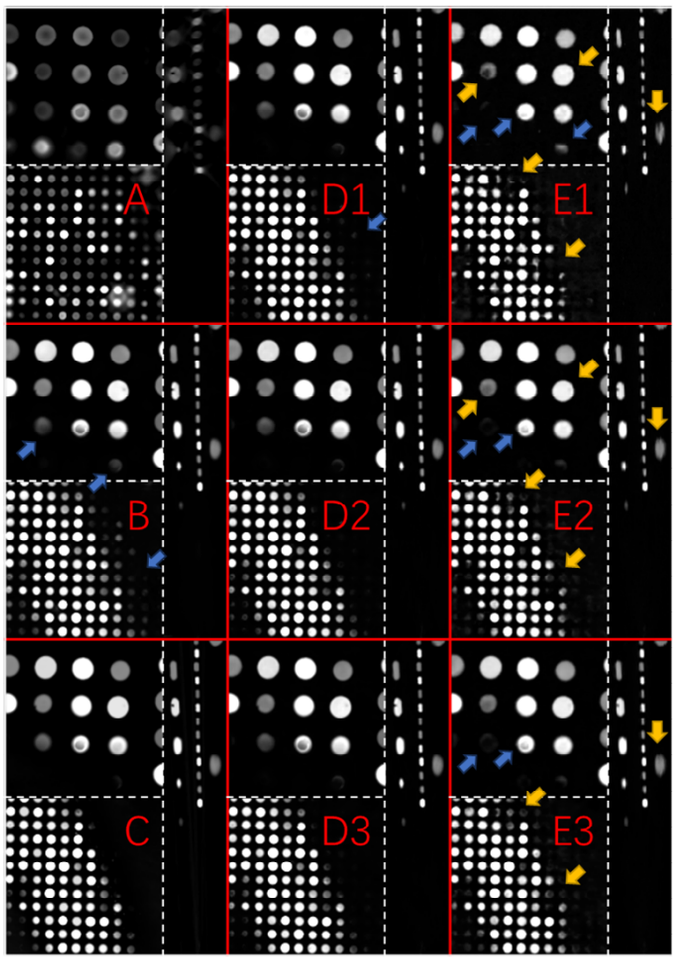


Fig. 5. Comparison between different methods' reconstruction results on the low-fidelity simulated data. A: FDK, B: INFACT Net, C: Ground Truth, D1~D3: INFACT INR trained using 1%, 5%, 10% projection data. E1~E3: Zero INR trained using 10%, 50% and 100% projection data. For each method, the left two images are transverse plane, the right one image is sagittal plane.

Fig. 5 and Fig. 6 shows the example reconstruction of the low-fidelity and high-fidelity data by each method. For INFACT INR (D1~D3), we show the result after fine-tuning on 1%, 5% and 10% data. For Zero INR (E1~E3), we show the result after training on 10%, 50% and 100% data.

In Fig. 5, FDK shows severe superimposed artifacts. INFACT Net is very close to the ground-truth, preserving all the detailed structures, but there are a few value-shifted solders, labeled in blue arrows. INFACT INR fine-tuned on only 1% data can correct most artifacts, and the result becomes better during training. Zero INR exhibits both structural changes (in yellow arrows) and value-shifting (in blue arrows), demonstrating its weaker ability to learn fine details and reducing ill-posed artifacts, compared to the proposed method.

In Fig. 6, reconstruction by FDK shows severe value-shifting and superimposed artifacts, but the soldering defects are preserved. INFACT Net provides high-quality reconstruction with almost no structural loss, but it changes some mist-like structures (mainly artifacts in ground truth image). INFACT INR gradually reduces the mist-like wrong structures during fine-tuning. Zero INR faces severe structural loss (in yellow arrows) and value-shifting even after training on 100% data.

As conclusion, the proposed framework performs well in

ill-posed CT reconstruction like CL. It solves the detail loss and inefficiency problem in trained-from-scratch INR, allowing fast generalize to unseen cases.

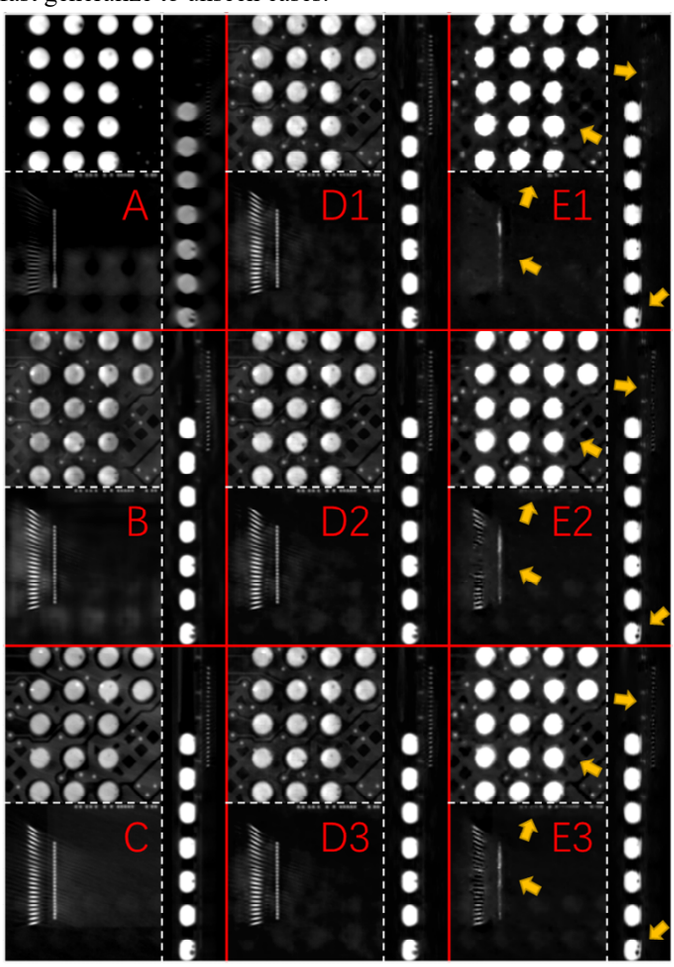


Fig. 6. Comparison between different methods' reconstruction results on the high-fidelity simulated data. A: FDK, B: INFACT Net, C: Ground Truth, D1~D3: INFACT INR trained using 1%, 5%, 10% projection data. E1~E3: Zero INR trained using 10%, 50% and 100% projection data. For each method, the left two images are transverse plane, the right one image is sagittal plane.

IV. Conclusion

In this study, we addressed the challenges of inefficiency and difficulty in recovering all details when training INR from scratch in CT reconstruction. We proposed a novel framework, called INFACT, which contains a pre-trained network suitable for various scanning geometries, and an INR fine-tuning method, enhancing its versatility for application in diverse scenarios. To build a geometry-generalizable network, we utilize several geometric-invariant and geometric-aware block designs, including projection feature extraction, feature back-projection, feature aggregation, prior feature extraction, feature fusion and final value prediction. To alleviate the challenge in learning details, we integrate detailed information from prior images into the method, thus transforming the goal of INR into easier tasks like low-frequency information filling and ill-posed artifact reduction. INFACT allows fast convergence with fine detail recovery. Preliminary experiments on PCB data and CL geometry indicate that the proposed pre-trained network itself exhibits good capabilities in solving ill-posed reconstruction problem and shows a certain degree of out-of-domain generalization. The proposed INR framework significantly reduces training steps and improves the details notably, which is expected to facilitate the application of INRs in CT.

References

[1] 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," Mar. 2020.

[2] 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, Cham, 2022, pp. 442–452.

[3] J. Xu, D. Moyer, B. Gagoski, J. E. Iglesias, P. E. Grant, P. Golland, and E. Adalsteinsson, "NeSVoR: Implicit Neural Representation for Slice-to-Volume Reconstruction in MRI," IEEE Transactions on Medical Imaging, vol. 42, no. 6, pp. 1707–1719, Jun. 2023.

[4] G. Zang, R. Idoughi, R. Li, P. Wonka, and W. Heidrich, "IntraTomo: Self-supervised Learning-based Tomography via Sinogram Synthesis and Prediction," in 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 2021, pp. 1940–1950.

[5] Q. Wu, L. Chen, C. Wang, H. Wei, S. K. Zhou, J. Yu, and Y. Zhang, "Unsupervised Polychromatic Neural Representation for CT Metal Artifact Reduction." arXiv, 27-Jun-2023.

[6] Y. Zhang, H.-C. Shao, T. Pan, and T. Mengke, "Dynamic cone-beam CT reconstruction using spatial and temporal implicit neural representation learning (STINR)," Phys. Med. Biol., Jan. 2023.

[7] 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, pp. 1–15, Jul. 2022.

Efficient simulation of biological tissue variability using Midgley's decomposition

Martina Nassi, Mikhail Mikerov, Alessandra Tomal, Koen Michielsen, and Ioannis Sechopoulos

***Abstract*—The aim of this study is to develop an accurate and computationally efficient method for introducing inter- and intra-patient tissue biological variability in x-ray imaging simulations. When considering the actual inhomogeneities of tissues, the varying elemental composition needs to be individually specified for each voxel, leading to more complex and computationally demanding simulations. In this study, we propose Midgley's model as an alternative approach. Through this parameterization we can model tissue variability using only five coefficients, reducing simulations' computational demand. As a use case, we present and evaluate this approach using the simulation of variable fibro-glandular tissue for digital breast tomosynthesis. Our preliminary findings demonstrated that Midgley's parameterization results in accurate simulations of fibro-glandular tissue variability, resulting in a median absolute percentage error of 0.0021% compared to the results of conventional simulations with indexed materials. Therefore, the proposed approach proved to be able to increase imaging simulation complexity without excessive computational demands.**

***Index Terms*—tissue variability, attenuation model, image simulation, x-ray imaging, tomosynthesis.**

I. INTRODUCTION

The advancement of x-ray imaging systems needs the use of digital phantoms for development, optimization, and evaluation purposes. Although homogeneous models have been used for decades for this purpose, their main drawback is the inability to reflect the actual inhomogeneous properties of tissues. This limitation leads to simplified phantoms and may introduce biases in the results. Including inter- and intra-patient tissue variability within phantoms offers a potential solution for improving the accuracy of simulations of medical imaging systems, since the interaction of tissues with radiation is mimicked more realistically [1, 2]. However, the inclusion of tissue variability within phantoms poses a significant computational challenge in simulations since each voxel needs to be assigned a unique value.

M. Nassi, M. Mikerov, K. Michielsen, and I. Sechopoulos are with the Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands.

A. Tomal is with the Universidade Estadual de Campinas - Instituto de Física Gleb Wataghin. Campinas - SP - Brazil. 13083-859

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 challenge, we propose Midgley's parameterization [3] as an efficient alternative for simulating tissue variability. In his work, Midgley developed a novel parameterization of x-ray interaction cross-sections to describe the x-ray linear attenuation coefficient for mixtures. This parameterization scheme addressed the Z-dependence of elemental cross-sections, accurately describing them for the entire periodic table using not more than five coefficients [3]. Therefore, this approach overcomes the limitations of simulation methods derived from basis material indexing, as it allows to accommodate arbitrary sample compositions that lie within the broad range of validity of the parameterization, using a finite number of coefficients.

This study aims to investigate the potential of Midgley's attenuation model to parameterize tissue variability using only five coefficients, resulting in more efficient and less computationally expensive simulations. While this study focuses on simulating fibro-glandular tissue variability in Digital Breast Tomosynthesis (DBT), it serves as an illustrative case, showcasing the potential of the proposed approach to simulate the variability of all tissue types in different x-ray imaging systems.

II. MATERIALS AND METHODS

A. Midgley's parameterization

The parameterization proposed by Midgley [3] decomposes the linear attenuation coefficient into energy-dependent S-parameters, each weighted by composition-dependent M-parameters. This decomposition is mathematically expressed as follows:

$$\mu(Z,E) = \sum_{k=1}^{N} S(E,k)\, M(Z,k). \qquad (1)$$

Here, the mixture parameters, denoted as $M(Z,k)$, are functions of the atomic number Z and particle density $n(Z)$, as expressed in equation (2). Specifically, the first mixture parameter $\sum n(Z)Z$ represents the electron density, and the remaining parameters are the higher order statistical moments describing the elemental distribution in the sample.

$$M(Z,k) = \sum_{Z} n(Z)Z^{k} \qquad (2)$$

The number of S-parameters, denoted as N, is not fixed; increasing the number of parameters improves model accuracy. However, for diagnostic x-ray energies, Midgley demonstrated that five S-parameters resulted in a maximum error below 2% at energies above the K-edge (from 6 keV to 125 MeV) [3].

The S-parameters were determined by solving a least-squares problem for a parameterization within the defined energy and compositional range of interest. Specifically, our investigation was delimited to an energy range between 10 keV and 49 keV [4] and the compositional range found in breast tissues, consisting of elements H, C, N, O, Na, Al, Si, P, S, K and Ca. Attenuation profiles for the elements of interest were obtained from XrayDB [5, 6].

B. Phantoms

A dataset of 10 digital compressed breast phantoms [7], with thickness ranging from 30 mm to 82 mm, was employed in this study. These phantoms consisted of three tissue types: skin, adipose, and fibro-glandular tissue. The elemental compositions of these tissues were extracted from the tabulations provided by Hammerstein et al. [8].

In addition to simply replacing the tissues with their Midgley components, we also accounted for fibro-glandular tissue variability. To consider variability in fibro-glandular tissue composition, we inserted four 2 cm cubes with different compositions into the compressed breast phantoms. For this purpose, the reference glandular attenuation curve was varied by ±5% [1] to accommodate different tissue densities, and its steepness was adjusted to generate two additional curves: one starting at the beginning point of the curve scaled by a factor of 1.05 and terminating at the endpoint of the curve scaled by 0.95, and vice versa for the second curve (as illustrated in Fig. 1).

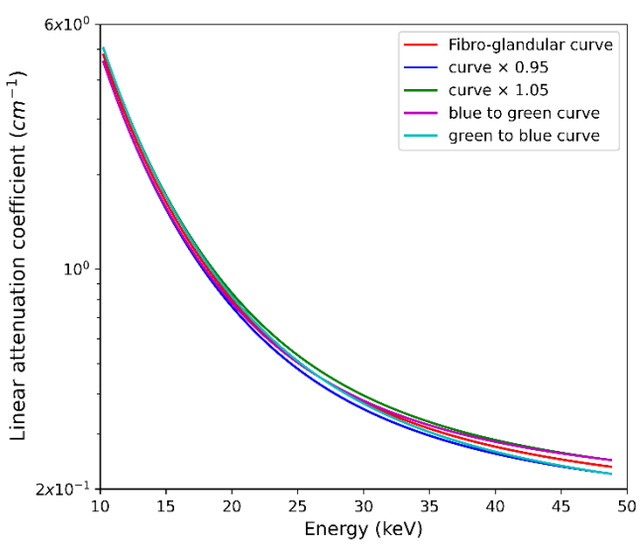


Fig. 1. Theoretical fibro-glandular attenuation curve (red) and its possible variations: less dense (blue), denser (green), less steep (magenta), and steeper (cyan) curves.

Midgley's parameterization resulted in a set of five mixture parameters for both standard and modified phantoms. These parameters were then used to set the representation of phantom materials to five coefficients only, resulting in the simulation computational cost no longer depending on the number of tissues being simulated.

C. Image simulation & Reconstruction

The simplified phantoms resulting from Midgley's parameterization were employed to simulate DBT projections, using the Lambert-Beer law:

$$\vec{y} = \sum_e \vec{b}_e \, e^{-L \cdot \vec{\mu}(e)} \tag{3}$$

where $\vec{b}_e$ represents the energy weighted spectrum, L is the projection matrix, and $\vec{\mu}(e)$ is the linear attenuation coefficient at energy e.

All three tissue types and the fibro-glandular compositional variations were simulated through the proposed approach.

For comparison, conventional simulations based on indexed materials were performed with both phantoms with constant tissue composition and those including fibro-glandular tissue variability.

Images resulting from both the conventional simulation method and the proposed approach were reconstructed using the ML-TR method [9] with 30 iterations.

D. Analysis

The accuracy of the fit performed by Midgley's parameterization was assessed by computing the mean absolute error (MAE) between the theoretical values of linear attenuation coefficient of the three breast tissue types evaluated from the mixture rule and the predicted values obtained from the linear combination of S-curves with the five mixture parameters resulting from the fit (equation (1)).

To evaluate the effect of simulating breast tissues using Midgley's components, we computed the Bland-Altman plot [9] between sinogram data generated by the conventional simulation method, and data obtained through the proposed approach. The median and interquartile range (IQR) of the differences represented in this plot were averaged across all cases, excluding open field voxels.

In the reconstruction domain, we computed the absolute percentage error between images reconstructed through the conventional simulation approach and those obtained through Midgley's parameterization. The average median and IQR of this error were then calculated across all phantoms, again excluding the open field from the analysis.

III. RESULTS

Fig. 2 shows the S-parameters obtained within the previously specified energy and compositional range of interest.

The evaluation of least-squares fitting is displayed in Table 1, reporting the average MAE between the theoretical linear attenuation coefficients and Midgley's modeled values.

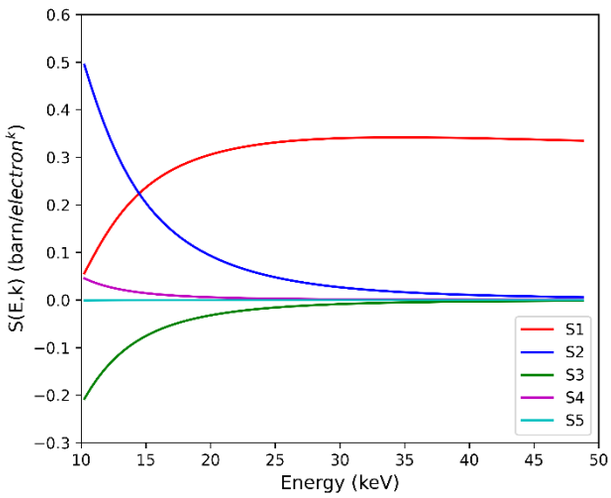


Fig. 2. S-curves computed by solving a least-square problem within the specified energy and compositional range of interest.

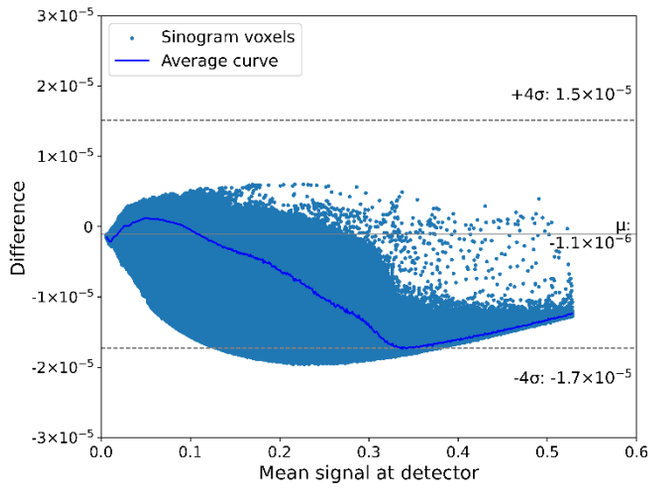


Fig. 3. Bland-Altman plot between the sinogram generated using the conventional simulation approach and that obtained through the method proposed in this work. The error of each voxel is reported with blue dots; the blue line represents the average curve. Open field voxels were excluded from this analysis.

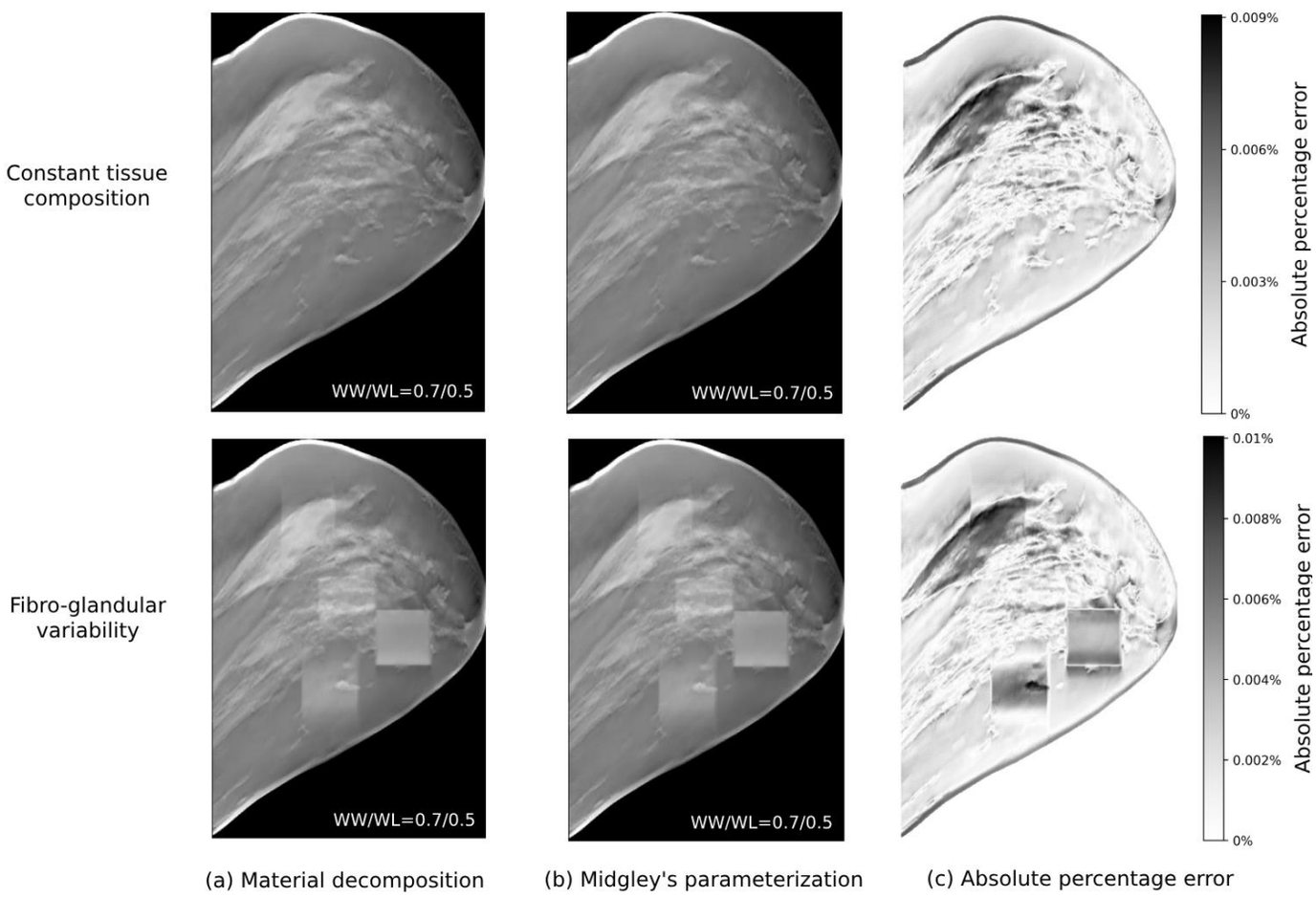


Fig. 4: Reconstructed axial slice obtained through the conventional simulation approach (a), corresponding slice of the image reconstructed using Midgley's model (b), and absolute percentage error (c) computed between these two slices. The top row displays a phantom with constant tissue composition, whereas the bottom row shows the corresponding phantom incorporating fibro-glandular tissue variability.

Table 1. Minimum, mean, and maximum MAE between theoretical and modeled linear attenuation coefficients for adipose, fibro-glandular and skin tissues, in cm^{-1}.

	Adipose	Fibro-glandular	Skin
Min.	0	0.2×10^{-5}	0.2×10^{-5}
Mean	2.0×10^{-5}	1.0×10^{-4}	1.1×10^{-4}
Max.	8.6×10^{-5}	4.7×10^{-4}	4.9×10^{-4}

The Bland-Altman plot in Fig. 3 showcases the difference between the sinograms obtained through the standard simulation method and those computed using Midgley's parameterization. The figure illustrates the difference for a single phantom with varying fibro-glandular composition. The difference computed through this plot showed an average median value of -1.9×10^{-6} (average IQR = [-3.2×10^{-6}, -4.0×10^{-7}]) across all phantoms with constant tissue composition, and -2.5×10^{-6} (average IQR = [-3.8×10^{-6}, -8.6×10^{-7}]) across all phantoms including fibro-glandular variability.

Finally, Fig. 4 shows the absolute percentage error between an axial slice of the standard reconstructed image and the same slice of the image obtained using Midgley's model. The calculation of this error resulted in an average median value of 0.0018% (average IQR: [0.0009%, 0.0028%]) across all phantoms with constant tissue composition, and 0.0021% (average IQR: [0.0011%, 0.0033%]) across all phantoms including fibro-glandular variability. The maximum absolute percentage error was found to be 0.0141% and 0.0149%, respectively.

IV. DISCUSSION

As illustrated in Fig. 3 and Fig. 4, an excellent agreement was observed between the conventional simulation method and the approach proposed in this study, in both projection and reconstruction domains. However, the negative bias observed in the Bland-Altman plot needs further examination. The average curve shows that this bias tends to increase as the mean signal at the detector increases, indicating an escalation as we approach the border of the breast.

The expected biological tissue variability we want to simulate is approximately 4%, as found by previous works [1]. The error between the reference attenuation curve and its Midgley fit was found to be <0.015% in all examined cases. Therefore, since the errors consistently remained below the intended variability, the proposed approach proved to be sufficiently accurate for our purpose.

In terms of computational efficiency, the proposed approach has a fixed cost since it reduces simulations to a fixed number of five parameters. Conversely, the conventional simulation method, reliant on material indexing, incurs a linear increase in computational cost proportional to the number of materials considered. This study did not measure computational efficiency for the examined phantoms since, including only four variations in fibro-glandular tissue composition, no gain in computational performance is expected. However, when simulating inter- and intra-patient variability of all breast tissues in every phantom voxel, the proposed approach will save computational time and power.

Therefore, our preliminary study demonstrated Midgley's decomposition to be an accurate tool for simulating tissue variability, without excessive computational demands.

Although this work focused on breast imaging, the proposed approach may have wider applicability for whole-body imaging.

V. CONCLUSION

This study demonstrated the efficiency of Midgley's parameterization to accurately simulate tissue variability, providing a faithful representation of breast tissue inhomogeneities.

In future work, we plan to include inter- and intra-patient variability of all breast tissues, expand our dataset, analyze the computational gain, and devise a more realistic method for incorporating tissue inhomogeneities into phantoms. Additionally, we aim to evaluate the proposed approach in whole-body x-ray imaging and analyze the resulting computational gain.

REFERENCES

[1] A. Tomal, I. Mazarro, E. M. Kakuno, and M. E. Poletti, "Experimental determination of linear attenuation coefficient of normal, benign and malignant breast tissues," *Radiat Meas, vol. 45, no. 9, pp. 1055–1059, 2010, doi: 10.1016/j.radmeas.2010.08.008.*

[2] E. Fredenberg, P. Willsher, E. Moa, D. R. Dance, K. C. Young, and M. G. Wallis, "Measurement of breast-tissue x-ray attenuation by spectral imaging: fresh and fixed normal and malignant tissue," *Phys Medicine Biology, vol. 63, no. 23, pp. 235003, 2018, doi: 10.1088/1361-6560/aaea83.*

[3] S. M. Midgley, "A parameterization scheme for the x-ray linear attenuation coefficient and energy absorption coefficient". *Physics in Medicine & Biology, vol. 49, no. 2, pp. 307, 2004.*

[4] M. Mikerov, K. Michielsen, J. G. Nagy, and I. Sechopoulos, "Effect of attenuation model on iodine quantification in contrast-enhanced breast CT." In *7th International Conference on Image Formation in X-Ray Computed Tomography, vol. 12304, pp. 635-640. SPIE, 2022*.

[5] "Xraydb," https://github.com/xraypy/XrayDB.

[6] W. T. Elam, B. D. Ravel, and J. R. Sieber. A new atomic database for x-ray spectroscopic calculations. *Radiation Physics and Chemistry, vol. 63, no. 2, pp. 121–128, February 2002, doi: 10.1016/S0969-806X(01)00227-.*

[7] E. García, C. Fedon, M. Caballo, R. Martí, I. Sechopoulos, and O. Diaz, "Realistic compressed breast phantoms for medical physics applications," in *Proc. SPIE, in 15th International Workshop on Breast Imaging (IWBI2020), vol. 11513, pp. 1151304, 2020, doi: 10.1117/12.2564273.*

[8] G. R. Hammerstein, D. W. Miller, D. R. White, M. E. Masterson, H. Q. Woodard, and J. S. Laughlin, "Absorbed radiation dose in mammography". *Radiology, vol. 130, no. 2, pp. 485–491, 1979, doi: 10.1148/130.2.485*

[9] J. Nuyts, B. De Man, P. Dupont, M. Defrise, P. Suetens, and L. Mortelmans, "Iterative reconstruction for helical CT: a simulation study," *Phys Med Biol, vol. 43, no. 4, pp. 729–737, 1998, doi: 10.1088/0031-9155/43/4/003.*

Principal Component Analysis for Assessing Spectral Performance of Photon-Counting Detectors – Spectral Distance Index

Zhicong Yu, Georgios Prekas, and Daniel Gagnon

***Abstract*—Photon-Counting CT has shown impressive clinical use-cases over the past few years. With this technology being matured, photon-counting detectors will be more and more available from different vendors. Therefore, it is important to compare across-vendor spectral performance. However, such a task has been difficult in both projection and reconstructed image domains. In this work, we propose a figure-of-merit called Spectral Distance Index using principal component analysis. This index only requires availability of projection data without any advanced reconstruction and material decomposition processes. Via both computer simulations and real data experiments from a tabletop system, we demonstrate that this index can effectively reflect spectral performance of photon-counting detectors and is robust to noise, scan geometry and spectrum estimation error.**

***Index Terms*—PCD, PCA, Spectral Distance Index**

I. Introduction

Photon-counting CT (PCCT) has demonstrated improved image quality at reduced dose, offering new potentials in both diagnostic imaging and radiation therapy [1]. It is foreseeable that PCCT will be commercially available from different vendors, and it is valuable to compare spectral performance across different photon-counting detectors (PCD). Such a task has been challenging as different vendors may have different system configurations and image formation processes. One approach is to use an isotope or synchrotron to measure the full width at half maximum of a single energy response. However, this approach is demanding in terms of x-ray source and accessibility to the detector. Another approach is to measure the material separation capability, but it requires utilization of various physics correction and advanced image reconstruction techniques, making it difficult to fairly compare the PCD performance across different PCCT systems.

This work introduces a Spectral Distance Index (SDI) for PCD spectral performance assessment using principal component analysis (PCA). PCA has been used in spectral CT previously for alternative image representations [2,3]. To the best of our knowledge, it has not been used for spectral content measurement. SDI exploits the fact that multi-bin images from a poor energy discriminating detector would present a high correlation between each energy bin, while an excellent spectral detector would show more unique feature therebetween. SDI uses projection data of line integrals (before reconstruction), allowing for a fair and practical comparison among different PCDs. The rest of this paper is dedicated to SDI definition and its effectiveness for assessing PCD spectral performance via computer simulations and real data from a tabletop system.

Zhicong Yu, Georgios Prekas, and Daniel Gagnon (formerly) are with Imaging Technologies, Accuray Incorporated.

II. Methods and Materials

A. Principal Component Analysis (PCA)

PCA can be defined differently depending on whether the measured data are centered (zero mean) or not. Let M be the number of energy bins and N be the detector pixels. Let $A = [\underline{g}_1, \underline{g}_2, \dots, \underline{g}_N]$ be the measured projection data of size $[M, N]$. Here, $\underline{g}_n$ is a column vector of line integrals, containing M measurements from the nth detector pixel as defined below.

$$\underline{g}_n = [g_{1n}, g_{2n}, \dots, g_{Mn}]^T$$

The centered PCA is then defined as follows.

Step 1. Center the measurement A so that each row has mean 0.

$$A_c = A - A \times \text{ones}(N,1)/N \times \text{ones}(1,N)$$

Step 2. Calculate the empirical covariance matrix of A_c.

$$C = A_c A_c^T/(N-1)$$

Step 3. Perform singular value decomposition on C, where $C = USV^T$ and $U = V$ since C is symmetric.

$$[U,S,V] = \text{svd(C)}$$

Step 4. Let r be a positive integer number and $r \leq \min(M,N)$. The principal vectors (columns of U_r) and their variances (non-negative S_r components in descending order) are as follows.

$$U_r = U(:,1{:}r) = [\underline{e}_1, \underline{e}_2, \dots, \underline{e}_r]$$

$$S_r = \text{diag}(S(1{:}r, 1{:}r) = [\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2]$$

A non-centered PCA is obtained by replacing A_c by A in Step 2. For the case of a single pixel, A becomes a column vector, and the PCA can be calculated using a modified empirical covariance matrix as defined below.

$$C_{singlePixel} = AA^T$$

A Couple of Toy Stories

Recall from the previous section that A stores line integrals. Therefore, PCA results are dependent on the following factors.

1. Imaging chain, e.g., x-ray spectrum and detector response,
2. Phantom material composition, and

3. Phantom dimensions (x-ray path lengths).

To utilize PCA results for spectral content measurement, we need to understand what information each PCA component represents. Below are a couple of toy examples.

Toy Example A: If there is only a single material with a single path length with four energy bins, then phantom material and path length does not play a role. What PCA does is finding the main axis that spreads the line integrals most. This main axis is roughly parallel to the linear regression of the linear attenuation curves of the material under consideration.

Toy Example B: Now let's consider two materials, each material with fixed path length and the same four energy bins. In this case, the first principal axis is still the one spread all line integrals most, but the second axis is perpendicular to the first principal axis and separates the two materials most.

Inspired by the above understanding, we are interested in developing figures-of-merit to measure spectral and material distance. SDI is inspired by Toy Example A.

B. Spectral Distance Index

We now define the SDI. First, we consider a standard CT/Cone-Beam CT single projection data acquisition setup as shown in Figure 1A. We assume the intrinsic filter corresponding to the tube and housing is known or well calibrated, e.g., 3 mm equivalent of aluminum. The detector is placed orthogonal to the source-iso-center line. A slab with known material and thickness is placed between the source and the detector such that the slab is also perpendicular to the source-iso-center line. We also assume the geometry is known or well calibrated, e.g., source-to-detector distance (SDD) and source-to-iso-center distance (SID). Then a single projection or average of multiple projections of line integrals are taken. All projection data are stored in A as specified in Section II.A. The SDI can be obtained via the following steps.

Step 1. Model the imaging chain with known or calibrated parameters and ideal detector response. Some key information can be obtained as below.

a. X-ray spectrum $\phi(E)$: We assume modern x-ray tubes of CT scanners share equivalent spectrum, which can be obtained via software or library, e.g., SpekCalc.
b. Linear attenuation coefficients: They can be obtained from the National Institute of Standards and Technology.
c. Any inherent filtration: This shall be available via vendor specification or experimental calibration.

Using the model, the line integral for each energy bin with ideal PCD response is calculated; the first principal axis $\underline{e}_{\text{ref}}$ and its corresponding variance σ^2_{ref} can be obtained via non-centered PCA using $C_{singlePixel}$ as specified in Section II.A.

Step 2. Perform non-centered PCA on measurement A to obtain its first principal axis $\underline{e}$ and the corresponding variance $\underline{\sigma}^2$.

Step 3. The SDI, denoted by (r_σ, θ), is defined in Equation [1]. Here, r_σ and θ are called the relative Euclidean distance and angular distance, respectively. Their geometric meanings are illustrated in Figure 1B. By definition, $r_\sigma \in [0, +\infty)$ and $\theta \in [0, \pi)$, with $(r_\sigma, \theta) = (0,0)$ representing the best spectral performance.

$$r_\sigma = \sqrt{\left|\sigma^2_{\text{ref}}\underline{e}_{\text{ref}} - \sigma^2\underline{e}\right|}/\sigma_{\text{ref}}$$
$$\theta = \text{acos}\left(\underline{e}^T\underline{e}_{\text{ref}}\right) \quad [1]$$

To enable cross-system PCD comparison, data acquisition setup shall be matched as much as possible. Meanwhile, it is also crucial to make the setup accessible or require bear minimum amount of effort. For example, 120 kVp and 15 mm aluminum slab. Aluminum has been a common material used for tube housing filtration specification as well as for flat filters. It is also recommended that the aluminum slab thickness is not too large to avoid poor photon statistics.

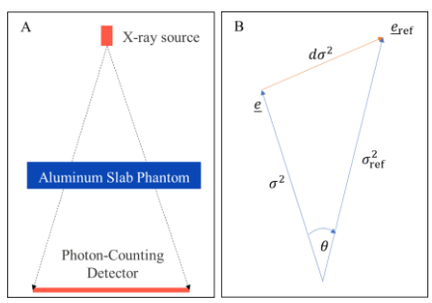


Figure 1: SDI data acquisition scheme (A) and geometric meaning (B).

C. Computer Simulation

SDI is designed to compare detector spectral performance. In an ideal situation, other than the detector, the rest of the imaging chains of the two systems under comparison shall be identical. This is impossible, and certain tolerance should be allowed. To this end, we conducted computer simulations to answer the following questions.

1. How can SDI reflect detector spectral performance?
2. How sensitive is SDI to noise, scan geometry, and estimation error in imaging chain?

For all computer simulations, we generated a 120 kVp spectrum with 1.43 mm intrinsic aluminum filter via SpekCalc. The PCD is configured with 4 thresholds at 25 kev, 45 kev, 60 kev and 80 kev. We set the SID and SDD 1040 mm and 1554 mm, respectively. We used an aluminum slab thickness of 15 mm. We simulated the PCD using the following parameters.

- Semiconductor material: CdTe
- Material thickness: 1.6 mm
- Bias voltage: 800 V
- Native pixel size: 150 μm.
- Detector matrix: 15 rows x 512 channels
- Temperature: 300 Kelvin
- Relative permittivity: 10.6
- Energy per quanta: 4.43 ev
- Mobility: 1000

We included two physics effects, charge sharing (CS) and K -fluorescence and -escape. For CS, we divided the material thickness into 100 segments over 1.6 mm. Both diffusion and repulsion were taken into account [4-7]. The final charge clouds were approximated by 2D uniform disks. For K-fluorescence and -escape, we only considered photoelectric effects, since it is dominant for CdTe in the diagnostic energy range (~94%). K shell emission (α and β, absorbed and escaped) of both Cd and Te were included.

To show how SDI reflects detector spectral performance, we compared the following four detector configurations.

1. Detector pixel size of 150 *um* (1x1 binning) with CS and K-fluorescence and -escape both included.
2. Detector pixel size of 150 *um* (1x1) with CS only.
3. Detector pixel size of 300 *um* (2x2) with CS only.
4. Detector pixel size of 450 *um* (3x3) with CS only.

The rationale is that the detector spectral performance would be degraded with increased charge sharing (reduced pixel size) and addition of K-fluorescence and -escape; see Figure 2.

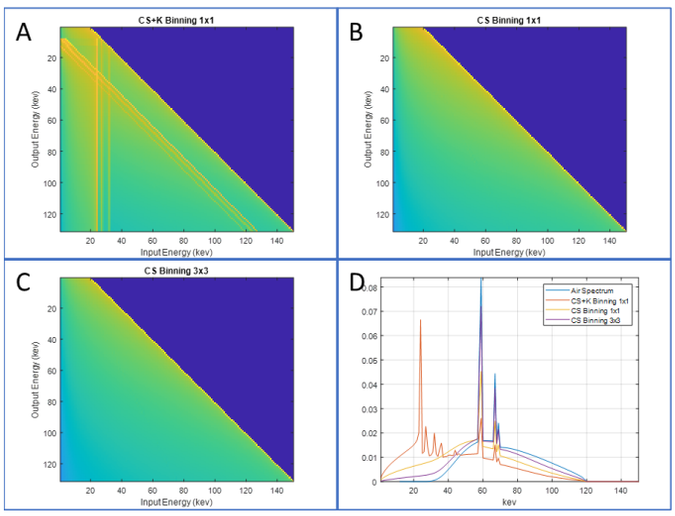


Figure 2: Simulated PCD response with charge sharing and K-fluorescence and -escape at pixel size of 150 um (A), with charge sharing only at pixel size of 150 um (B), and with charge sharing only at pixel size of 450 um (C). As an example, we also show the PCD response to an input spectrum in D, demonstrating degradation of the detector spectral performance with increased charge sharing (reduced pixel size) and inclusion of K-fluorescence and -escape.

To investigate SDI sensitivity to noise, we simulated projection data at dose levels from 0.1 mAs to 2 mAs per view, using 150 um pixel size with both charge sharing and K-fluorescence and -escape included. We added Poisson noise for each projection data. No electronic noise was added. For the kth energy bin $[E_{kL}, E_{kH}]$, the x-ray spectrum $\phi(E)$ convolved with the detector response $\eta(E_i, E_o)$, denoted by $\phi_k(E)$, was calculated using Equation [2]. Poisson noise was added to each bin using the total counts of the bin spectrum attenuated by the phantom for each path length. SDI was then calculated for each noise realization and compared.

$$\phi_k(E) = \int_{E_{kL}}^{E_{kH}} \phi(E)\eta(E, E_o)dE_o \qquad [2]$$

Next, we simulated representative scan geometries for both diagnostic CT and radiation therapy imaging; see Table 1. The PCD pixel size was set to 150 um with both CS and K-fluorescence and -escape included. We then compared SDI among those geometries.

Lastly, we investigated SDI sensitivity to estimation error of the x-ray spectrum by introducing additional aluminum filter to the imaging chain and compared the corresponding SDI.

Table 1: List of Scan Geometries Used for SDI Investigation

mm	Geo I	Geo II	Geo III	Geo IV	Geo V
SID	1040	541	570	595	600
SDD	1554	949	1040	1085	1072

D. Tabletop Experiments

We also demonstrate SDI using real data from a tabletop system equipped with a photon counting detector (Thor, Varex Imaging Inc.). This PCD consists of 16 tiles with each tile containing 256x128 isotropic pixels of size 0.1 mm x 0.1 mm. No anti-scatter grid was installed on the detector. The PCD comes with hardware-based anti-coincidence correction (ACC) for reduction of charge sharing and K-fluorescence and -escape.

We compared the SDI of two scans, one was with ACC off and the other on; see Table 2. Both scans used the same imaging chain of 120 kVp, 1.6 mm aluminum filter (tube housing) and 0.512 mm copper filter, with the PCD operating in a high sensitivity mode with bias refreshing on. An aluminum slab of thickness 18.415 mm was used as the attenuation object and placed close to the source to reduce scatter as much as possible. Both scans were taken at 72 frames per second, 5 ms pulse width and 20 mA. All object scans were taken with a narrow aperture of 6 mm. It is expected that the ACC should improve spectral resolution and thus we expect the SDI of the scan with ACC on to be noticeably smaller than that with ACC off.

Table 2: List of Scan Configurations for Thor Tabletop Experiments

Scan ID	Object Scan	Air Scan	ACC
A	1000 frames at 20 mA	200 frames at 20 mA	Off
B			On

All projection data were converted to line integrals with bad pixel correction, tile edge correction, shadow-based scatter correction, air normalization and negative logarithm. No other advanced correction steps like linearization were performed.

III. Results

A. Computer Simulations

The SDI, (r_σ, θ), of different detector modellings are shown in Table 3. As expected, SDI is smallest $(0.127, 3.721°)$ for 450 um (3x3) pixel size with charge sharing only and largest $(0.384, 13.487°)$ for 150 um (1x1) pixel size with both intense charge sharing and K-fluorescence and -escape.

Table 3: SDI for Various Levels of Modelling of Charge Sharing and K-Fluorescence and -Escape.

	CS + K 150 um	CS 150 um	CS 300 um	CS 450 um
r_σ	0.384	0.267	0.171	0.127
$\theta°$	13.487	8.383	5.137	3.721

In terms of SDI sensitivity to noise, we noticed that while the variance of the first principal axis indicates spectral distance, the sum of the rest of the variances is strongly correlated to noise. Figure 3 shows that the error bars (red) are ignorable relative to the signal themselves (blue), indicating that the noise has negligible impact on SDI for a wide range of dose levels.

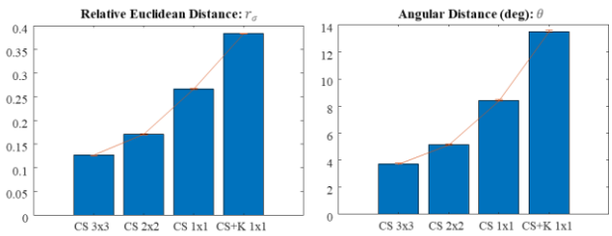


Figure 3: Impact of noise on SDI. Here CS and K stand for charge sharing and K-fluorescence and -escape, respectively. Also, 3x3, 2x2 and 1x1 are binning modes, corresponding to 450 um, 300 um and 150 um pixel sizes, respectively. The error bars (red) in both r_σ and θ plots are negligible relative to their magnitudes(blue).

In terms of SDI sensitivity to scan geometries, results are listed in Table 4. SDI (r_σ, θ) changed less than $(0.01, 0.06^\circ)$ across the representative geometries, demonstrating robustness.

In terms of SDI sensitivity to spectrum estimation error, with addition of 0.5 mm aluminum error, the SDI varied less than $(0.01, 0.6^\circ)$, which is negligible compared to PCD responses.

Table 4: SDI for Representative Scan Geometries in CT and RT

	Geo I	Geo II	Geo III	Geo IV	Geo V
r_σ	0.382	0.379	0.377	0.378	0.379
θ°	13.521	13.559	13.579	13.564	13.557

B. Tabletop Experiments

Using the same imaging chain, we performed energy sweeping with ACC on and off, and obtained the corresponding spectra as shown in Figure 4A. We also averaged the line integrals over all pixels within the open aperture region; the results are shown in Figure 4B. Both Figure 4A and B show that ACC can improve Thor spectral resolution. The SDI results are shown in Figure 4C and D. As expected, the SDI with ACC on is noticeably smaller than that with ACC off.

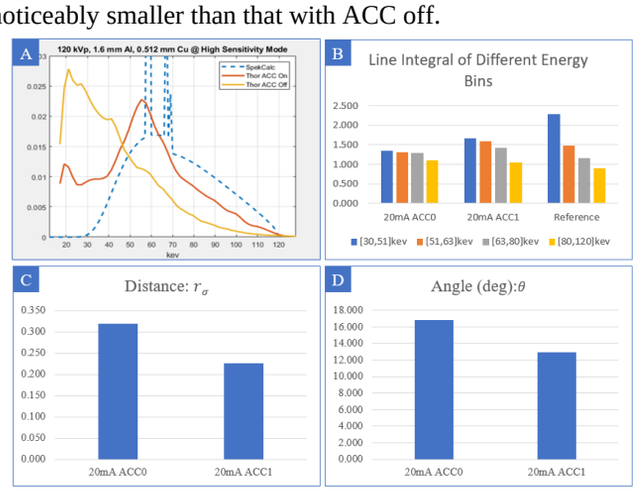


Figure 4: A: Spectra obtained via energy sweeping for the tabletop system with Thor operating in a high sensitivity mode. B: Average of line integrals over all pixels within the beam opening. C & D: SDI results for the scans with ACC on and off.

IV. Discussion

PCA is a popular tool for data representation. As mentioned previously, the PCA results of PCCT data sets could have principal axes that correspond to the directions of imaging chain spectral spread, material spread, and noise spread. In typical PCCT scan scenarios, such as those mentioned in this work, the first principal axis corresponds to the imaging chain spectral spread. If the scan configurations are controlled to have negligible impact for different PCDs, the first principal axis and its variance can be used to compare PCD spectral performance. The proposed SDI reflects how the measured data deviates from the modelling in terms of Euclidean and angular distances relative to the first principal axis. While SDI is designed for cross-system PCD performance comparison, we are also interested in developing a material separation index based on the second principal axis and its variance (Toy Example B in Section II.A.. Such an index may be used for imaging chain and scan protocol design for a dedicated PCCT system.

There are several limitations of the current work, and we plan to work on them in the future. First, it is not clear how sensitive SDI is to scatter and pulse pileup. Scatter and pulse pileup may distort spectrum in different directions relative to what charge sharing might do. It would also be interesting to see how SDI could reflect the spectrum distortion caused by charge sharing and pulse pileup together. We would also like to compare SDI before and after data linearization, as it is natural to expect such data correction to be included in the accessible line integral data from different vendors. Second, it is not clear how the number and value of energy thresholds impact SDI. This is important since it is possible that a PCD may work better in some energy ranges than others. Lastly, the current SDI is designed with flat filters only. However, it might be the case that a bowtie filter is not removable, and SDI shall be adapted to such cases.

V. Summary

We have proposed a figure-of-merit called SDI for cross-system comparison of PCD spectral performance. Via both computer simulations and tabletop experiments, we have shown that SDI can successfully reflect the negative impact of CS and K-fluorescence and -escape on PCD spectral response. For typical setups in diagnostic imaging and radiation therapy imaging, SDI is robust to noise, scan geometry and imaging chain estimation error, demonstrating its practical value.

References

[1] Kishore et al. "First clinical photon-counting detector CT system: technical evaluation." Radiol. 303, no. 1 (2022): 130-138.

[2] Xie et al. "Principal component analysis in projection and image domains—another form of spectral imaging in photon-counting CT." IEEE Trans. Biomed. Eng. 68, no. 3 (2020): 1074-1083.

[3] Butler et al. "Processing of spectral X-ray data with principal components analysis." Nucl. Instrum. Methods Phys. Res., Sect. A 633 (2011): S140-S142.

[4] Koch-Mehrin et al. "A spectroscopic Monte-Carlo model to simulate the response of pixelated CdTe based detectors." Nucl. Instrum. Methods Phys. Res., Sect. A 976 (2020): 164241.

[5] Benoit, Mathieu, and L. A. Hamel. "Simulation of charge collection processes in semiconductor CdZnTe γ-ray detectors." Nucl. Instrum. Methods Phys. Res., Sect. A 606, no. 3 (2009): 508-516.

[6] Gatti et al. "Dynamics of electrons in drift detectors." Nucl. Instrum. Methods Phys. Res., Sect. A 253, no. 3 (1987): 393-399.

[7] Iniewski et al. "Modeling charge-sharing effects in pixellated CZT detectors." In 2007 IEEE NSS/MIC Conference Record, vol. 6, pp. 4608-4611.

The accuracy of CT Numbers for DIRA with indata from a Real Dual-Source DECT-scanner

Maria Magnusson, Mikael Sandborg, Åsa Carlsson Tedgren, and Alexandr Malusek

***Abstract*—Accurate CT numbers obtained from Dual-energy computed tomography (DECT), may improve the accuracy of radiation treatment planning. Our Dual energy Iterative Reconstruction Algorithm (DIRA) utilizes material decomposition in the iterative loop to produce accurate CT numbers of the imaged object. DIRA produces two mono-energy images. These mono-energies correspond to the effective energies of the two X-ray spectra. In simulations, the resulting CT numbers in the reconstructed images have shown to agree well with tabulated values.**

The aim of this work is to see if DIRA can also be used on real data. Compared to simulated data, the following adjustments were needed; i) Elimination of patient table. ii) Bow-tie filter modeling and adaptation to this in DIRA. iii) Application of inverse water beam hardening correction.

A Siemens SOMATOM Force scanner was used to scan a cylindrical PMMA phantom with four rod inserts made of different materials. For simlicity, an axial scan was used. 6 iterations of DIRA was applied, where the 0th iteration is our own implementation of filtered backprojection (FBP). DIRA at iteration 6 gave values close to the tabulated ones.

DIRA also performed well on an anthropomorphic head phantom (CIRS model 731-HN), with specified material composition.

I. Introduction

Computed tomography reconstructs linear attenuation coefficients (LACs). To suppress artifacts resulting from their strong dependence on photon energy, a water beam-hardening correction is always performed on the projection data. This ensures that a phantom consisting of water only will have no beam-hardening artifacts. The LACs are presented to the end-user in the form of CT numbers,

$$H = 1000 \cdot (\mu/\mu_w - 1)\,, \tag{1}$$

where μ and μ_w are the linear attenuation coefficients (LAC) of the material and water, respectively, at a specific photon energy. Consequently, water will have the CT number 0 HU.

The X-ray tube produces photon beams with a spectrum of energies and the scanned object normally contains multiple materials. The strong energy dependence of LACs causes beam-hardening artifacts in single-energy CT (SECT). In dual-energy CT (DECT), two sets of data are obtained, each corresponding to a different tube voltage. DECT has the potential to better suppress beam-hardening artifacts, obtain accurate CT numbers, and extract knowledge of the material composition of the imaged object. One application of interest is radiation treatment planning, where DECT may improve the accuracy [1].

The different DECT algorithms can be divided into projection-based basis material decomposition (PBBMD) and image-based basis material decomposition (IBBMD) methods [2]. PBBMD methods, such as the well-known Alvarez-Macovski's base material decomposition (AMBMD) [3], require geometrically consistent projections and only two base materials can be assigned for the whole object. IBBMD methods can use any number of tissue-specific material doublets or triplets, nevertheless the methods suffer from beam-hardening artifacts.

In [4], two monoenergetic images are produced via AMBDM with the water-iodine base material doublet. Then a three-material decomposition is applied, where the pixel-specific material bases are selected according to the values of reconstructed LACs.

A clinical application is the Monoenergetic Plus algorithm [5], which we investigated in [9]. It performed well, but there were small discrepancies in the CT-numbers, probably caused by the two-material decomposition to iodine and water. (Iodine is slightly problematic due to its K-edge at 33.2 keV [10].) Also, the suppression of the beam-hardening artifact was not complete.

Two methods that work simultaneously on projections and images are the material decomposition from inconsistent rays (MDIR) [8] and DIRA [6]. Both can work with inconsistent rays from a dual-source DECT scanner. MDIR is an iterative image reconstruction method that produces two base material images. It uses FBP and the original measured projections inside the iterative loop. The original 2D DIRA algorithm is an iterative reconstruction method which uses FBP inside its iterative loop. Within its iterative loop, it uses two- and three-material decomposition to material doublets and triplets for the characterization of body tissues. The DIRA-3D algorithm is adopted to dual-energy dual-source 3D helical CT [11]. In simulations, DIRA and DIRA-3D has shown to give accurate CT numbers.

This paper aims to evaluate whether DIRA can also perform well on real data, by comparing the measured CT numbers with tabulated values derived from the EPDL97 library [12].

II. Theory

A. Material Decomposition

Two-material decomposition (2MD) assumes that a mixture is composed of two materials. 2MD determines mass fractions,

Maria Magnusson[1,2,3] (e-mail: maria.magnusson@liu.se), Åsa Carlsson Tedgren[2,3,4] and Alexandr Malusek[2,3] are with [1]Dept. of Electrical Engineering, [2]Dept. of Health, Medicine and Caring Sciences, [3]Center for Medical Image Science and Visualization (CMIV), Linköping University, SE-581 83 Linköping, Sweden and [4]Dept. of Medical Radiation Physics and Nuclear Medicine, Karolinska University Hospital, SE-171 77 Stockholm, Sweden.

w_1 and w_2, of the two base materials and the mass density, ρ, of the mixture. Three-material decomposition (3MD) assumes that a mixture consists of three base materials. 3MD determines mass fractions, w_1, w_2 and w_3, and the mass density of the mixture, which is calculated as $\rho^{-1} = \sum_{k=1}^{3} w_k/\rho_k$, where ρ_k and w_k are the mass density and mass fraction, respectively, of the kth material. In both methods, the mass fractions are normalized so that $\sum_k w_k = 1$. More information on the resulting systems of linear equations is in [6]. In the current implementation, bone tissues are decomposed into bone, bone marrow and density, soft tissues are decomposed into lipid, protein, and water, and low attenuation materials such as air are decomposed into lipid, water and density.

B. Forward projection generation

The logarithm of attenuation, here referred to as the polyenergetic projection P, is calculated as

$$P = \ln(I_0/I), \tag{2}$$

where I and I_0 are the detector responses with and without, respectively, the imaged object. The intensity I_0, is calculated for an energy integrating detector as

$$I_0 = \int_0^{E_{max}} D(E)N(E)dE, \tag{3}$$

where $D(E)$ is the detector responsivity, $N(E)\,dE$ is the number of photons with energies in the interval $(E,\, E+\, dE)$ emitted from the x-ray tube. An ideal detctor had $D(E) = E$. The intensity I is calculated as

$$I = \int_0^{E_{max}} D(E)N(E) \exp\left[-\int_{\mathcal{L}} \mu(x,y,E)dl\right] dE, \tag{4}$$

where $\mu(x, y, E)$ is the LAC of the pixel at position (x, y) with energy E and $\int_{\mathcal{L}} dl$ is a line integral through the object. This calculation is time consuming since the line integrals must be calculated for all energies in the energy spectrum. The calculation of projections changes slightly when material decomposition is introduced. The line integrals are calculated through volume fractions of the different base materials. The intensity I is then

$$I = \int_0^{E_{max}} D(E)N(E) \exp\left[-\sum_k \mu_k(E)l_k\right] dE, \tag{5}$$

where μ_k is the LAC of the kth base material and l_k is computed as

$$l_k = \rho_k^{-1} \int_{\mathcal{L}} \rho(x,y)w_k(x,y)dl, \tag{6}$$

where ρ_k is the tabulated density of the kth base material, $\rho(x, y)$ is the calculated density in the pixel at position (x, y) and $w_k(x, y)$ is the mass fraction of the kth material in the pixel at position (x, y). The density ρ and the mass fractions w_k are obtained from the two-material or three-material decomposition. The line integral $\int_{\mathcal{L}} dl$ can be calculated using Josephs method [7].

A monoenergetic projection P_{mono}, at a specific energy E, is calculated as

$$P_{mono} = \sum_k \mu_k(E)l_k. \tag{7}$$

C. Calculation of the two effective energies to be used in DIRA

Two effective energies $E_{\mathrm{eff,L}}$ and $E_{\mathrm{eff,H}}$, were L stands for low and H stands for high, are chosen in DIRA by using the following method. The effective attenuation coefficient μ_{eff} for water is calculated by using the weighted mean of the X-ray spectrum and the attenuation curve $\mu(E)$ for water,

$$\mu_{\mathrm{eff}} = \frac{\int_0^{E_{max}} D(E)\,N(E)\,\mu(E)\,dE}{\int_0^{E_{max}} D(E)\,N(E)\,dE}. \tag{8}$$

Then, using the $\mu(E)$-curve for water, the energy value corresponding to μ_{eff} is taken as the effective energy.

D. Summary of DIRA

The complete DIRA algorithm is illustrated in Figure 1 and described below.

1) Measured projections $P_{\mathrm{M,L}}$ and $P_{\mathrm{M,H}}$ are obtained for low- and high-energy spectra, respectively.
2) They are reconstructed with filtered backprojection (FBP) to corresponding μ_{L} and μ_{H}. The initial reconstruction (iteration 0) is preceded by a conventional water beam-hardening correction (not shown in the figure).
3) For the phantoms considered in this work, a threshold segmentation is performed to separate regions of bone tissue, soft tissue and low attenuation materials such as air. The results are stored in 2×3 images $\mu_{\mathrm{T,L}}$ and $\mu_{\mathrm{T,H}}$.
4) Tissues are decomposed using the base material decomposition methods (section II-A). Density and base material mass fraction images μ_B are recieved.
5) Monoenergetic forward projections, $P_{\mathrm{mono,L}}$ and $P_{\mathrm{mono,H}}$, are generated for the low and high effective energies, $E_{\mathrm{eff,L}}$ and $E_{\mathrm{eff,H}}$, respectively. See (7) and (8). They are then reconstructed to $\mu_{\mathrm{mono,L}}$ and $\mu_{\mathrm{mono,H}}$.
6) Polyenergetic forward projections P_{L} and P_{H} are generated, see (2), (3) and (5). They are then compared with the measured projections. Reconstruction by FBP gives updates $\Delta\mu_{\mathrm{L}}$ and $\Delta\mu_{\mathrm{H}}$. The addition to $\mu_{\mathrm{mono,L}}$ and $\mu_{\mathrm{mono,H}}$ gives μ_{L} and μ_{H} for the next iteration.

Points 2-6 are iterated a predefined number of times. In [10], we show that two-material water-bone decomposition is sufficient for human tissues with atomic number $Z \leq 20$.

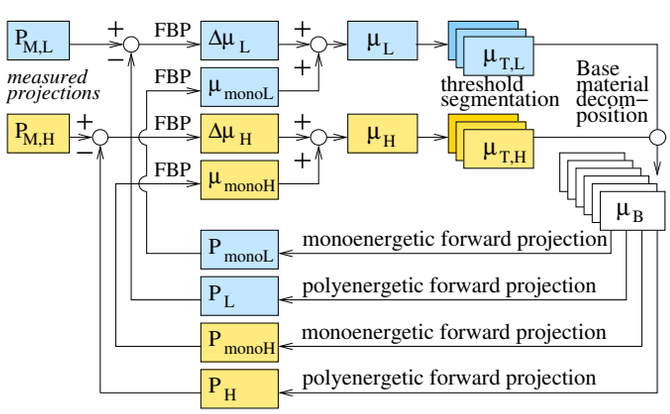


Fig. 1. Flowchart of the dual-energy iterative reconstruction algorithm DIRA.

III. Materials and Methods

A. CT Scanner and phantom data

A Siemens SOMATOM Force scanner (Siemens Healthineers) was used to scan a cylindrical polymethyl methacrylate (PMMA) phantom of diameter 160 mm with four rod inserts of diameter 20 mm made of aluminium, polytetrafluoroethylene (PTFE, Teflon), and low-density polyethylene (LDPE), see Figure 3, top left. Peripheral holes for a pencil ion chamber were filled with air. A clinical-like, axial, dual-energy protocol with 80 kV and 150 kV with an added Sn filter (denoted as Sn150kV) was used. The corresponding effective energies are 50 keV and 93 keV, respectively, see equation (8).

For comparison, images were reconstructed with the ADMIRE algorithm (Qr36d kernel, strength 2) in the Siemens' syngo.via 5.1 software. ADMIRE suppresses noise much better than FBP. To minimize the noise in our own reconstructions with DIRA and the rod phantom, several detector rows were averaged. This is reasonable since the rod phantom has no variation in the longitudinal direction.

Moreover, an anthropomorphic head phantom (CIRS model 731-HN), see Figure 2, was exposed to an axial scan through the teeth. The CIRS head phantom has a specified material composition corresponding to brain, soft tissue, spinal cord, trabecular bone, cortical bone, enamel, dentin, and more.

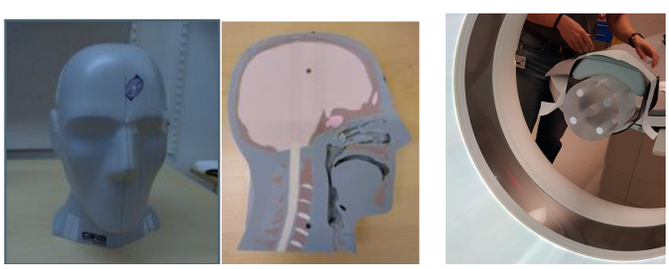


Fig. 2. Left: The CIRS Model 731-HN head phantom. Right: The rod phantom floating in the center of the CT scanner.

B. Adjustments in DIRA to match real data

See Figure 1. The following procedure gave measured projections $P_{\mathrm{M,L}}$ and $P_{\mathrm{M,H}}$. The CT scanner was set to give axial scans and rawdata were produced. We can regard these data as fanbeam projections. The rawdata have been exposed to water beam hardening correction (WBHC). This is unfortunate because DIRA requires data that has not been exposed to WBHC. In the past, we tried to set the scanner in service mode to remove WBHC, but these experiments were not successful. Instead, we came up with the idea to perform inverse WBHC on the rawdata.

To remove the influence of the patient table, the rod phantom can be attached to the end of the patient table to let it float in the center of the CT scanner, see Figure 2. Another option is to first scan the phantom on the patient table and then do another scan of only the patient table and finally subtract the second scan from the first. We tried both methods with the rod phantom and they gave similar results.

The fanbeam projections were then rebinned to parallel projections with 511 channels per projection angle. These now constitute the measured projections $P_{\mathrm{M,L}}$ and $P_{\mathrm{M,H}}$.

Observe the effective spectrum $D(E)N(E)$ in equations (3), (4), and (5). The bowtie filter is made of aluminum and is thinner in the middle. As the X-rays travel through the bowtie filter, they are attenuated and undergo beam-hardening. Therefore, the shape of the spectrum varies depending on the angle at which the rays were emitted from the x-ray tube. In other words, the effective spectrum $D(E)N(E)$ is different for different angles. We adapted DIRA to handle this.

The CIRS head phantom was placed on a pillow on the patient table. First, a scan was done with CIRS+pillow+table. Then a scan was done with pillow+table. The two scans were then subtracted. The pillow changed place and shape slightly between the two scans and it was later found that this gave rise to a weak artifact in the back of the head.

IV. Results and Discussion

Figure 3 shows images of the rod phantom reconstructed by ADMIRE and DIRA at iterations 0 and 6 for the 80 kV spectrum with effective energy 50 keV. Note the similar results obtained with ADMIRE and DIRA, iteration 0. Note that DIRA, at iteration 6, was able to remove the weak beam hardening artifact between the Al and PTFE rods. In [9], we found that the suppression of the beam-hardening artifact was not complete with Monoenergetic Plus [5]. (There are no experiments with Monoenergetic Plus in this paper, however.) Table I further shows that the good agreement between ADMIRE and DIRA, iteration 0. In addition, the table shows a good agreement between tabulated values and DIRA, iteration 6. Figure 4 shows profiles along the Al and PTFE rods in

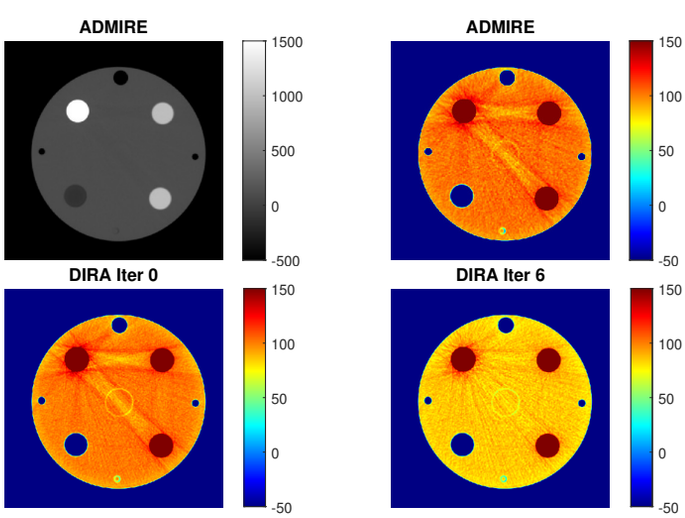


Fig. 3. Images of the rod phantom reconstructed by ADMIRE and DIRA. 80 kV spectrum with effective energy 50 keV.

the PMMA rod phantom at 80 kV spectrum with effective energy 50 keV and Sn150 kV spectrum with effective energy 93 keV. Note how the Al rod approaches the tabulated values after 6 iterations with DIRA. For Al and 80 kV, note how the downward bend in iteration 0 straightens out in iteration 6.

TABLE I
CT NUMBERS [HU] CALCULATED IN REGIONS OF RADIUS 10 PIXELS THE DIFFERENT MATERIALS IN THE ROD PHANTOM, ALONG WITH TABULATED VALUES. 80 KV SPECTRUM WITH EFFECTIVE ENERGY 50 KEV.

Material	ADMIRE	DIRA, iter=0	DIRA, iter=6	Tabulated
Al	2711	2700	3341	3363
PTFE	983	980	1010	1026
PMMA	104	102	85	83
LPDE	-111	-114	-134	-135

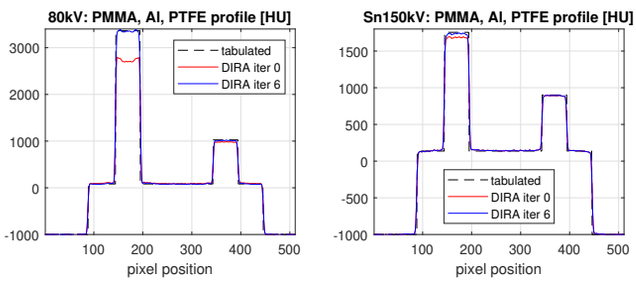


Fig. 4. Profiles along the Al and PTFE rods in the PMMA rod phantom for the two spectra, 80 kV and Sn150 kV.

Figure 5 shows images of the CIRS head phantom reconstructed by DIRA at iteration 0 and 12 for the 80 kV spectrum with effective energy 50 keV. The images are shown with different contrast windows and different color tables. Table II shows the improvement of DIRA from iteration 0 to 6 and 12, to a rather good agreement with values calculated from the CIRS phantom specification. Note that the enamel in the teeth improves considerably with iterations. Also note that the soft tissue in the tongue (halved) and neck converges to almost the same value (30 HU and 29 HU).

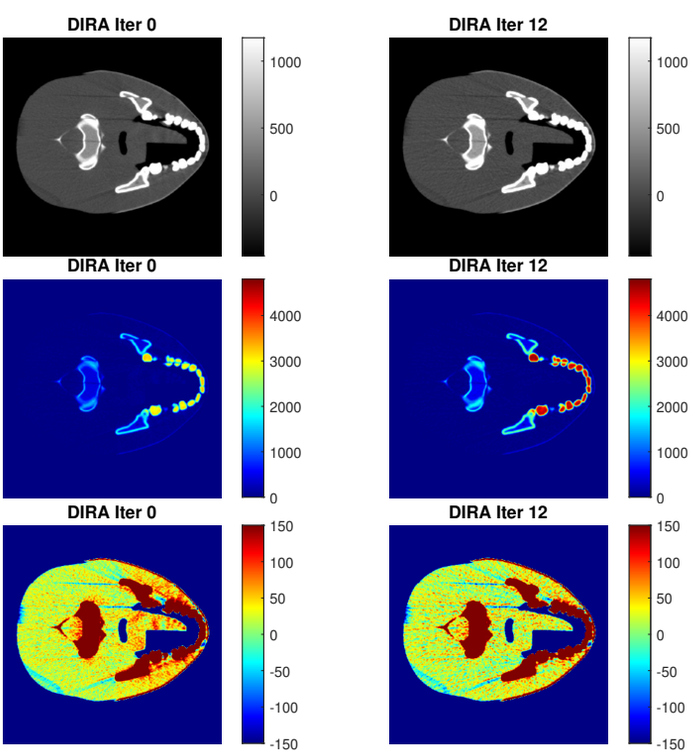


Fig. 5. Images of the CIRS head phantom reconstructed by DIRA at iterations 0 and 12. 80 kV spectrum with effective energy 50 keV.

TABLE II
CT NUMBERS [HU] CALCULATED WITH DIRA IN CIRCULAR REGIONS OF RADIUS 10 PIXELS (6 FOR ENAMEL) IN THE CIRS PHANTOM, ALONG WITH SPECIFIED VALUES. 80 KV SPECTRUM WITH EFF. ENERGY 50 KEV.

Tissue	DIRA iter=0	DIRA i=6	DIRA i=12	Specified
Trabec. bone	426	526	526	488
Enamel	3183	4560	4595	4720
Soft: tongue	78	34	30	19
Soft: neck	28	30	29	19

V. CONCLUSION

The DIRA algorithm performed well on real data. Obtained CT numbers were close to tabulated values derived from the EPDL97 library. Image examples showed that DIRA was able to suppress beam hardening artifacts.

In terms of noise, the images reconstructed with DIRA are comparable to images reconstructed with FBP and cannot compare with images reconstructed with ADMIRE.

VI. ACKNOWLEDGMENT

Thanks to Jonas Nilsson Althén and Lilian Henriksson for help with scanning the phantoms. Thanks to Karl Stierstorfer and Sebastian Faby for providing information about the Siemens SOMATOM Force scanner, including bowtie filter.

REFERENCES

[1] E. Bär, A. Lalonde, G. Royle, H.-M. Lu, and H. Bouchard. The potential of dual-energy CT to reduce proton beam range uncertainties. *Medical Physics*, 44(6):2332–2344, June 2017.

[2] B. J. Heismann, B. T. Schmidt, and T. Flohr. *Spectral Computed Tomography*. SPIE Press, Bellingham, Wash. (1000 20th St. Bellingham WA 98225-6705 USA), 2012.

[3] R. E. Alvarez and A. Macovski. Energy-selective reconstructions in X-ray computerised tomography. *Physics in Medicine and Biology*, 21:733–744, September 1976.

[4] P. R. S. Mendonca, P. Lamb, and D. V. Sahani. A Flexible Method for Multi-Material Decomposition of Dual-Energy CT Images. *IEEE Transactions on Medical Imaging*, 33(1):99–116, 2014.

[5] P. Wohlfahrt, C. Möhler, V. Hietschold, S. Menkel, S. Greilich, M. Krause, M. Baumann, W. Enghardt, and C. Richter. Clinical Implementation of Dual-energy CT for Proton Treatment Planning on Pseudo-monoenergetic CT scans. *International Journal of Radiation Oncology*Biology*Physics*, 97(2):427–434, February 2017.

[6] A. Malusek, M. Magnusson, M. Sandborg, and G. Alm Carlsson. A model-based iterative reconstruction algorithm DIRA using patient-specific tissue classification via DECT for improved quantitative CT in dose planning. *Medical Physics*, 44(6):2345–2357, 2017.

[7] P. M. Joseph. An improved algorithm for reprojecting rays through pixel images. *IEEE transactions on medical imaging*, 1(3):192–196, 1982.

[8] C. Maaß, E. Meyer, and M. Kachelrieß. Exact dual energy material decomposition from inconsistent rays (MDIR). *Medical Physics*, 38(2):691, 2011.

[9] M. Magnusson, M. Sandborg, G. Alm Carlsson, L. Henriksson, Å Carlsson Tedgren, and A. Malusek. Accuracy of CT numbers obtained by DIRA and Monoenergetic Plus algorithms in dual-energy computed tomography. *Radiation Protection Dosimetry*, 195(3-4):212-217, 2021.

[10] M. Magnusson, G. Alm Carlsson, M. Sandborg, Å Carlsson Tedgren, and A. Malusek. On the choice of base materials for Alvarez-Macovski and DIRA dual-energy reconstruction algorithms in CT. *In: Photon Counting Computed Tomography. Ed. by Kris Iniewski and Hsieh Scott. Springer*, 2023.

[11] M. Magnusson, M. Björnfot, Å Carlsson Tedgren, G. Alm Carlsson, M. Sandborg, and A. Malusek. DIRA-3D—a model-based iterative algorithm for accurate dual-energy dual-source 3D helical CT. *Biomed. Phys. Eng. Express 5, 065005*, 2019.

[12] D. Cullen, J. Hubbell, L. Kissel EPDL97: the Evaluated Photon Data Library,'97 Version. *Livermore, CA 94550: University of California, Lawrence Livermore National Laboratory, Report No.: UCRL-50400*, 1997

Pseudo Inverse Diffusion Models for Generative CT Image Reconstruction from Low Dose Data

Matthew Tivnan, Dufan Wu, and Quanzheng Li,

***Abstract*—Score-based diffusion models have significantly advanced generative deep learning for image processing. Measurement conditioned models have also been applied to inverse problems such as CT reconstruction. However, the conventional approach, culminating in white noise, often requires a high number of reverse process update steps and score function evaluations. To address this limitations, we propose an alternative forward process in score-based diffusion models that aligns with the noise characteristics of low-dose CT reconstructions, rather than converging to white noise. This method significantly reduces the number of required score function evaluations, enhancing efficiency and maintaining familiar noise textures for radiologists, Our approach not only accelerates the generative process but also retains CT noise correlations, a key aspect often criticized by clinicians for deep learning reconstructions. In this work, we rigorously define a matrix-controlled stochastic process for this purpose and validate it through computational experiments. Using a dataset from The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC), we simulate low-dose CT measurements and train our model, comparing it with a baseline scalar diffusion process and conditional diffusion model. Our results demonstrate the superiority of our pseudo-inverse diffusion model in terms of efficiency and the ability to produce high-quality reconstructions that are familiar in texture to medical professionals in a low number of score function evaluations. This advancement paves the way for more efficient and clinically practical diffusion models in medical imaging, particularly beneficial in scenarios demanding rapid reconstructions or lower radiation exposure.**

***Index Terms*—Score Based Generative Models, Diffusion Models, Tomographic Image Reconstruction, Inverse Problems, Deep Learning**

I. Introduction

Score-based diffusion models have been introduced as a powerful framework for deep generative modeling and have revolutionized the field of image generation [1]. These models are based on a forward stochastic process where an image is progressively degraded in quality. Then, a score-matching neural network can be used to approximate the reverse time stochastic process, generating high-quality images from degraded inputs. This approach has been particularly successful due to its ability to model complex data distributions and generate images with high fidelity. We propose a novel forward process for score-based diffusion models that converges not to white noise, but to low-dose CT reconstructed images. This approach offers a more direct and efficient path to high-quality image generation. Our method reduces the computational burden, requiring fewer evaluations of the score function. Additionally, the intermediate images correspond to reconstructions from intermediate radiation dose levels in terms of the noise correlations and magnitude.

There have been many recent studies investigating the application of diffusion models to solving noisy inverse problems. Conditional diffusion models leverage an auxiliary input for measurements to facilitate the generation of high-quality images from lower-quality measurements [2]. However, these models typically conclude their diffusion process in white noise, contrasting our method which effectively ends with a pseudo inverse low dose reconstruction. Another method for measurement conditioned diffusion for solving inverse problems is Diffusion Posterior Sampling (DPS) [3]. DPS has been successfully applied to CT reconstruction in various studies [4], [5], [6], presenting a framework that combines a pre-trained prior model with a physics-driven measurement likelihood. Unlike these approaches, our model is trained using paired data and supervised learning. Instead of modifying the unconditional reverse process with a measurement likelihood term, our method commences the reverse process directly with reconstructed images to produce posterior samples. Come-Closer-Diffuse-Faster (CCDF) approach addresses a critical downside of traditional diffusion models — their inherently slow sampling rate [7]. CCDF circumvents this by initiating the reverse process from a point that is not pure Gaussian noise, but a noisy version of the target image, thereby reducing the number of sampling steps required. While this approach shows improvements in efficiency, it relies on an initialization which is the ground truth plus white noise, which is not immediately available. In our method, we use the same noise correlations inherent in CT reconstructions. This distinction underlines the efficiency and practical applicability of our method in the medical imaging domain.

Our work is most closely aligned with our previous work on Fourier Diffusion Models [8]. These models replace scalar operations in the forward process with shift-invariant convolutions and additive stationary noise, enabling control over Modulation Transfer Function (MTF) and Noise Power Spectrum (NPS). However, Fourier Diffusion Models are restricted to shift invariant blur and stationary noise, which is not realistic for CT. In this work, we extend this concept to encompass realistic noise correlations found in CT reconstructions.

II. Methods

A. Nonlinear CT Measurement Likelihood Model

A statistical model of CT measurements, $\mathbf{Y}$, given true attenuation images, $\mathbf{X}$, is defined by the following conditional probability density function

$$p(\mathbf{y}|\mathbf{x}) = \mathcal{P}\Big(\mathbf{y}; \bar{\mathbf{y}}(\mathbf{x})\Big) \tag{1}$$

$$\bar{\mathbf{y}}(\mathbf{x}) = I_0 \exp\Big(-\mathbf{A}\mathbf{x}\Big) \tag{2}$$

where $\mathbf{x} \in \Re^{(N\times 1)}$ is the true image of linear attenuation coefficients, $\mathbf{A} \in \Re^{(M\times N)}$ is the forward projector, which we assume is a full-rank matrix with $M > N$, $I_0 \in \Re$ is the expected number of incident photons per pixel per view, $\mathbf{y} \in \Re^{(M\times 1)}$ is the vector of CT measurements assuming an ideal photon counting detector, and $\mathcal{P}(\mathbf{y}; \bar{\mathbf{y}}(\mathbf{x}))$

All authors were with the Department of Radiology, Massachusetts General Hospital and Harvard Medical School and , Boston, MA, 02114 USA.

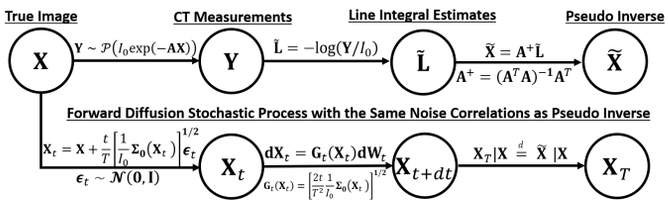


Fig. 1. Probabilistic graphical model of pseudo inverse CT reconstruction and forward diffusion stochastic process. All probability density functions displayed in this figure are rigorously defined in the sections to follow. The final time step of the forward diffusion process has the same distribution as pseudo inverse reconstructions, so the reverse diffusion process can be initialized with the pseudo inverse.

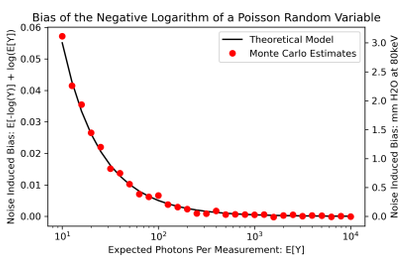


Fig. 2. Expected error due to negative logarithm applied to a Poisson distributed random variable. Theoretical model of expected error was computed via numerical integration of the log Poisson distribution. Monte Carlo estimates were computed via sample expectation of 10,000 Poisson samples.

is the probability density function of a jointly independent Poisson random vector with expectation, $\bar{\mathbf{y}}(\mathbf{x}) \in \Re^{(M\times 1)}$ and covariance matrix, $D\{\bar{\mathbf{y}}(\mathbf{x})\}$.

B. Line Integral Estimation with Negative Logarithm

It is a common preprocessing step to estimate attenuation line integrals by taking the negative logarithm of the gain corrected measurements, $\tilde{\mathbf{L}} = -\log(\mathbf{Y}/I_0)$. In our probabilistic framework, this deterministic data processing operation is represented by the following Dirac delta distribution,

$$p(\tilde{\boldsymbol{\ell}}|\mathbf{y}) = \delta\Big(\tilde{\boldsymbol{\ell}} - (-\log(\mathbf{y}/I_0))\Big). \tag{3}$$

We can analyze the mean signal and noise covariance transfer through the nonlinear elementwise logarithm operation using a first order Taylor approximation centered on $\mathbf{y}/I_0$. Using second order statistics, we can approximate the random vector of unitless line integral estimates with the following conditional Gaussian probability density function,

$$p(\tilde{\boldsymbol{\ell}}|\mathbf{x}) \approx \mathcal{N}\Big(\tilde{\boldsymbol{\ell}}; \mathbf{A}\mathbf{x}, D\{1/I_0 \exp(-\mathbf{A}\mathbf{x})\}\Big). \tag{4}$$

This approximation is appropriate for many practical applications in CT imaging. Figure 1 shows that the noise induced bias for unitless attenuation line integral estimates is less than 0.01 if the expected photon counts are greater than 100.

C. Pseudo Inverse CT Image Reconstruction

For the purposes of this work, we assume the CT reconstruction algorithm exactly implements the Moore-Penrose pseudo inverse, $\tilde{\mathbf{X}} = \mathbf{A}^{+}\tilde{\mathbf{L}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\tilde{\mathbf{L}}$. Since $\mathbf{A}$ is defined as full-rank and $M > N$, we know that $\mathbf{A}^T\mathbf{A}$ is invertible. In the future, we are interested to extend our methods to include sparse CT acquisitions where $\mathbf{A}^T\mathbf{A}$ is not invertible, but we take this well conditioned case as a starting point. The pseudo inverse reconstruction is a deterministic function of line integrals, so it is modeled by the Dirac delta distribution,

$$p(\tilde{\mathbf{x}}|\tilde{\boldsymbol{\ell}}) = \delta\Big(\tilde{\mathbf{x}} - \mathbf{A}^{+}\tilde{\boldsymbol{\ell}}\Big) = \delta\Big(\tilde{\mathbf{x}} - (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\tilde{\boldsymbol{\ell}}\Big). \tag{5}$$

Further, the pseudo inverse is a linear operator, and the input is a Gaussian distributed random vector defined in (4), so the output is also a Gaussian distributed random vector given by the conditional probability density function,

$$p(\tilde{\mathbf{x}}|\mathbf{x}) \approx \mathcal{N}\Big(\tilde{\mathbf{x}}; \mathbf{x}, \frac{1}{I_0}\boldsymbol{\Sigma}_0(\mathbf{x})\Big) \tag{6}$$

$$\boldsymbol{\Sigma}_0(\mathbf{x}) = \mathbf{A}^{+} D\{1/\exp(-\mathbf{A}\mathbf{x})\}\mathbf{A}^{+T} \tag{7}$$

This end-to-end image quality model describes the correlated noise in pseudo inverse CT reconstructed images as a function of the dose, I_0, and the object-dependent spatial correlations as defined by, $\boldsymbol{\Sigma}_0(\mathbf{x})$. Note that the object dependent statistical weights and corresponding noise correlations can be well approximated by the raw measurements; that is, $\boldsymbol{\Sigma}_0(\mathbf{x}) \approx \boldsymbol{\Sigma}_0(\tilde{\mathbf{x}}) = \boldsymbol{\Sigma}_0(\mathbf{A}^{+}\mathbf{y})$, which is a necessary property for our diffusion model derivations and similar to approximations made in several previous works [9] [10].

D. Forward and Reverse Diffusion Stochastic Process

In this section, we define the continuous time stochastic process, $\mathbf{X}_t$ that begins with true images at $\mathbf{X}_0 = \mathbf{X}$ and ends with the same expectation and covariance as the pseudo inverse reconstructed images, $\mathbf{X}_T = \tilde{\mathbf{X}}$. To that end, we define the forward stochastic process as follows

$$\mathbf{X}_t|\mathbf{X}_0 = \mathbf{X}_0 + \frac{t}{T}[\frac{1}{I_0}\boldsymbol{\Sigma}_0(\mathbf{X}_0)]^{1/2}\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t = \frac{\mathbf{W}_t}{\sqrt{t}} \tag{8}$$

where $\boldsymbol{\Sigma}_0(\mathbf{X}_0)$ are the object dependent noise correlations defined in the previous section and $\mathbf{W}_t$ is the standard Wiener process with zero mean and covariance, $t\,\mathbf{I}$, so $\boldsymbol{\epsilon}_t$ is zero mean and identity covariance white noise. Therefore, the noise standard deviation is proportional to t. The corresponding probability density function is

$$p(\mathbf{x}_t|\mathbf{x_0}) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_0, \boldsymbol{\Sigma}_t) \tag{9}$$

$$\boldsymbol{\Sigma}_t = \frac{t^2}{T^2}\frac{1}{I_0}\boldsymbol{\Sigma}_0(\mathbf{x}_0) \tag{10}$$

This forward diffusion stochastic process can be well approximated by the following stochastic differential equation, using Itô calculus notation,

$$\mathbf{d}\mathbf{X}_t = \mathbf{G}_t(\mathbf{X}_t)\mathbf{d}\mathbf{W}_t \tag{11}$$

$$\mathbf{G}_t(\mathbf{X}_t) = [\frac{2t}{T^2 I_0}\boldsymbol{\Sigma}_\mathbf{0}(\mathbf{X}_t)]^{1/2} \approx [\frac{2t}{T^2 I_0}\boldsymbol{\Sigma}_\mathbf{0}(\mathbf{X}_0)]^{1/2} \tag{12}$$

where $\boldsymbol{\Sigma}_\mathbf{0}(\mathbf{X}_t) \approx \boldsymbol{\Sigma}_\mathbf{0}(\mathbf{X}_0)$ is a commonly used approximation for the object-dependent statistical weights and noise correlations as stated in the previous section. The corresponding conditional probability density function is

$$p(\mathbf{x}_{t+dt}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t+dt}; \mathbf{x}_t, \frac{2\,t\,dt}{T^2 I_0}\boldsymbol{\Sigma}_\mathbf{0}(\mathbf{X}_t)]^{1/2}) \tag{13}$$

Following the general framework for score based generative models in [1], based on the formula originally proposed by [11], we can write the reverse diffusion stochastic process as

$$\mathbf{d}\mathbf{X}_t = -\nabla\cdot\mathbf{G}_t(\mathbf{X}_t)\mathbf{G}_t(\mathbf{X}_t)^T dt \tag{14}$$
$$-\mathbf{G}_t(\mathbf{X}_t)\mathbf{G}_t(\mathbf{X}_t)^T\nabla\log p(\mathbf{X}_t)]dt + \mathbf{G}_t(\mathbf{X}_t)\mathbf{d}\bar{\mathbf{W}}_t.$$

We can ignore the divergence term, $\nabla \cdot \mathbf{G}_t(\mathbf{X}_t)\mathbf{G}_t(\mathbf{X}_t)^T$, because the derivative with respect to $\mathbf{X}_t$ is nearly zero, using the same assumption previously used in (12) that $\mathbf{G}_t(\mathbf{X}_t) \approx \mathbf{G}_t(\mathbf{X}_0)$. We also substitute a score-matching neural network, $\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{x_t}, t) \approx \nabla \log p(\mathbf{X}_t)$, for the score function

$$\mathbf{d}\mathbf{X}_t = -\mathbf{G}_t(\mathbf{X}_t)\mathbf{G}_t(\mathbf{X}_t)^T \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{X_t}, t)dt + \mathbf{G}_t(\mathbf{X}_t)\mathbf{d}\bar{\mathbf{W}}_t \quad (15)$$

This reverse process can be initialized directly with pseudo inverse reconstructed images, $\tilde{\mathbf{X}}$, and running the reverse process from $t = T$ to $t = 0$ will result in a sample from the posterior distribution $\mathbf{X}|\tilde{\mathbf{X}}$. One important feature of our model is that the intermediate time points correspond to intermediate dose levels. This could be useful to control the tradeoff between noise and hallucinations through early stopping of the reverse process.

Algorithm 1 Pseudo Inverse Diffusion Training

Input: T, K, I_0, $p(\mathbf{x})$
Initialize $\boldsymbol{\theta}$
for each epoch **do**
  for each sample **do**
    $\mathbf{X}_0 \sim p(\mathbf{x})$
    $t \sim \mathcal{U}(0, T)$
    $\tilde{\boldsymbol{\epsilon}}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ # projection domain white noise
    $\mathbf{X}_t \leftarrow \mathbf{X}_0 + \frac{t}{T}\sqrt{\frac{1}{I_0}}\mathbf{A}^+ D\{\frac{1}{\sqrt{\exp(-\mathbf{A}\mathbf{X}_0)}}\}\tilde{\boldsymbol{\epsilon}}_t$
    compute $\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{X}_t, t)$
    optimizer step on training loss using (18)
  end for
end for
Return: $\boldsymbol{\theta}$

Algorithm 2 Pseudo Inverse Diffusion Sampling (SDE)

Input: T, Δt, I_0, $\mathbf{Y}$
$\tilde{\mathbf{L}} \leftarrow -\log(-\mathbf{Y}/I_0)$
$\tilde{\mathbf{X}} \leftarrow \mathbf{A}^+\tilde{\mathbf{L}}$
$\mathbf{X}_T \leftarrow \tilde{\mathbf{X}}$
for each n in [N,0) **do**
  $\mathbf{G}_t(\mathbf{X}_t)\mathbf{G}_t(\mathbf{X}_t)^T \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{X_t}, t) \leftarrow \frac{2}{t}(\mathbf{X}_t - \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{X}_t, t))$
  $\mathbf{d}\tilde{\mathbf{W}}_t \sim \mathcal{N}(\mathbf{0}, \Delta t\mathbf{I})$ # projection domain white noise
  $\mathbf{G}_t(\mathbf{X}_t)\mathbf{d}\bar{\mathbf{W}}_t \leftarrow \sqrt{\frac{2t}{T^2 I_0}}\mathbf{A}^+ D\{\frac{1}{\sqrt{\exp(-\mathbf{A}\mathbf{X}_t)}}\}\mathbf{d}\tilde{\mathbf{W}}_t$
  $\Delta\mathbf{X}_t \leftarrow -\mathbf{G}_t(\mathbf{X}_t)\mathbf{G}_t(\mathbf{X}_t)^T \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{X_t}, t)\Delta t + \mathbf{G}_t(\mathbf{X}_t)\mathbf{d}\bar{\mathbf{W}}_t$
  $\mathbf{X}_{t-\Delta t} \leftarrow \mathbf{X}_t - \Delta\mathbf{X}_t$
end for
Return: $\mathbf{X}_0$

E. Training and Sampling Algorithms

The score function for a Gaussian random vector is defined as $\nabla \log p(\mathbf{X}_t) = -\boldsymbol{\Sigma}_t^{-1}(\mathbf{X}_t - \mathrm{E}[\mathbf{X}_t])$, so we use the following parameterization of the score matching neural network

$$\mathbf{s}_{\boldsymbol{\theta}}(\mathbf{X}_t, t) = -\boldsymbol{\Sigma}_t^{-1}(\mathbf{X}_t - \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{X}_t, t)) \quad (16)$$

Therefore, the first term of (15) can be written as

$$-\mathbf{G}_t(\mathbf{X}_t)\mathbf{G}_t(\mathbf{X}_t)^T \mathbf{s}_{\boldsymbol{\theta}}(\mathbf{X_t}, t) = \frac{2}{t}(\mathbf{X}_t - \boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{X}_t, t)) \quad (17)$$

The advantage of this parameterization is that all $\boldsymbol{\Sigma}_0(\mathbf{X}_t)$ and I_0 are cancelled out. The training loss function is

$$E_{\mathbf{X_0}, \mathbf{X_t}, t}\Big[(\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{X}_t, t)) - \mathbf{X}_0))^T \boldsymbol{\Sigma}_t^{-K}(\boldsymbol{\mu}_{\boldsymbol{\theta}}(\mathbf{X}_t, t)) - \mathbf{X}_0))\Big] \quad (18)$$

For an L2 norm on mean estimates, one should set $K = 0$ which results in uniform time dependent weights. For an L2 norm on score estimates, one should set $K = 2$ which will result in time dependent weights proportional to $1/t^4$. Experimentally, we have found this places too small of a weight on times near $t = T$, and we choose the compromise $K = 1$, which is log-likelihood loss with $1/t^2$ time dependence. The training algorithm is summarized in Algorithm 1.

The second term of (15) requires a large matrix square root which would be computationally expensive. We propose an exact yet efficient formula to compute this stochastic term,

$$\mathbf{G}_t(\mathbf{X}_t)\mathbf{d}\bar{\mathbf{W}}_t = \sqrt{\frac{2t}{T^2 I_0}}\mathbf{A}^+ D\{\frac{1}{\sqrt{\exp(-\mathbf{A}\mathbf{X}_t)}}\}\mathbf{d}\tilde{\mathbf{W}}_t \quad (19)$$

where $\mathbf{d}\tilde{\mathbf{W}}_t \in \Re^{(M\times 1)}$ is projection domain white noise with zero mean and covariance, $dt\mathbf{I}$. This represents applying elementwise statistical weights to the white noise in the projection domain followed by pseudo inverse reconstruction and scaling by time and dose. This formula is exact because the left and right sides of the equation are image domain Gaussian random vectors with zero mean and covariance, $dt\mathbf{G}_t(\mathbf{X}_t)\mathbf{G}_t(\mathbf{X}_t)^T$. A similar approach can be used to sample from $p(\mathbf{x}_t|\mathbf{x_0})$ during training as follows

$$\mathbf{X}_t|\mathbf{X_0} = \mathbf{X}_0 + \frac{t}{T}\sqrt{\frac{1}{I_0}}\mathbf{A}^+ D\{\frac{1}{\sqrt{\exp(-\mathbf{A}\mathbf{X}_0)}}\}\tilde{\boldsymbol{\epsilon}}_t \quad (20)$$

where $\tilde{\boldsymbol{\epsilon}}_t$ is zero-mean identity-covariance white noise in the projection domain. The sampling algorithm, using the Euler-Maruyama method, is shown in Algorithm 2.

F. Computational Experiment

We conducted a comprehensive evaluation of our proposed method using a dataset from The Cancer Genome Atlas Liver Hepatocellular Carcinoma (TCGA-LIHC). This dataset consists of normal dose abdominal CT images. We extracted 8000 training samples and 2000 validation samples of 256x256 axial cross sections 1.6mm voxel spacing which were treated as the ground truth. We simulated low dose CT measurements using the LEAP CT forward projector [12]. We set the number of view angles to 720 covering 360 degrees and modeled a single row of pixels with 1024 columns for the detector array. The pixel spacing was 2.0mm in both height and width. . The source-to-object distance was set at 600.0 mm, and the source-to-detector distance was 1200.0 mm. Projection domain Poisson was added corresponding to 10,000 photons per pixel per view. For the score matching neural network, we used an open source convolutional U-Net with 32 base channels, a depth of 4 down sampling layers, and channelwise self attention modules in the two deepest layers. We trained the model with a learning rate of 0.001 with the Adam optimizer for 1000 epochs and 128 batches per epoch and 16 images per batch [13]. For a baseline, we trained a conditional diffusion model following the methods in [2]. We trained the pseudo inverse diffusion models using Algorithm 1.



Fig. 3. The pseudo inverse reconstructions are generated directly from the measurements and used to initialize the reverse diffusion process. The intermediate time points have the same noise correlations but lower noise magnitude than initial pseudo inverse CT reconstruction.

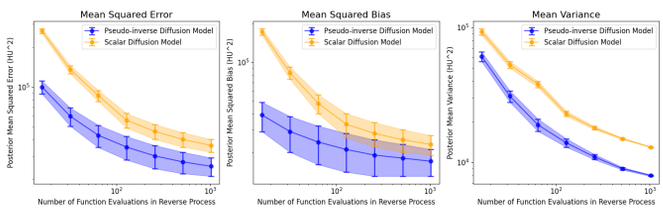


Fig. 4. Performance metrics for posterior estimates of ground truth normal dose images given simulated low dose CT reconstructions. Orange curves show conditional diffusion models using a scalar diffusion process and blue curves show the proposed pseudo inverse diffusion models. The solid line shows the mean and the error bars and filled region shows one standard deviation over all voxels and validation samples.

III. Results

The results of our computational experiments are shown in Figures 3 and 4. In Figure 3, we showcased the reverse diffusion process initiated from pseudo-inverse reconstructed images. The intermediate stages of this reverse process corresponded to varying radiation dose levels. Figure 4 presented a comprehensive quantitative analysis of the model's performance. The results show that pseudo inverse diffusion models can out perform the conditional diffusion models, particularly for a low number of function evaluations. By breaking down mean squared error into mean squared bias and mean variance components, we can see that the majority of the improvement comes from lowering the posterior variance.

IV. Discussion

The Moore-Penrose pseudo inverse reconstruction is not common due to stability concerns and computational limitations. The stability issue is related to a low condition number of $\mathbf{A}^T\mathbf{A}$. In the future, we are interested to explore penalized likelihood reconstruction using $(\mathbf{A}^T\mathbf{A} + \mathbf{R})^{-1}\mathbf{A}^T$, but this biased estimator is not compatible with our current model. The computational implementation of pseudo inverse reconstruction can also be a challenge. One method is to store $(\mathbf{A}^T\mathbf{A})^{-1}$ as a dense image domain matrix operator, as we did in our experiments, but this requires a large amount of memory. It is also possible to implement the pseudo inverse via iterative reconstruction methods, but it could require many iterations to converge. Filtered back projection (FBP) reconstruction is a promising approximation because it is computationally efficient in terms of both time and memory. Furthermore, FBP and similar analytical methods are based on the exact inverse in a spatially continuous framework [14]. In the future, we are interested to apply the FBP approximated approach in large three-dimensional images.

V. Conclusion

Our results demonstrate the efficacy of pseudo-inverse diffusion models in reconstructing CT images, particularly from low-dose data. A key aspect of our method is the ability to maintain noise correlations and only reduce noise magnitude in the reverse process such that intermediate images correspond to the noise covariance of pseudo inverse reconstructions at various radiation dose levels. This characteristic addresses a common concern among radiologists regarding the altered texture in deep learning-based reconstructions. This advancement represents a significant stride in the realm of score based generative models for CT reconstruction.

References

[1] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-based generative modeling through stochastic differential equations," *arXiv preprint arXiv:2011.13456*, 2020.

[2] G. Batzolis, J. Stanczuk, C.-B. Schönlieb, and C. Etmann, "Conditional image generation with score-based diffusion models," *arXiv preprint arXiv:2111.13606*, 2021.

[3] H. Chung, J. Kim, M. T. Mccann, M. L. Klasky, and J. C. Ye, "Diffusion posterior sampling for general noisy inverse problems," *arXiv preprint arXiv:2209.14687*, 2022.

[4] W. Xia, Y. Shi, C. Niu, W. Cong, and G. Wang, "Diffusion prior regularized iterative reconstruction for low-dose ct," *arXiv preprint arXiv:2310.06949*, 2023.

[5] S. Li, M. Tivnan, Y. Shen, and J. W. Stayman, "Diffusion posterior sampling for nonlinear ct reconstruction," *arXiv preprint arXiv:2312.01464*, 2023.

[6] C. Huang, P. Vaska, Y. Gao, S. Chang, T. W. Holmes, A. Pourmorteza, and J. Liang, "Proceedings of the 17th virtual international meeting on fully 3d image reconstruction in radiology and nuclear medicine," *arXiv preprint arXiv:2310.16846*, 2023.

[7] H. Chung, B. Sim, and J. C. Ye, "Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12 413–12 422.

[8] M. Tivnan, J. Teneggi, T.-C. Lee, R. Zhang, K. Boedeker, L. Cai, G. J. Gang, J. Sulam, and J. W. Stayman, "Fourier diffusion models: A method to control mtf and nps in score-based stochastic image generation," *arXiv preprint arXiv:2303.13285*, 2023.

[9] H. Erdogan and J. A. Fessler, "Monotonic algorithms for transmission tomography," in *5th IEEE EMBS International Summer School on Biomedical Imaging, 2002.* IEEE, 2002, pp. 14–pp.

[10] S. Tilley, J. H. Siewerdsen, and J. W. Stayman, "Model-based iterative reconstruction for flat-panel cone-beam ct with focal spot blur, detector blur, and correlated noise," *Physics in Medicine & Biology*, vol. 61, no. 1, p. 296, 2015.

[11] B. D. Anderson, "Reverse-time diffusion equation models," *Stochastic Processes and their Applications*, vol. 12, no. 3, pp. 313–326, 1982.

[12] H. Kim and K. Champley, "Differentiable forward projector for x-ray computed tomography," *arXiv preprint arXiv:2307.05801*, 2023.

[13] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[14] A. Katsevich, "Theoretically exact filtered backprojection-type inversion algorithm for spiral ct," *SIAM Journal on Applied Mathematics*, vol. 62, no. 6, p. 2012–2026, 2002.

Assessment of Proton Relative Stopping Power From Photon-Counting CT Data

Guyue Hu, Katharina Niepel, Franka Risch, Christopher Kurz, Matthias Würl, Thomas Kröncke, Florian Schwarz, Guillaume Landry, Katia Parodi

***Abstract*—We performed photon-counting CT (PCCT) measurements with customised plastic phantoms housing tissue-equivalent inserts. To estimate quantitative information for radiation therapy treatment planning, calibration of measured CT numbers was conducted relying on a pair of low- and high-energy virtual monoenergetic images reconstructed at a PCCT scanner. Relative electron density, effective atomic number, and proton relative stopping power were estimated for phantoms of varying sizes and differently positioned inserts. The accuracy of the estimation was evaluated by root-mean-square errors (RMSEs) over the phantom inserts.**

***Index Terms*—photon-counting CT, proton therapy, relative stopping power, virtual monoenergetic images.**

I. Introduction

PHOTON-counting CT (PCCT), which provides CT data with intrinsic spectral information, is a topic of increasing interest in radiation therapy. PCCT images reconstructed at different virtual monoenergies allow us to directly estimate relative electron density (RED) and effective atomic number (EAN) of materials, and subsequently calculate the proton relative stopping power (RSP). In this work, we assessed the quantitative determination of different physical quantities relevant for radiation therapy, namely RED, EAN, and RSP, using virtual monoenergetic images (VMIs) of phantoms containing tissue-equivalent inserts acquired at a clinical PCCT scanner.

II. Materials and methods

A. Phantom scans

Customised cylindrical plastic phantoms housing tissue-equivalent inserts were scanned at a first-generation clinical PCCT scanner (NAEOTOM Alpha, Siemens Healthineers, Forchheim, Germany), using scan parameters as listed in table I.

TABLE I
Photon-counting CT scan protocol.

Tube voltage	120 kVp
Acquisition mode	normal resolution
$CTDI_{vol}$	20.1 mGy
Collimation width	19.2 mm
Spiral pitch factor	0.7
Exposure	102 mAs
Tube current	142 mA

A cylindrical polymethyl methacrylate (PMMA) phantom with a diameter of 15 cm housing twelve Gammex 467 inserts (Gammex, Inc., Middleton, WI, USA) along the central axis of the cylinder was used as a calibration phantom (*Gammex PMMA* phantom). A similar phantom with a diameter of 13 cm containing seven CIRS 062M inserts (CIRS, Inc., Norfolk, VA, USA), positioned both at the phantom centre and at a given radial distance to the centre, was used for evaluation (*CIRS* phantom). In addition, the Gammex 467 inserts were placed in cylindrical polyoxymethylene (POM) phantoms with different diameters (15, 20, and 25 cm). For each diameter, two phantoms, with inserts at the phantom centre and at a given radial position, were used. Figure 1 shows the two different position configurations of the inserts, along the central axis of the phantom for three phantoms of different diameters (*Gammex C15*, *Gammex C20*, and *Gammex C25* phantoms), and at a same radial distance from the centre for another three phantoms of different diameters (*Gammex R15*, *Gammex R20*, and *Gammex R25* phantoms).

For each phantom, virtual monoenergetic images (VMIs) at energies ranging from 40 to 190 keV in 10 keV increments were reconstructed with the Qr36f kernel at the scanner, resulting in 512×512-pixel images with $0.5 \times 0.5\ \mathrm{mm}^2$ pixel size and 1 mm slice thickness.

B. Conversion of CT numbers into RSP

1) Calibration: In our study, calibration was conducted on two phantoms, *Gammex PMMA* (diameter 15 cm) and *Gammex C25* (diameter 25 cm), respectively, in order to test the impact of the calibration phantom in two extreme scenarios.

Using a pair of low- and high-energy VMIs of the chosen calibration phantom, the mean measured CT numbers were calculated for the inserts. Following the method reported in [1], RED, denoted by ρ_e, was calibrated against a weighted

This work was supported by German Research Foundation (Research Training Group GRK2274).

G. Hu, K. Niepel, and K. Parodi are with the Department of Medical Physics, Faculty of Physics, LMU Munich, Garching, Germany, e-mail: Guyue.Hu@physik.uni-muenchen.de.

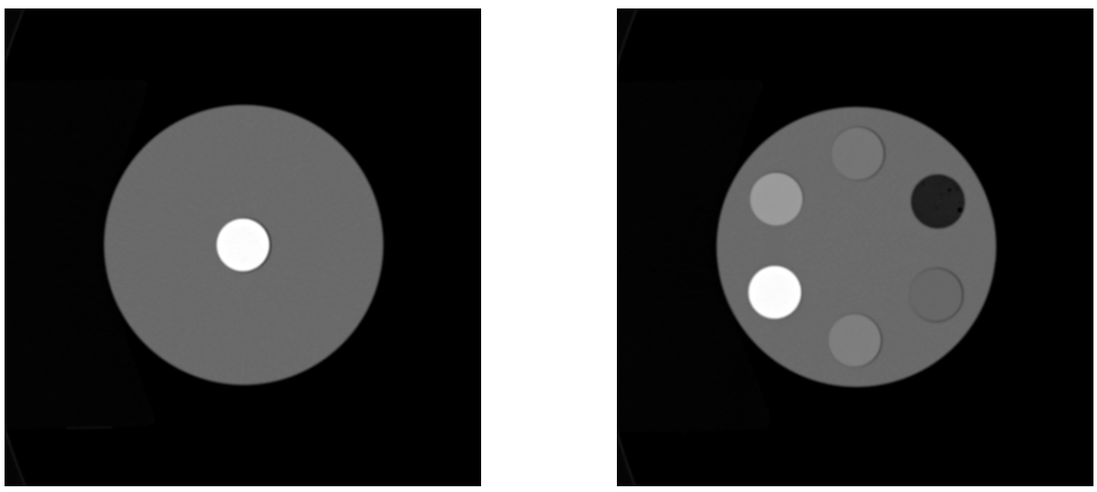
F. Risch and T. Kröncke are with the Department of Diagnostic and Interventional Radiology, Universitätsklinikum Augsburg, Augsburg, Germany.

C. Kurz and G. Landry are with the Department of Radiation Oncology, LMU Klinikum, Munich, Germany.

M. Würl was with the Department of Medical Physics, Faculty of Physics, LMU Munich, Garching, Germany.

F. Schwarz was with the Department of Diagnostic and Interventional Radiology, Universitätsklinikum Augsburg, Augsburg, Germany.

G. Landry and K. Parodi are also with German Cancer Consortium (DKTK), Partner Site Munich, Munich, Germany.



(a) (b)

Fig. 1. CT images of the two POM phantoms with a diameter of 15 cm and two insert position configurations: (a) central inserts and (b) radial inserts, reconstructed at a virtual monoenergy of 60 keV. Viewing window is set between HU of -1000 (black) and 1200 (white). The same position configurations, i.e., inserts at the phantom centre and inserts at a fixed radial position same as in (b), were also used for the two phantoms with 20 cm diameter and the two with 25 cm diameter.

difference between the low- and the high-energy CT number given in Hounsfield units:

$$\rho_{\mathrm{e}} = a\frac{(1+\alpha)\mathrm{HU_H} - \alpha\mathrm{HU_L}}{1000} + b. \quad (1)$$

The ratio of EAN of the Gammex insert materials to that of water ($Z_{\mathrm{eff}}/Z_{\mathrm{eff,w}}$) was fitted to the reduced CT number $u_{\mathrm{L}} = \mathrm{HU_L}/1000 + 1$ and RED [2]:

$$\left(\frac{Z_{\mathrm{eff}}}{Z_{\mathrm{eff,w}}}\right)^{3.3} - 1 = \gamma_{\mathrm{L}}\left(\frac{u_{\mathrm{L}}}{\rho_{\mathrm{e}}} - 1\right) + \gamma_0. \quad (2)$$

Mean excitation energy (I) was calibrated against EAN for soft-tissue and bone-tissue inserts, respectively [3]:

$$\ln\frac{I}{I_{\mathrm{w}}} = C_1^{\mathrm{soft/bone}}\left[\left(\frac{Z_{\mathrm{eff}}}{Z_{\mathrm{eff,w}}}\right)^{3.3} - 1\right] - C_0^{\mathrm{soft/bone}}. \quad (3)$$

For calibration, the RED, EAN, and mean excitation energy values were calculated with the known mass density and elemental composition of the inserts. The fit parameters a, b, α, γ_0, γ_1, $C_0^{\mathrm{soft/bone}}$, and $C_1^{\mathrm{soft/bone}}$ were obtained by least-squares fit.

2) Evaluation: For all the phantoms, RED, EAN, and mean excitation energy were estimated based on calibration of the measured CT numbers of the Gammex inserts. The RSP maps were calculated from RED and mean excitation energy using the Bethe-Bloch formula [3]:

$$\mathrm{RSP} = \rho_{\mathrm{e}}\frac{\ln\frac{2m_{\mathrm{e}}c^2\beta^2}{I(1-\beta^2)} - \beta^2}{\ln\frac{2m_{\mathrm{e}}c^2\beta^2}{I_{\mathrm{w}}(1-\beta^2)} - \beta^2}. \quad (4)$$

C. Evaluation of RED/EAN/RSP accuracy

To evaluate the accuracy of the estimation, we calculated the mean estimated RED/EAN/RSP values within a region of interest covering the central volume of each insert. The mean RED and EAN values were compared to reference values calculated from vendor-provided mass densities and elemental compositions of the inserts. For RSP accuracy evaluation, the reference values were obtained through Bragg peak measurements [4]. RMSEs of the estimated RED/EAN/RSP were evaluated over the phantom inserts.

III. Results

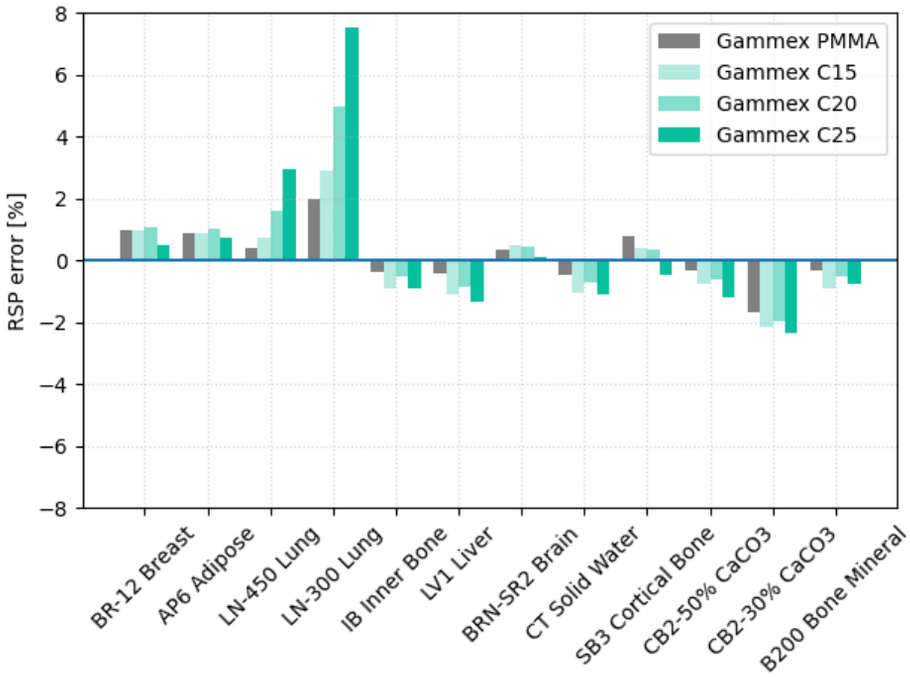
Figure 2 shows comparison of insert RSP errors between Gammex phantoms of different sizes. For both central inserts (figure 2a) and radial inserts (figure 2b), the estimated RSP values mostly agree with the reference values within 1%, except for the two lung inserts.

Table II shows RMSEs of RSP evaluated over the phantom inserts for different phantoms based on calibration results of the *Gammex PMMA* phantom and the *Gammex C25* phantom, respectively. The lung inserts showed relatively high errors, as observed in previous studies [4][5]. Therefore, the RMSEs were also calculated excluding the lung inserts.

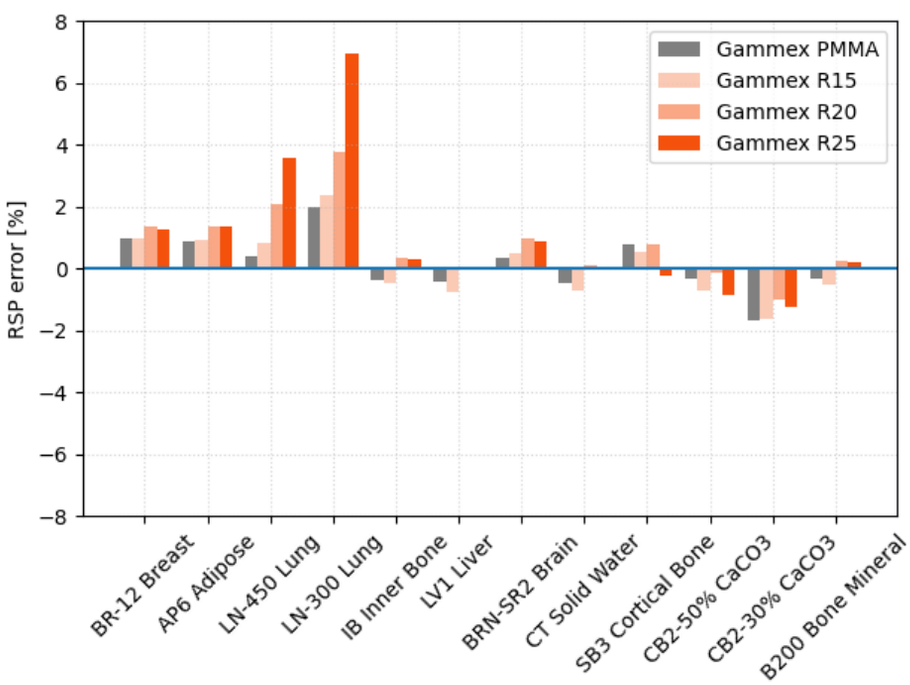
The two calibration phantoms contain the same sets of Gammex inserts, both centrally positioned, but have a difference in diameter, with 15 cm for *Gammex PMMA* and 25 cm for *Gammex C25*. They both yielded an RSP accuracy better than 0.8%, shown in bold in table II, when evaluated using the self-calibration results. Compared to calibration on *Gammex PMMA*, the RSP accuracy for the three phantoms containing radial inserts is slightly worse using calibration with *Gammex C25*. For the three phantoms with central inserts, similar RMSEs of RSP around 1% were obtained using the two different calibration phantoms.

IV. Discussion

Comparing the results for RSP estimation between the two calibration phantoms of different sizes, the *Gammex PMMA*



(a)



(b)

Fig. 2. RSP errors of (a) the central inserts and (b) the radial inserts for phantoms of different sizes, calibrated on the *Gammex PMMA* phantom.

phantom, with a diameter of 15 cm, showed in general better performance, leading to slightly more accurate determination of RSP.

V. Conclusion

We have performed measurements of tissue-equivalent materials at a first-generation clinical PCCT scanner. Using the VMIs provided by the scanner, the RED, EAN, and proton RSP maps were estimated for phantoms of different sizes and with differently positioned inserts. When excluding lung inserts, an RSP accuracy of around 1% was achieved irrespective of the size of the calibration phantom. This work has so far only studied tissue-equivalent materials. Therefore, account

TABLE II
RMSEs OF RSP FOR DIFFERENT PHANTOMS, CALCULATED OVER ALL THE PHANTOM INSERTS ('RMSE' COLUMNS) AND EXCLUDING LUNG INSERTS ('RMSE*' COLUMNS).

Calibration	Gammex PMMA		Gammex C25	
	RMSE	RMSE*	RMSE	RMSE*
Gammex PMMA	0.92%	**0.78%**	1.23%	0.98%
CIRS	0.70%		1.16%	
Gammex R15	1.05%	0.83%	1.12%	0.91%
Gammex R20	1.45%	0.81%	1.16%	1.24%
Gammex R25	2.39%	0.82%	1.25%	1.03%
Gammex C15	1.30%	1.06%	1.08%	0.98%
Gammex C20	1.73%	0.92%	0.87%	0.94%
Gammex C25	2.55%	1.11%	1.15%	**0.79%**

should be taken of the need for further validation of the RSP estimation approach with biological tissues.

REFERENCES

[1] M. Saito. Potential of dual-energy subtraction for converting CT numbers to electron density based on a single linear relationship. *Medical Physics* 39(4):2021-2030, 2012.

[2] M. Saito, S. Sagara. A simple formulation for deriving effective atomic numbers via electron density from dual-energy CT data in the human body. *Medical Physics* 44(6):2293-2303, 2017.

[3] M. Saito, S. Sagara. Simplified derivation of stopping power ratio in the human body from dual-energy ct data. *Medical Physics* 44:4179–87, 2017.

[4] N. Hudobivnik, F. Schwarz, T. Johnson, L. Agolli, G. Dedes, T. Tessonnier, et al. Comparison of proton therapy treatment planning for head tumors with a pencil beam algorithm on dual and single energy CT images. *Medical Physics* 43:495, 2016.

[5] G. Landry, F. Dörringer, S. Si-Mohamed, P. Douek, J. Abascal, F. Peyrin, et al. Technical note: Relative proton stopping power estimation from virtual monoenergetic images reconstructed from dual-layer computed tomography. *Medical Physics* 46:1821–8, 2019.

Sparse2Inverse: Self-supervised inversion of sparse-view CT data

Nadja Gruber, Johannes Schwab, Elke Gizewski and Markus Haltmeier

***Abstract*—Sparse-view computed tomography (CT) enables fast and low-dose CT imaging, an essential feature for patient-save medical imaging and rapid non-destructive testing. In sparse-view CT, only a few projection views are acquired, causing standard reconstructions to suffer from severe artifacts and noise. To address these issues, we propose a self-supervised image reconstruction strategy. Specifically, in contrast to the established Noise2Inverse, our proposed training strategy uses a loss function in the projection domain, thereby bypassing the otherwise prescribed nullspace component. We demonstrate the effectiveness of the proposed method in reducing stripe-artifacts and noise, even from highly sparse data.**

***Index Terms*—Self-supervised reconstruction, Sparse-view CT, deep learning.**

I. Introduction

Image reconstruction from sparse CT measurements suffers from data incompleteness and a large null space, resulting in severe non-uniqueness. In addition, the inversion is unstable, causing noise in the data to be amplified during the reconstruction process. These issues result in reconstructed images containing severe streak-artifacts and noise. Although conventional filtered back projection (FBP) is fast and accurate for full data, it produces low-quality images dominated by noise and artifacts in sparse-view CT. This highlights the importance of regularization that incorporates specific prior knowledge for high-quality reconstruction. Recently, machine learning methods have significantly improved reconstruction quality and have gained considerable popularity in this context [1], [2], [3], [4].

Classical variational and iterative reconstruction methods allow for the efficient incorporation of hand-crafted priors such as smoothness or total variation [5]. While accurate in many scenarios, they rely on idealized prior assumptions, are time consuming, and are highly sensitive to the choice of regularization parameters. Recently, deep learning methods have gained popularity because they naturally adapt to the available image structure and produce high quality reconstructions. However, these methods typically require supervised training data with specific input-target pairs tailored to a particular tomographic setting.

N. Gruber is with VASCage-Centre on Clinical Stroke Research, Innsbruck, Austria, e-mail: nadja.gruber@vascage.at.

N. Gruber and M. Haltmeier are with the Department of Mathematics, University of Innsbruck, Austria, e-mail: (markus.haltmeier)nadja.gruber@uibk.ac.at.

Elke Gizewski is with the Department of Neuroradiology, Medical University of Innsbruck, Austria, e-mail: elke.gizewski@i-med.ac.at.

J. Schwab is with MRC-LMB, Cambridge, United Kingdom.

Manuscript received April 19, 2005; revised August 26, 2015.

Self-supervised learning combines the advantages of both iterative and learning-based methods and has recently become an important direction in deep learning and image processing. It overcomes the need for data annotation while exploiting the power of learnable neural networks. Self-supervised methods developed for denoising tasks exploit inherent structure and redundancy in the data to remove noise from signals and images. Self-supervised methods can also be extended to reconstruct missing or corrupted parts of a signal or image, where the goal is to fill in missing information or to restore the original signal from noisy or incomplete observations. Existing studies on self-supervised CT image reconstruction operate either entirely in the image domain [6] or entirely in the data domain [7], [8]. Methods that operate only in the image domain do not perform well for operators with a large null space, such as sparse-view CT. On the other hand, methods that operate only in the data domain may not perform optimally because convolutional neural networks are built for image structure rather than projection data.

In this work, we propose a self-supervised learning strategy for image reconstruction in sparse-view CT. Our method allows the CNN to operate in the image domain and combine it with a loss in the data domain. In contrast to Noise2Inverse, our method can handle a large null space, allowing the simultaneous removal of noise and streak-artifacts while maintaining data consistency.

II. Related Work

A. Self-supervise denoising

In [9], the authors introduce Noise2Noise, a self-supervised denoising methods using only noisy image pairs $(x_t \odot \delta_{t,1}, x_t \odot \delta_{t,2})$ for $i = 1, \dots, N$. They showed that training solely with corrupted images, assuming independent and additive noise, achieves results comparable to noisy-clean pairs. However, obtaining multiple instances of a signal with different noise may also be challenging in practice. This limitation was eliminated with Noise2Self [10], which enables self-supervised training with only noisy measurements.

In Noise2Self, pixels from the input image are (randomly) masked out and the neural network is trained to predict values of the masked pixels from the noisy unmasked pixels. Training a neural network using the Noise2Self strategy approximates training with ground truth targets if the noise has mean zero and is element-wise independent across different pixels in the image.

B. Self-supervised image reconstruction

Training without the requirement of clean targets is also valuable for image reconstruction problems of the form $y = Ax \odot \delta$ such as CT or MRI reconstruction. In this section we review existing approaches that are relevant for our strategy.

a) Image and projection domain methods: Natural self-supervised strategies for image reconstruction are to apply Noise2Self either in the image domain or in the data domain. In image domain methods, a reconstruction operator B such as Filtered Backprojection (FBP) is applied in a first step, and then Noise2Self is applied, treating BAx as the ground truth and $B\delta$ as the reconstructed noise. However, $B\delta$ is statistically coupled, which severely degrades Noise2Self. Another option is to use self-supervised training schemes in the data domain [7], [8] where the noise is statistically independent. Since neural networks tend to perform better on real images than on sinograms, this degrades performance. In addition, both methods are unable to handle image components orthogonal to the nullspace of A.

b) Noise2Inverse: Some drawbacks of image and projection domain methods are addressed by Noise2Inverse [6], which extends Noise2Self to image reconstruction in the case of complete data. Here, the data is split in the projection domain, where the noise is statistically independent, and a network is trained in the reconstruction domain. Noise2Inverse aims to remove measurement noise, but is not designed to remove artifacts resulting from under-sampling in the case of sparse data. The reason for this is that the targets in this framework are defined using FBP and thus still fix the nullspace component of the reconstructions.

c) Extensions: In [11], the authors propose a self-supervised learning strategy for the inverse problem, assuming that the underlying signal distribution is invariant to the action of a group of transformations. In [12], the authors proposed an adaptation of the Noise2Self strategy that operates in the data domain combined with a differentiable learnable backprojection. In contrast to this approach, we adapt the sampling strategy proposed in [6] to the sparse data case combined with a different training loss.

d) MRI-reconstruction: Closely related to our work are methods for self-supervised MRI reconstruction from sparsely sampled and noisy data [13], [14], [15], [16]. However, while the Fourier transform of MRI is an isometry, the Radon transform used in sparse-view CT lacks this property. In addition, the sampling pattern used in these MRI studies is Cartesian (with the exception of [16]), whereas the sparse CT data is related to radial MRI data.

III. Method

In this section we describe the proposed self-supervised learning strategy for sparse-view CT reconstruction. For the sake of simplicity, we consider the parallel beam geometry in 2D. However, the overall concept can easily be extended to more general geomeries in 2D or 3D.

A. Problem Formulation

Let $x \in X = \mathbb{R}^m$ represent a discretized ground-truth image and denote by $y = A_\Sigma x \odot \delta \in Y$ the noisy sparse view CT data at equidistant angles $\Sigma \coloneqq \{\sigma_1, \dots, \sigma_l\}$ on the semicircle corresponding to normal vectors of the measurement lines. Here $\odot\delta$ denotes impact of measurement noise which might be non-additive. We consider the situation, where the Radon data are available only for a small number of directions, meaning that Σ only has a few elements.

Following the self-supervised learning paradigm we assume that we are given a collection of noisy data

$$y_t^\delta = A_\Sigma x_t \odot \delta_t \quad \text{for } t = 1, \dots, N \tag{1}$$

without access to the clean and artifact-free ground truth images $x_t \in X$. The aim is to construct a reconstruction operator $R\colon Y \to X$ using (1), mapping noisy sparse data to reconstructions close to unknown ground-truth images.

B. Proposed Sparse2Inverse

a) Network architecture: For given numbers $k, p \in \mathbb{N}$ we define subsets of projection directions $\Theta_i = \{\sigma_j \mid j \bmod k = i\}$, consider the partition $\boldsymbol{\Theta} \coloneqq \{\Theta_1, \dots, \Theta_k\}$ and we use $\mathcal{I}$ to denote the collection of all subsets $I \subset \boldsymbol{\Theta}$ with p elements. Further let A_Θ and B_Θ denote the Radon transform and corresponding FBP with directions in Θ and write $M_\Theta y$ for the restriction of data y to directions in Θ. Inspired by [6] we take the desired reconstruction operator as a parameterized network of the form

$$R_\psi(y^\delta) \coloneqq \frac{1}{|\mathcal{I}|} \sum_{I \in \mathcal{I}} R_{\psi,I}(y^\delta) \tag{2}$$

$$R_{\psi,I}(y^\delta) \coloneqq \Phi_\psi \left(\frac{1}{p} \sum_{\Theta \in I} B_\Theta (M_\Theta y^\delta) \right), \tag{3}$$

where $\Phi_\psi\colon X \to X$ is a neural network parametrized by ψ,

The final reconstructions at inference are taken as $R_{\psi^*}(y^\delta)$ where ψ^* optimizes the loss defined next.

b) Self-supervised loss: The proposed strategy, and its comparison with the classic Noise2Inverse approach is visualized in Figure 1. Note that opposed to Noise2Inverse, we use a loss in the projection domain instead of a loss in the image domain. More precisely, for Sparse2Inverse we construct the network parameters as $\psi^* \in \arg\min_\psi \mathcal{L}_Y(\phi)$ using the loss

$$\mathcal{L}_Y(\phi) = \sum_{t=1}^{N} \sum_{I \in \mathcal{I}} \| A_{(\bigcup I)^C}(R_{\psi,I}(y_t^\delta)) - M_{(\bigcup I)^C} y_t^\delta \|_2^2. \tag{4}$$

Thus, we apply the neural network in the reconstruction domain, and evaluate the loss in the projection domain for angles not belonging to any of the sets in I for any partition $I \in \mathcal{I}$ and any training data.

C. Comparison to Noise2Inverse

Note that due to the large nullspace of the sparse-view forward operator, our loss (4) significantly differs from the Noise2Inverse loss

$$\mathcal{L}_X(\phi) = \sum_{i=1}^{N} \sum_{I \in \mathcal{I}} \| R_{\psi,I}(y_t^\delta) - B_{(\bigcup I)^C} M_{(\bigcup I)^C} y_t^\delta \|_2^2. \tag{5}$$

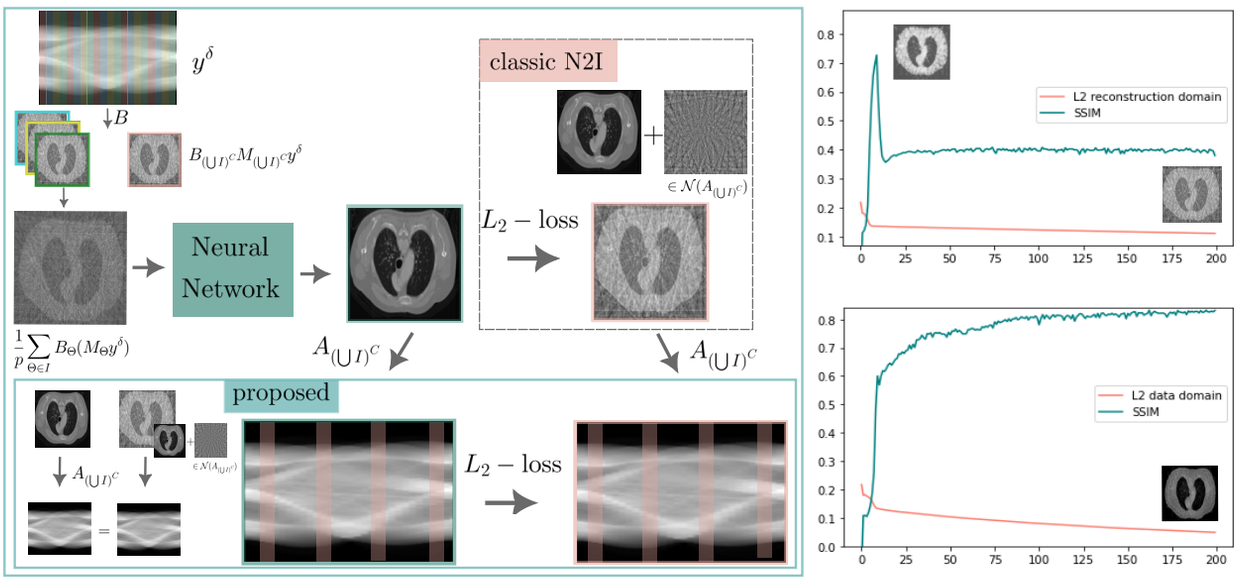


Fig. 1. The proposed training strategy of Sparse2Inverse, in comparison with the classic Noise2Inverse approach [6]. Both methods start with an initial data splitting in the projection domain, where the noise is pixel-wise independent. The neural network takes the mean sub-reconstruction as input, and the target is defined by the remaining data. Unlike the classic approach, where the loss is computed in the reconstruction domain, our strategy applies the Radon transform and computes the loss in the data domain. The effectiveness of this approach is evident in the curves on the right.

This discrepancy can be attributed to the fact that opposed to (4), the loss (5) targets $B_{(\bigcup I)^c} A_{(\bigcup I)^c}$ and thus learns streak-artifacts due to Σ. More formally, this behaviour of Noise2Inverse method can be summarized as follows: Any ground truth image x, can be decomposed as $x = x^* + v$, with $v \in \mathcal{N}(A_\Theta)$ and $x^* \in \mathcal{R}(A_\Theta^T)$. The minimum norm solution of a noise-free equation yields $B_\Theta(y) = x^*$. Looking at the Noise2Inverse approach [6], where the loss is computed in the reconstuction doamin, there is no chance for the network to learn the correct nullspace component v as it is fixed to $v = 0$. The target therefore specifies a fixed nullspace element, which the neural network Φ_Ψ aims to learn.

In the Sparse2Inverse approach, despite the neural network being fed images containing artifacts from the FBP, the optimization process of the Convolutional Neural Network (CNN) may converge to a solution with fewer artifacts that still adequately explains the measured data. Consequently, compared to the reconstruction domain loss, there is no inherent motivation for the network to alter the solution and prefer an alternative one.

IV. Experiments

The source code, data, and experiments are accessible on GitHub[1]. Our proposed method is evaluated on two datasets: a subset of 32 images from the CBCT Walnut dataset (Cone-Beam X-Ray CT Data Collection Designed for Machine Learning) [17] and 30 images from the Chest CT-Scan Images Dataset[2]. We employed the UNet architecture [18], with an image resolution of 336×336, and uniformly distributed projections between 0 and π. The neural network was trained for 2000 epochs using the Adam optimizer with a learning rate of 0.0002. Our method was compared with Noise2Inverse [6], FBP, and total variation minimization (TV). For the implementation of [6], we utilized the same UNet architecture and a learning rate of 0.0001.

Projection images underwent Poisson noise corruption with photon counts of 1000 (walnuts) and 3000 (lung CTs), following the methodology outlined in [6]. The sinograms were subsequently divided into four parts (denoted as $k = 4$). For the implementation of FBP and TV, we utilized tomosipo [19] and leveraged LION[3] for specifying the CT geometry.

In the assessment and comparison phase, we employed the Structural Similarity Index (SSIM) and Peak Signal to Noise Ratio (PSNR). To optimize the total variation (TV) minimization process, we conducted a grid search for the optimal regularization parameters λ for SSIM and PSNR. During the training of the proposed method and the classic Noise2Inverse approach, evaluation metrics were computed using the ground truth image, and reconstructions were selected to maximize SSIM and PSNR. Although the ground truth was used in this work we made the observation that SSIM and PSNR improve until the end of the training.

We further note that, during training we have observed that the SSIM of the classic Noise2Inverse reaches its maximum at a quite early point of training, and the results at that point have less artifacts than later, when the training loss is minimized. This is also underlined on the right side of Figure 1.

Figure 2 shows the reconstructions obtained from the different methods and three sparse view geometries. We observe that compared with the classical Noise2Inverse [6], the proposed approach leads to reconstructions with less streak-artifacts and finer details. In addition, the smoothing effect of the results is

[1]https://github.com/Nadja1611/Self-supervised-image-reconstruction-for-Sparse-view-CT.git

[2]https://www.kaggle.com/datasets/mohamedhanyyy/chest-ctscan-images

[3]https://github.com/CambridgeCIA/LION/tree/main

	Walnuts (Photon count 1000)						lung CT (Photon count 3000)					
	16 angles		32 angles		64 angles		16 angles		32 angles		64 angles	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
FBP	0.100	7.743	0.082	9.712	0.156	18.231	0.054	10.574	0.117	14.176	0.220	17.552
TV	**0.650**	17.882	0.788	20.867	0.838	22.606	**0.905**	**29.201**	0.925	30.613	0.935	31.627
Noise2Inverse [6]	0.408	15.402	0.500	18.800	0.736	21.793	0.586	25.775	0.673	28.356	0.743	30.663
Proposed	**0.650**	**19.103**	**0.798**	**22.347**	**0.857**	**24.373**	0.860	28.088	**0.928**	**31.394**	**0.956**	**34.006**

TABLE I
EVALUATION METRICS SSIM AND PSNR OF THE RECONSTRUCTIONS OBTAINED ON THE TWO DATASETS FOR THE DIFFERENT METHODS.

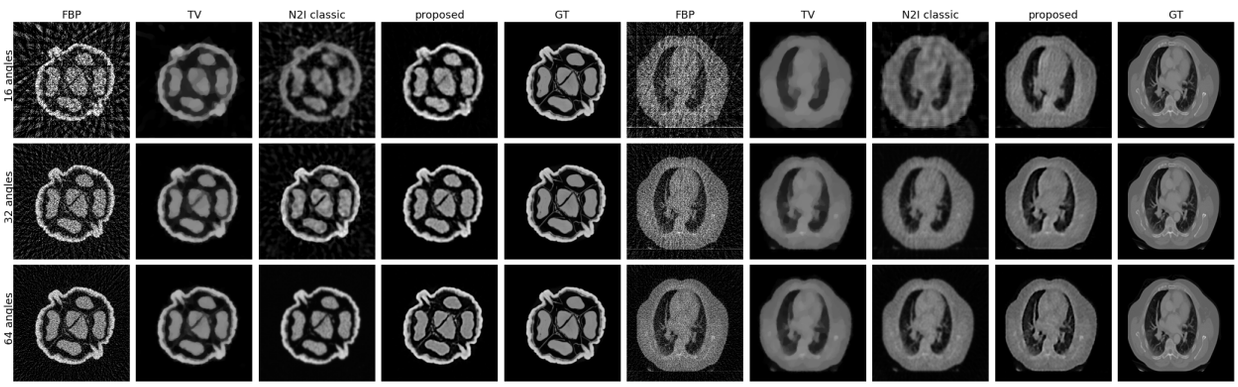


Fig. 2. **Comparison of the reconstructions obtained from the four different methods on the two different datasets for different numbers of projection angles.** We clearly see the streak-artifacts present in the classic Noise2Inverse reconstructions.

lower compared to [6], and additionally, we observe that for more views also the reconstructions obtained from [6] have fewer artifacts as the network may have problem in learning high-frequency artifacts.

V. CONCLUSION

In this work, we propose Sparse2Inverse, a self-supervised method for the reconstruction from sparse-view CT data. We present a strategy, where the neural network operates in the reconstruction domain, while the loss is computed in the projection domain, and show its superiority for severely undersampled data in comparison with the classic Noise2Inverse approach [6] and classical reconstruction algorithms like FBP and TV minimization.

REFERENCES

[1] H. Li, J. Schwab, S. Antholzer, and M. Haltmeier, "Nett: Solving inverse problems with deep neural networks," *Inverse Problems*, vol. 36, no. 6, p. 065005, 2020.

[2] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, 2019.

[3] J. Schwab, S. Antholzer, and M. Haltmeier, "Deep null space learning for inverse problems: convergence analysis and rates," *Inverse Problems*, vol. 35, no. 2, p. 025008, 2019.

[4] Y. Romano, M. Elad, and P. Milanfar, "The little engine that could: Regularization by denoising (red)," *SIAM Journal on Imaging Sciences*, vol. 10, no. 4, pp. 1804–1844, 2017.

[5] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational methods in imaging*. Springer, 2009, vol. 167.

[6] 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, pp. 1320–1335, 2020.

[7] L. Zhou, X. Wang, M. Hou, P. Li, C. Fu, Y. Ren, T. Shao, X. Hu, J. Sun, and H. Ye, "Low-dose ct reconstruction by self-supervised learning in the projection domain," *arXiv preprint arXiv:2203.06824*, 2022.

[8] Z. Zhang, X. Liang, W. Zhao, and L. Xing, "Noise2context: Context-assisted learning 3d thin-layer for low-dose ct," *Medical Physics*, vol. 48, no. 10, pp. 5794–5803, 2021.

[9] 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.

[10] J. Batson and L. Royer, "Noise2self: Blind denoising by self-supervision," in *International Conference on Machine Learning*. PMLR, 2019, pp. 524–533.

[11] D. Chen, J. Tachella, and M. E. Davies, "Equivariant imaging: Learning beyond the range space," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 4379–4388.

[12] M. O. Unal, M. Ertas, and I. Yildirim, "Self-supervised training for low-dose ct reconstruction," in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2021, pp. 69–72.

[13] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, "Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data," *Magnetic resonance in medicine*, vol. 84, no. 6, pp. 3172–3191, 2020.

[14] C. Millard and M. Chiew, "A theoretical framework for self-supervised mr image reconstruction using sub-sampling via variable density noisier2noise," *IEEE transactions on computational imaging*, 2023.

[15] ——, "Clean self-supervised mri reconstruction from noisy, sub-sampled training data with robust ssdu," *Authorea Preprints*, 2023.

[16] M. Blumenthal, G. Luo, M. Schilling, M. Haltmeier, and M. Uecker, "Nlinv-net: Self-supervised end-2-end learning for reconstructing undersampled radial cardiac real-time data," in *ISMRM annual meeting*, 2022.

[17] H. Der Sarkissian, F. Lucka, M. van Eijnatten, G. Colacicco, S. B. Coban, and K. J. Batenburg, "A cone-beam x-ray computed tomography data collection designed for machine learning," *Scientific data*, vol. 6, no. 1, p. 215, 2019.

[18] 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: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[19] A. Hendriksen, D. Schut, W. J. Palenstijn, N. Viganò, J. Kim, D. Pelt, T. van Leeuwen, and K. J. Batenburg, "Tomosipo: Fast, flexible, and convenient 3D tomography for complex scanning geometries in Python," *Optics Express*, Oct 2021. [Online]. Available: https://doi.org/10.1364/oe.439909

Development of a photon-counting micro-CBCT for a small animal proton irradiation platform

Maxin Chen, Jonathan Bortfeldt, Daniel Berthe, George Dedes, Franz Pfeiffer, Katia Parodi

***Abstract*—Photon-counting computed tomography is expected to provide improved visualization and relative stopping power estimation in image-guided proton therapy. We designed and developed a photon-counting micro-CBCT system that can perform standalone imaging experiments in a self-developed X-ray cabinet and can also be integrated in a small animal proton irradiation platform developed at our institute. A simulation micro-CBCT environment based on Geant4 has also been developed to reproduce the experimental setup but also offers the possibilities of modeling spectral and incoming photon types (i.e. primary, scatter and fluorescence) discrimination, aiming for a preliminary computational study. The characterization and commissioning of the imaging system are on-going and will be presented.**

***Index Terms*—micro-CBCT, photon-counting detector, pre-clinical image-guided proton therapy**

I. Introduction

IN image-guided proton therapy, X-ray Computed Tomography (CT) images are commonly used for treatment planning and dose calculation with conversion of CT numbers to Relative Stopping Power (RSP) through a semi-empirical calibration. The conversion error in clinical studies, however, has been reported to be up to 3% to 4% with conventional CT and below 2% with dual energy CT due to better material differentiation and soft tissue visualization capabilities [1-2]. Computed tomography imaging with photon-counting detectors can provide better visualization of tumors with virtual monochromatic images as a result of spectral photon discrimination. More importantly, it is expected to improve RSP prediction accuracy in proton therapy treatment planning with material decomposition images, which are expected to be more robust in tissue composition variations and have already shown promise with a clinical system [3].

In clinical radiation therapy, a cone beam CT (CBCT) integrated with the treatment beam delivery machine is normally used for patient positioning and verification of the patient anatomy in order to identify possible changes, e.g., tumor deformation. A good CBCT image could help with treatment plan adaptation and ensure treatment accuracy. CBCT with photon-counting detector would also be of interest for exploration of its potential in adaptive radiation therapy.

Pre-clinical research with small animals in image-guided proton therapy is used for exploring new imaging techniques and understanding complex tumor and normal tissue response, which could then contribute to optimization of treatment planning and delivery in a clinical scenario [4]. SIRMIO (Small Animal Proton Irradiator for Research in Molecular Image-guided Radiation-Oncology) is an ERC funded pre-clinical research platform aiming for precision small animal proton therapy [5]. It is a compact platform with several imaging techniques (e.g. PET, proton-imaging) integrated. To enrich its functionalities, a photon-counting micro-CBCT was designed and developed, which could perform standalone experiments as well as be integrated on the SIRMIO platform working together with other modalities.

II. METHODS

A. The photon-counting micro-CBCT system

The developed micro-CBCT system includes a compact X-ray tube (RTW MCBI-65B, max 65kVp, 50μm focal spot), a small photon-counting detector (Varex Imaging XC-ACTAEON, 26x26mm^2, 0.1mm pixel pitch), and six remote-controlled stages for phantom and detector movement. Holders of X-ray tube and detector were designed to be compatible with the SIRMIO platform and precise to be aligned with the ISO-center of the platform.

The X-ray tube is controlled via vendor provided GUI and the tube temperature is monitored by custom electronics reading of a temperature sensor (PT100). The detector can be controlled either via vendor provided GUI or self-developed C++ code for multiple projections acquisition in sync with motor movements. All stages are controlled by phyMOTION (Phytron GmbH, Maisach, Germany), via a self-developed software already used on SIRMIO, based on an open source device-controlled software platform TANGO [6].

B. The X-ray shielding cabinet

To perform standalone experiments in the lab, we also designed and set up a shielded cabinet. The X-ray cabinet exhibits a compact dimension to ensure a sufficient source to detector distance and a detector lateral shift distance to enlarge the reconstructed field of view for mouse imaging with displaced detector reconstruction. It is constructed with "sandwich" structure plates, i.e. aluminum-lead-aluminum, (MSR Röntgenraumtechnische Systeme GmbH, Jünkerath, Germany), aluminum profiles (Item Industrietechnik GmbH, Solingen, Germany), and safety relay circuit (K.A. Schmersal

The abstract was firstly submitted on Jan. 23rd, 2024, to 8th International Conference on Image Formation in X-Ray Computed Tomography. This work is supported by One Munich - EQAP project. The first year of the work was supported by the German Research Foundation (DFG) within the Research Training Group GRK 2274.

Maxin Chen, Jonathan Bortfeldt, George Dedes, and Katia Parodi are with the Department of Medical Physics, Ludwig-Maximilians-Universität München (LMU Munich), Garching bei München, Germany (e-mail: Maxin.Chen@lmu.de).

Daniel Berthe and Franz Pfeiffer are with the Department for Biomedical Physics, Technische Universität München (TUM), Garching bei München, Germany.

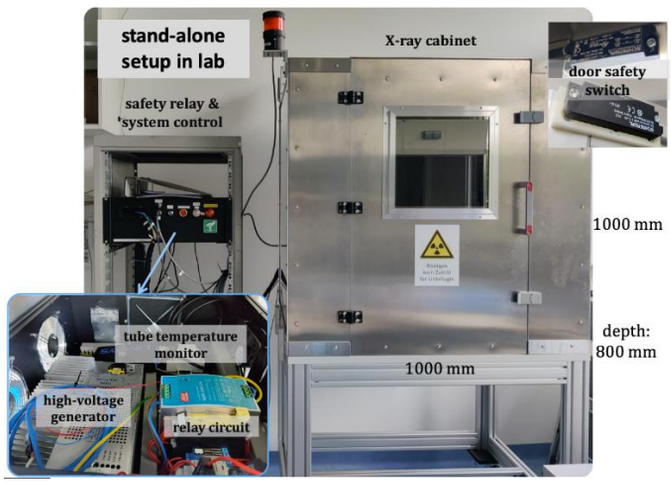


Fig 1. The stand-alone X-ray cabinet setup and safety circuits

GmbH & Co. KG, Wuppertal, Germany). The shielding was designed to be sufficient for operation also with an X-ray tube operating at 80kVp.

C. *Imaging study*

The basic system characteristics are being addressed in partly still ongoing tests, including the X-ray tube spectrum with and without beam filter, detector noise and energy response, along with system geometrical accuracy and stability.

We also plan to perform phantom imaging to study detector resolution, energy bin selection and material decomposition, using both commercial phantoms such as the SMART phantom (SmART Scientific Solutions B.V., Maastricht, Netherlands) and a bar-pattern phantom (QRM GmbH, Möhrendorf, Germany), along with self-designed and in-house manufactured step-wedge phantoms.

D. *Simulation*

A Geant4 (v11.0.2) based simulation platform of the spectral micro-CBCT system was also designed and developed. The physical parameters of the experimental system, i.e., geometry, X-ray beam angle and spectrum, detector dimension and resolution, were reproduced in the "simulated system". The simulation object can be either set directly in the code based on constructive solid geometry or loaded from DICOM images. Primary, scatter and fluorescence photons can be recorded on the detector plane separately with pre-defined energy bins. The simulation platform is used for preliminary, still ongoing reconstruction and correction algorithm studies.

III. Results

A. *The photon-counting micro-CBCT system and X-ray cabinet*

The cabinet and system design and realization have been recently completed. Figure 1 and Figure 2 show the cabinet and photon-counting micro-CBCT system setup in the lab.

B. *Imaging study*

Due to a delay caused by the need of energy re-calibration of the detector at the manufacturer, the characterization of the system is currently ongoing. Design details, capabilities and first imaging results will be presented.

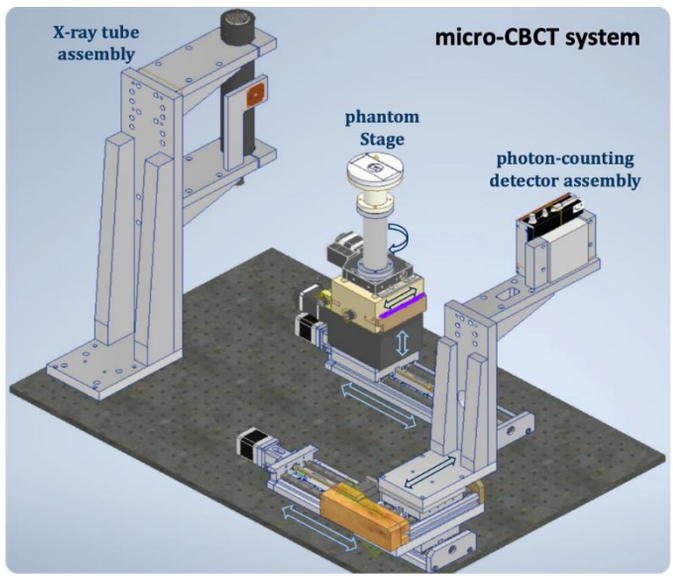


Fig. 2. The 3D sketch of the photon-counting micro-CBCT system in the cabinet.

C. *Simulation*

Figure 3 shows an example of the simulation result of the imaging of a SMART phantom. The input spectrum shown here is produced by SPEKTR [7] with 65kVp, Tungsten target, and 0.1mm Be filtration, the same as the parameters of the X-ray tube in the lab setup. A total of 10^8 of photons was used in the simulation for each angle. The energy distribution of the primary photons after the phantom shows the beam hardening effect. The energy distribution of scattered photons at each energy bin and its impacts on energy bin selection and image quality will be the next step to be studied and will also be presented.

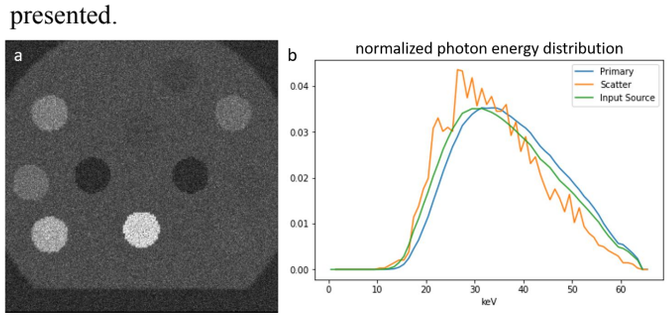


Fig. 3. Example of a SMART phantom simulation.
a). is the projection at 0 angle of energy bin (16,32] keV;
b). is the photon energy distribution comparison between input source, primary and scatter photons at detector plane

IV. Conclusion

We designed and developed a photon-counting micro-CBCT setup for use in a custom X-ray cabinet and a pre-clinical image guided proton therapy platform.

ACKNOWLEDGMENT

We would like to thank Romy Knab, Leonard Doyle, Vignesh Vaidyanathan, the mechanical and the electronics workshop from LMU, Dr. Hans Wirth, Christoph Rahm, Florian Saran and Dr. Johannes Wulz from LMU-CALA for the kind support in experimental platform setup. We would also like to thank Jonathan Schock, Fredrik Nyman and York Haemisch from Varex Imaging for the kind support in detector energy calibration.

REFERENCES

[1] Paganetti, Harald et al. “Roadmap: proton therapy physics and biology.” Physics in medicine and biology vol. 66,5 10.1088/1361-6560/abcd16. 26 Feb. 2021, doi:10.1088/1361-6560/abcd16

[2] Hudobivnik, Nace et al. “Comparison of proton therapy treatment planning for head tumors with a pencil beam algorithm on dual and single energy CT images.” Medical physics vol. 43,1 (2016): 495. doi:10.1118/1.4939106

[3] Hu, Guyue et al. “Assessment of quantitative information for radiation therapy at a first-generation clinical photon-counting computed tomography scanner.” Frontiers in oncology vol. 12 970299. 14 Sep. 2022, doi:10.3389/fonc.2022.970299

[4] Clark, D P, and C T Badea. “Advances in micro-CT imaging of small animals.” Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB) vol. 88 (2021): 175-192. doi:10.1016/j.ejmp.2021.07.005

[5] Parodi, Katia et al. “Towards a novel small animal proton irradiation platform: the SIRMIO project.” Acta oncologica (Stockholm, Sweden) vol. 58,10 (2019): 1470-1475. doi:10.1080/0284186X.2019.1630752

[6] Chaize, J-M.et al. TANGO Controls source code (Version 9.3.4) [Source code]. https://www.tango-controls.org/

[7] Punnoose, J et al. “Technical Note: spektr 3.0-A computational tool for x-ray spectrum modeling and analysis.” Medical physics vol. 43,8 (2016): 4711. doi:10.1118/1.4955438

Continuous Domain Reconstruction in CT Imaging with Coordinate-based Neural Networks

Gregory Ongie

***Abstract*—The majority of iterative algorithms for CT reconstruction rely on discrete-to-discrete modeling, where both the sinogram measurements and image to be estimated are discrete arrays. However, tomographic projections are ideally modeled as line integrals of a continuous attenuation function, i.e., the true inverse problem is discrete-to-continuous in nature. Recently, coordinate-based neural networks (CBNNs), also known as implicit neural representations, have gained traction as a flexible type of continuous domain image representation in a variety of inverse problems arising in computer vision and computational imaging. Using standard neural network training techniques, a CBNN can be fit to measurements to give a continuous domain estimate of the image. In this study, we empirically investigate the potential of CBNNs to solve the continuous domain inverse problems in CT imaging. In particular, we experiment with reconstructing an analytical phantom from its ideal sparse-view sinogram measurements. Our results illustrate that reconstruction with a CBNN are more accurate than filtered back projection and algebraic reconstruction techniques at a variety of resolutions, and competitive with total variation regularized iterative reconstruction.**

***Index Terms*—Coordinate based neural networks, implicit neural representations, CT reconstruction, continuous domain inverse problems.**

I. Introduction

Discrete image representations, which encode images as fixed-resolution arrays of pixels or voxels, form the backbone of modern computational imaging. However, estimating discrete images with high spatial resolution from indirect measurements is computationally demanding and often ill-posed because the number of unknown parameters (i.e., pixels or voxels) far exceeds the available measurements. This often necessitates introducing complicated regularization functionals to stabilize the recovery, which then require custom iterative solvers to implement.

Recently, a non-linear method of representing images as neural networks has gained traction in solving a variety of inverse problems in computer vision and computational imaging [1], [2]. These so-called *coodinate-based neural networks* (CBNNs)—also known as *implicit neural representations*—parameterize an image as a small-scale multi-layer perceptron taking spatial coordinates as inputs. Compared to traditional pixel/voxel representations, CBNNs are typically far more parameter-efficient, and scale well to multi-dimensional imaging problems. Furthermore, they can be fit to measurements using standard neural network algorithms, bypassing the need for custom iterative solvers.

G. Ongie is with the Department of Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI, USA e-mail: gregory.ongie@marquette.edu

In the context of computed tomography (CT) reconstruction, one advantage of CBNNs is that they allow for more accurate modelling of the ideal continuous-to-discrete measurement process. In particular, since the CBNN is a continuous function, samples of its continuous domain Radon transform can be approximated to an arbitrary degree of accuracy.

Several studies have proposed specialized CBNN approaches for CT reconstruction, e.g., [3], [4]. Yet, to the best of our knowledge, no work has investigated whether CBNNs enable the solution of the ideal continuous domain CT reconstruction problem. In this study, we explore this question empirically. For simplicity, we focus on a 2D imaging scenario and analytical phantoms where computing exact samples of Radon transform is possible. We investigate the recovery of sparse-view sinogram data with CBNNs using a close-to-exact continuous domain forward model versus recovery in a pixel basis with model-based iterative methods using a discretized forward model.

II. Problem Formulation

Recovery of a continuous-domain image $f : \mathbb{R}^2 \to \mathbb{R}$ from discrete sinogram measurements $\boldsymbol{y} \in \mathbb{R}^M$ can be posed as an optimization over a function space:

$$\min_{f \in \mathcal{F}} \|\mathcal{A}f - \boldsymbol{y}\|^2 \tag{1}$$

where $\mathcal{F}$ is some class of functions, and the continuous-to-discrete linear operator $\mathcal{A} : \mathcal{F} \to \mathbb{R}^m$ returns a vector of samples of the continuous Radon transform of f, i.e., if $\boldsymbol{y} = \mathcal{A}f$ then its ith entry is $y_i = \mathcal{R}f(\theta_i, t_i)$ for some angle/offset pair $(\theta_i, t_i) \in [0, 2\pi) \times \mathbb{R}$ where

$$\mathcal{R}f(\theta, t) = \int_{-\infty}^{\infty} f(s\sin(\theta)+t\cos(\theta), -s\sin(\theta)+t\cos(\theta))ds.$$

Here, for simplicity, we focus on the least-squares loss in (1), though other loss functions may be more appropriate in settings with a statistical noise model, e.g., the negative log likelihood.

A. Reconstruction in a pixel basis

When reconstructing in a pixel basis, the function class $\mathcal{F}$ in (1) is chosen to be the space of images $f(x, y) = \sum_{i,j} f_{i,j}\chi_{P_{i,j}}(x, y)$, where $\chi_{P_{i,j}}$ is the indicator function for a square region $P_{i,j}$ representing the i, jth pixel. In this case, (1) reduces to the least squares problem

$$\min_{\boldsymbol{f} \in \mathbb{R}^N} \|\boldsymbol{A}\boldsymbol{f} - \boldsymbol{y}\|_2^2 \tag{2}$$

where $\boldsymbol{f} \in \mathbb{R}^N$ is the vector of all pixel intensities $f_{i,j}$, and the columns of $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ are given by $\mathcal{A}\chi_{P_{i,j}}$ for all i, j.

One drawback to this approach is that when discretizing at a high resolution the number of pixels N can easily exceed the number of sinogram measurements M. In this case, since $\boldsymbol{A}$ is a $M \times N$ matrix, there are infinitely many solutions to (2), i.e., reconstruction is ill-posed. To overcome this issue, a common strategy is to solve the regularized least squares problem instead:

$$\min_{\boldsymbol{f} \in \mathbb{R}^N} \|\boldsymbol{A}\boldsymbol{f} - \boldsymbol{y}\|^2 + \lambda R(\boldsymbol{f}), \tag{3}$$

where $R(\boldsymbol{f})$ is a regularizer designed to promote desirable image properties and $\lambda > 0$ is a tunable regularization strength parameter. One common choice is total variation regularization

$$R_{TV}(\boldsymbol{f}) = \sum_{i,j} \sqrt{|f_{i+1,j} - f_{i,j}|^2 + |f_{i,j+1} - f_{i,j}|^2},$$

which promotes reconstructing images that are piecewise constant.

B. Reconstruction with CBNNs

The philosophy behind the CBNN approach is to choose $\mathcal{F}$ in (1) to be a parametric class of functions $f_{\boldsymbol{\theta}} : \mathbb{R}^2 \to \mathbb{R}$ described by a fixed neural network architecture with parameters $\boldsymbol{\theta} \in \mathbb{R}^p$. Image recovery in this class is then posed as an optimization problem over the network parameters:

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^P} \|\mathcal{A}f_{\boldsymbol{\theta}} - \boldsymbol{y}\|_2^2. \tag{4}$$

Typically, the number of trainable parameters P can be chosen to be far less than the number of sinogram measurements M. In this case, problem (4) is not necessarily ill-posed, in the sense that it may have a unique global minimizer.

However, finding a global minimizer of (4) is challenging in general. This is because, in contrast to the pixel-basis approach, the mapping from parameters $\boldsymbol{\theta}$ to an image $f_{\boldsymbol{\theta}}$ is no longer linear, which means that (4) is a non-linear least squares problem whose objective function is typically non-convex. Nevertheless, standard neural network training methods, such as stochastic gradient descent, can be used to find a local minimizer that fits the data well [5]. Following this fitting procedure, a rasterized image can be generating by evaluating the CBNN $f_{\boldsymbol{\theta}}$ on a pixel grid of any desired resolution.

III. Simulation Details

Our goal is to compare the performance of pixel-basis formulations (2), (3) based on a discrete-to-discrete model, and the CBNN formulation (4) based on a discrete-to-continuous model, for the recovery of ideal CT data. In particular, we focus on recovery of an analytical phantom for which exact Radon transform samples can be computed. The phantom we use, shown in Figure 1, is a modified version of the Shepp-Logan phantom that includes a detail insert consisting of an 8×8 grid of discs of varying radii and contrasts. In addition, we include two high contrast ovals near the center of the phantom. These ovals induce streaking artifacts that intersect with the detail insert when reconstructing with filtered back projection, which generally makes the faithful recovery of the detail insert features more challenging.

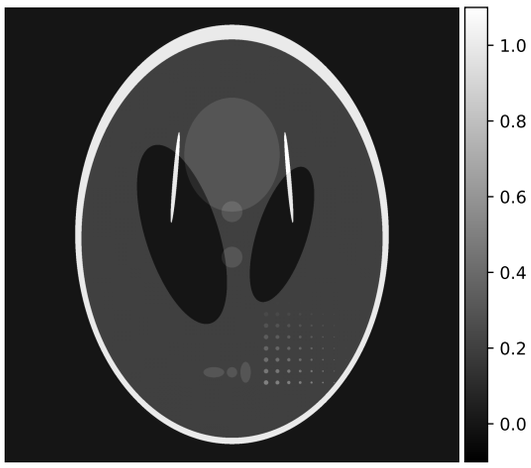


Fig. 1. Modified Shepp-Logan phantom used in our experiments.

For sinogram measurements, we simulate a circular fan-beam acquisition of the phantom. We assume the object lies in a circular region-of-interest (ROI) with radius 9 cm, and the x-ray source follows a circular trajectory with a radius of 50 cm centered in the middle of the ROI. The source-to-detector-center distance is 100 cm, and the detector is modeled as a linear array. We simulate 128 equally distributed views over a 360 degree scan and the projections are sampled on a 1024-pixel detector, resulting in a total of $M = 128 \times 1024 = 131,072$ sinogram measurements. Analytic line integrals are calculated for the phantom using exact formulas for the Radon transform of the indicator function of an ellipse.

As baselines, we compare against analytic reconstruction using filtered back projection (FBP) and two model-based iterative methods: least squares (LS), and regularized least squares with total variation regularization (TV-RLS). FBP is implemented using a weighted pixel-driven back-projection and a ramp filter [6]. For LS and TV-RLS, we use a standard line-intersection model for the x-ray transform of a discrete image to define the discrete ray transform matrix $\boldsymbol{A}$. The images are pixelated on a square $n \times n$ pixel grid, where we vary $n \in \{128, 256, 512\}$. For LS we find an approximate solution of (2) by applying the conjugate gradients algorithm to the normal equations with zero initialization and early stopping at 100 iterations, which approximates a pseudoinverse solution. For PLS-TV, we solve (3) using the Chambolle-Pock primal-dual algorithm [7]. To set the regularization strengths λ, we performed a grid search over $\lambda \in \{1, 0.1, 0.01, 0.001\}$ and chose the value that gave the lowest root mean-squared error in an ROI containing the detail insert.

For the CBNN approach, we approximate the continuous domain Radon transform using a Riemann approximation by densely sampling each ray uniformly along a trajectory through the circular ROI with 100 evaluation points per linear cm, such that the longest ray is approximated with 1800 evaluations. As our CBNN architecture, we use a multi-layer perceptron (MLP) with a Fourier features layers as proposed

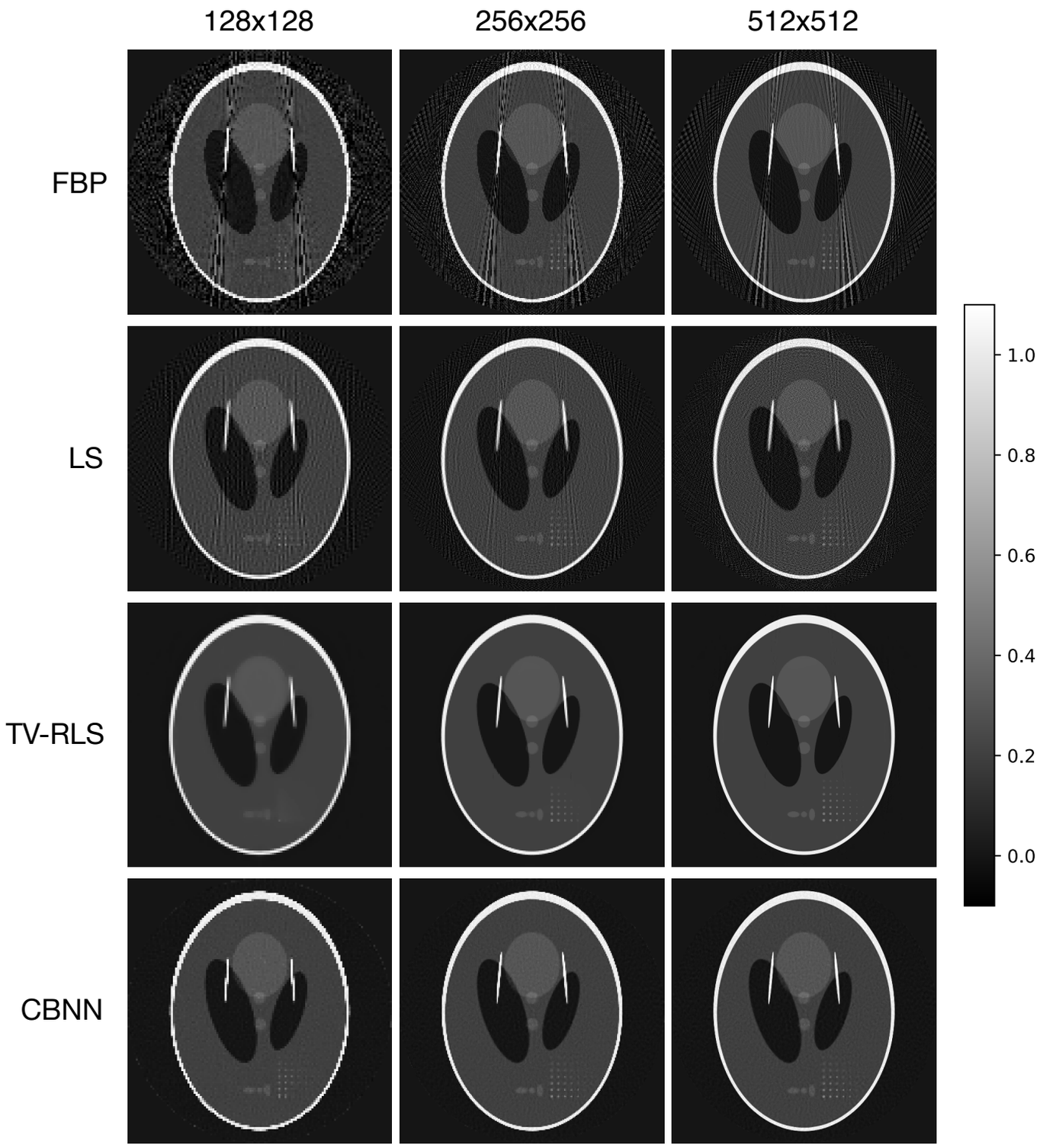


Fig. 2. Reconstruction of analytical phantom from a simulated sparse-view acquisition at different pixel resolutions. Sinogram data is simulated under a circular fan-beam geometry with 128 equi-spaced views and 1024 detector bins using an exact line integral model. Key: **FBP=**Filtered Back Projection, **LS=**Least Squares, **TV-RLS=**Total Variation Regularized Least Squares, **CBNN=**Coordinate-based neural network. Note that the CBNN reconstructions are rasterizations of the same continuous domain function onto different grid sizes.

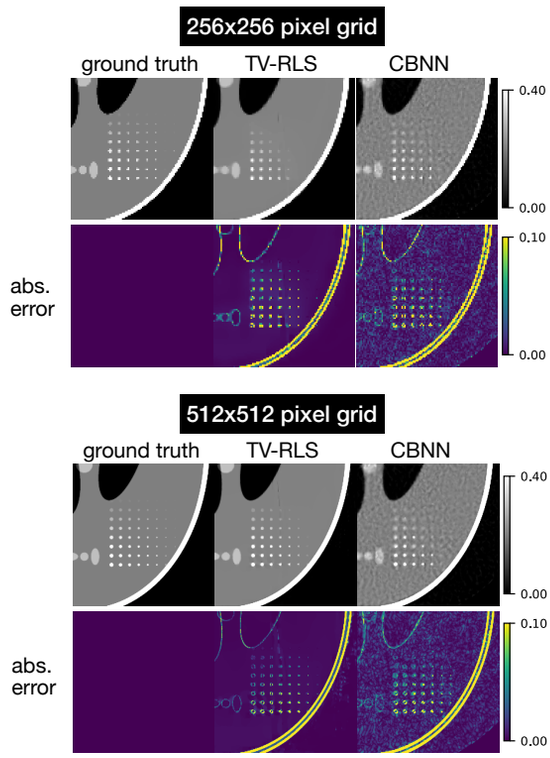


Fig. 3. Blow-up an ROI containing the contrast detail insert.

in [5]. We use an MLP with three hidden-layers with a ReLU activation and width 50 in each layer, resulting in 30801 trainable parameters. For the random Fourier features layer, we sample 256 random frequencies from a standard Gaussian with bandwidth parameter $\sigma = 1.0$ cm^{-1}. To train the CBNN, we use a stochastic gradient descent approach where 8 random views are sampled at each iteration and used to approximate the least squares loss (4). The network is trained for a total of 30000 SGD iterations using the Adam optimizer with an initial learning rate of 0.001, which is decreased by a power of 10 every 10000 iterations. Finally, to compare with the discrete reconstruction methods, we rasterize the CBNN on a variety of pixel grid sizes.

IV. Results

Figure 2 shows the reconstructions obtained by all methods at three pixel resolutions. Additionally, Figure 3 shows a blow-up of an ROI containing the detail insert comparing TV-RLS and CBNN reconstructions at 256×256 and 512×512 pixel resolutions. First, we observe that the FBP and LS reconstructions contain significant streaking artifacts at all pixel resolutions, as expected from the sparse-view sampling. Additionally, visible aliasing-like artifacts occur in the FBP and LS reconstructions at the 128×128 resolution. Relative to FBP and LS, both TV-RLS and CBNN reconstructions show essentially no visible streaking or aliasing artifacts. At the two lowest resolutions, there is some loss of fine detail in the TV-RLS reconstructions. In particular, at a 256×256 resolution, some detail insert features are wiped out in the TV-RLS reconstruction due to over-regularization (see Figure 3). However, compared to TV-RLS at a 512×512 resolution, the CBNN reconstruction has less sharp boundaries and preserves fewer of the discs in the detail insert. Also, the CBNN reconstruction has a uniform noise-like texture that is absent from the TV-RLS reconstruction.

V. Discussion and Conclusion

We investigate the possibility of recovering a continuous domain solution to a sparse-view CT inverse problem using CBNNs and an analytical phantom. Our experiments illustrate that reconstruction with a CBNN can be more accurate than filtered back projection and standard iterative reconstruction techniques, especially when performed on a low resolution pixel grid. However, we observe the CBNN recovery is less accurate than a total variation regularized reconstruction on a sufficiently high resolution grid. This may be in part because the phantom under consideration is a piecewise constant image, which is well-aligned with the total variation prior. An open question is whether CBNNs are better suited to alternative object models, e.g., piecewise smooth images.

Additionally, in our experiments we focused on only one type of CBNN architecture based on [5]. Several alternative architectures have been proposed (e.g., [8], [9]), which primarily differ in their activation functions. Exploring the full impact of changes to the CBNN width, depth, activation function, and training protocol on image estimation is an interesting direction for further study.

Acknowledgements

This work was supported by NSF CRII award CCF-2153371.

References

[1] V. Sitzmann, M. Zollhoefer, and G. Wetzstein, "Scene representation networks: Continuous 3d-structure-aware neural scene representations," 2019, pp. 1121–1132.

[2] 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," in *European Conference on Computer Vision (ECCV)*. Springer, 2020, pp. 405–421.

[3] Y. Sun, J. Liu, M. Xie, B. Wohlberg, and U. S. Kamilov, "CoIL: Coordinate-based internal learning for tomographic imaging," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 1400–1412, 2021.

[4] B. Song, L. Shen, and L. Xing, "PINER: Prior-informed implicit neural representation learning for test-time adaptation in sparse-view ct reconstruction," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 1928–1938.

[5] 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," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

[6] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*. SIAM, 2001.

[7] 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*, vol. 57, no. 10, p. 3065, 2012.

[8] V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," *Advances in neural information processing systems*, vol. 33, pp. 7462–7473, 2020.

[9] V. Saragadam, D. LeJeune, J. Tan, G. Balakrishnan, A. Veeraraghavan, and R. G. Baraniuk, "WIRE: Wavelet implicit neural representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 507–18 516.

Evaluation of Photon-Counting Detector CT on 3D-Printed Lung Phantom with Digital Lesions

Kai Mei, Shobhit Sharma, Leonid Roshkovan, Steven Ross, Thomas Labno, Zhou Yu, Richard Thompson, and Peter B. Noël

***Abstract*— The recent advancement of photon-counting detector computed tomography (PCD-CT) has notably improved spatial resolution in CT imaging. CT phantoms, mimicking real-life image textures and densities, are vital for evaluating and confirming CT efficacy in clinical settings. This study involved designing a phantom featuring synthetic lung lesions with three levels of spiculation (low, moderate, and high), and each in four sizes (diameters from approximately 6 cm to 1 cm). Utilizing PixelPrint software, the phantom was 3D-printed and then scanned with a CdZnTe-based PCD-CT prototype. Scans were conducted at varying dose levels (1.7 - 20 mGy), using a typical lung imaging protocol. The resulting CT images were processed and aligned with the ground truth image of the synthetic lesion. We calculated discrepancies in the volume and shape of lesion segmentations. When exposure levels were reduced, root mean squared errors of the CT image to the ground truth image increased for all lesion sizes, and the type of the lesions also became more challenging to characterized. Our findings indicate that PCD-CT enable visualization and identification of lung lesion types down to 15 mm with submillimeter features, even under low dose conditions (1.7 mGy). In summary, our research validates the potential of PCD-CT in lung imaging through characterization of the aggressive features in lung lesions of various sizes with significantly reduced radiation dose. This will be of great clinical value for better and more precise diagnosis and treatment planning for patients.**

***Index Terms*— Computed tomography, 3D-printing, phantom, pulmonary imaging, lung lesion, resolution, photon-counting.**

I. Introduction

THE recent development of photon-counting detector CT (PCD-CT) offers a significantly increased spatial resolution for CT imaging and comparably higher image quality under low dose circumstances. These new CT systems are already available in clinical settings and have advantages for a wide range of clinical indications [1]. Lung nodule detection and characterization is a common and important task performed primarily with the help of chest CT and is a key initial step in management of both benign and malignant thoracic pathology including primary lung malignancy. Key factors in distinguishing benign from malignant nodules include their growth rate, shape, and surface characteristics. In this area, realistic phantoms prove essential for researchers to assess the effectiveness of PCD-CT systems for evaluating lung cancer.

Phantoms play an essential rule in CT research and clinical practice. Both academic and clinical CT communities benefit from using phantoms to evaluate the performance of CT and facilitate the development of CT technology. A rapid and cost-effective manufacturing method to produce customizable and lifelike phantoms can target specific tasks and provide solutions to numerous research challenges.

In order to access the spatial resolution performance of a PCD-CT prototype, we utilized the PixelPrint [2] - [4] technique to manufacture a realistic lung lesion phantom. The phantom contains three different types of realistic synthetic solid lung lesion models [5] with different levels of spiculation. Each type of the lesions was printed in four sizes. A CdZnTe-based prototype photon-counting CT system (Canon Medical Systems Corporation) was used to scan the phantom at 120 kVp with varying dose levels (max: 20 mGy, min: 1.7 mGy). Comparisons between the CT images across various dose levels and the ground truth images of the lesions were conducted and benchmarked. Lesion segmentations were performed using thresholding techniques, and the volumetric size and visualization of the lesion shapes were subsequently extracted and analyzed. The findings indicate that PCD-CT can effectively visualize solid lung lesions down to 15 mm with submillimeter features.

II. Materials and Methods

A. Phantom Design

Three distinct lung lesion shapes [5] were included in this study, depicted in Fig. 1. Lesion A (spiculation: high) is notably spiculated, characterized by a rough surface with mass extending radially from its margin in all directions. This spiculated nature contributes to its larger radius. Lesion B (spiculation: medium) presents a less spiky margin, being comparatively rounder and possessing a more substantial central mass. Lesion C (spiculation: low), is an oval-shaped lesion, features a predominantly smooth surface.

These three lesions were scaled in four different sizes in the phantom, 4x, 3x, 2x and 1x. The pixel sizes in each scale were 0.8, 0.6, 0.4 and 0.2 mm, respectively. The lesion images were

This work was supported by Canon Medical Research USA.

K. M., L.R. and P. B. N. are with Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. (kai.mei@pennmedicine.upenn.edu)

S. S., S. R., T. L. Z. Y., and R. T. are with Canon Medical Research USA, Vernon Hills, IL, USA.

.

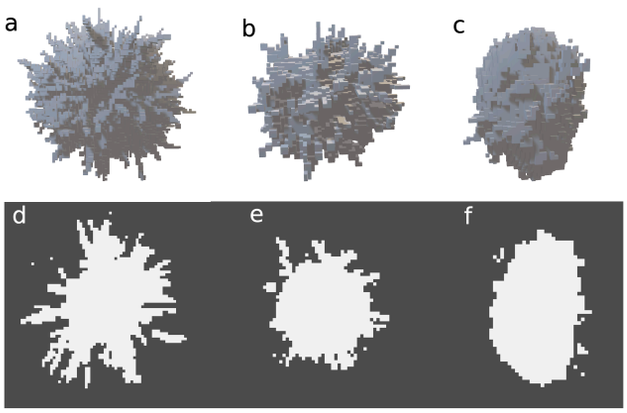


Fig. 1. Three different shape of digital lung lesions investigated in this study. Lesion A has the spikiest surface (a/d), Lesion B is medium spiculated (b/e) and Lesion C is an oval shape with smooth surface (c/f).

represented as binary and the method used in rescaling is first order (nearest-neighbor interpolation). At the largest scale (4x) of 0.8 mm resolution, Lesion A has a diameter of about 45 mm, Lesion B has diameter of about 30 mm, and Lesion C is about 30 mm along the major axis and 20 mm along the minor axis. At the smallest scale (1x) of 0.2 mm resolution, all lesions are $(1/4)^3$ of the largest size. At this smallest scale, Lesion A is about 11 mm in diameter, Lesion B is 7 mm in diameter and Lesion C is oval about 7 mm by 5 mm. Fig. 2 illustrates the central slice of the phantom, containing Lesion A, B and C in 4x, 3x and 1x scales. 2x is not shown in this slice position.

The whole phantom is a cylinder measuring 150 mm in diameter and 67 mm in height, tailored to fit the bore space of the Multi-Energy CT Phantom (Gammex, Sun Nuclear). The

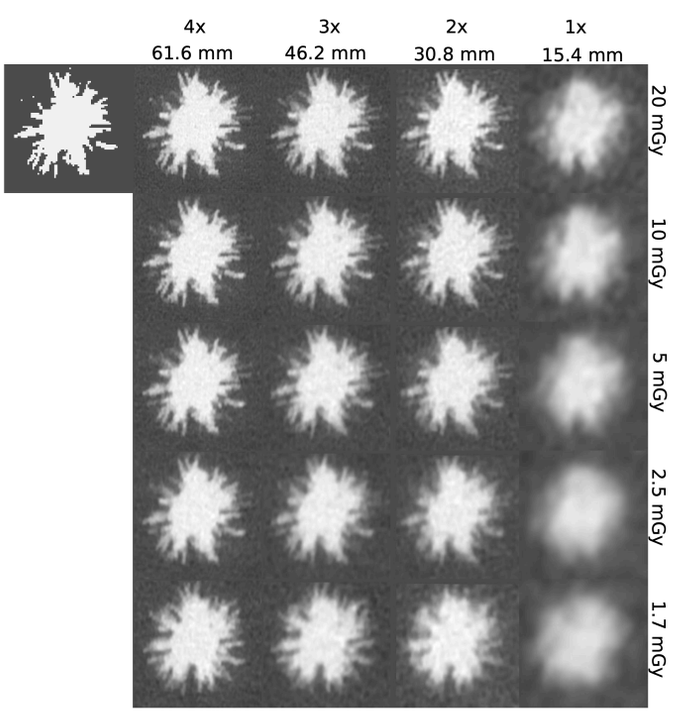


Fig. 3. CT images of Lesion A in different scales under different dose levels. 0.8, 0.6, 0.4 and 0.2 mm resolutions are noted as 4x, 3x, 2x and 1x. The length on the top represents the length of the edge of the images.

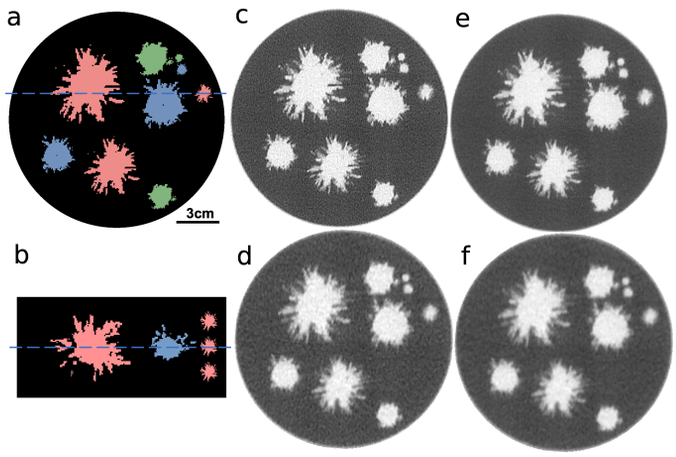


Fig. 2. Simulated lung lesion with different scaling in the 3D-printed phantom. (a/b) depicts the design of the phantom with axial and coronal views. Each of the three lesion types are illustrated in red, blue and green, respectively. 4x, 3x, and 1x resolution are shown. (c) CT image of the phantom with sharp filter and highest dose (20 mGy), and (d) lowest dose (1.7 mGy) (e) CT image of the phantom with standard filter and highest dose (20 mGy), and (f) lowest dose (1.7 mGy).

lesion exhibits a uniform Hounsfield unit of 0 HU, while the background is set at -700 HU. No extra noise has been introduced to the phantom, as shown in Fig. 2a/b.

B. PixelPrint

For printing, we used a fused-filament 3D printer (Lulzbot TAZ 6 with M175 tool head, Fargo Additive Manufacturing Equipment 3D, LLC Fargo, ND, USA) with a 0.25 mm brass nozzle. The filament was polylactic acid (PLA) with a diameter of 1.75 mm (MakeShaper, Keene Village Plastics, Cleveland, OH, USA), which was extruded at the nozzle temperature of 190 °C. Build plate was heated to 50 °C to improve adhesion.

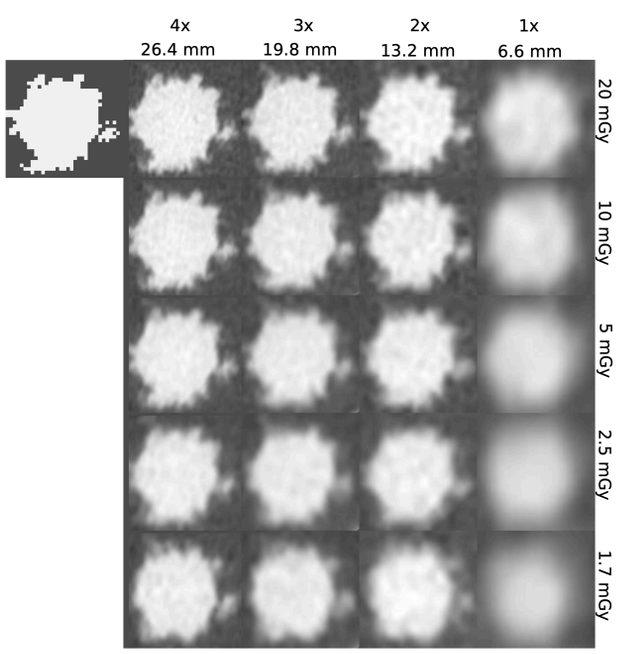


Fig. 4. CT images of Lesion C in different scales under different dose levels. 0.8, 0.6, 0.4 and 0.2 mm resolutions are noted as 4x, 3x, 2x and 1x. The length on the top represents the length of the edge of the images.

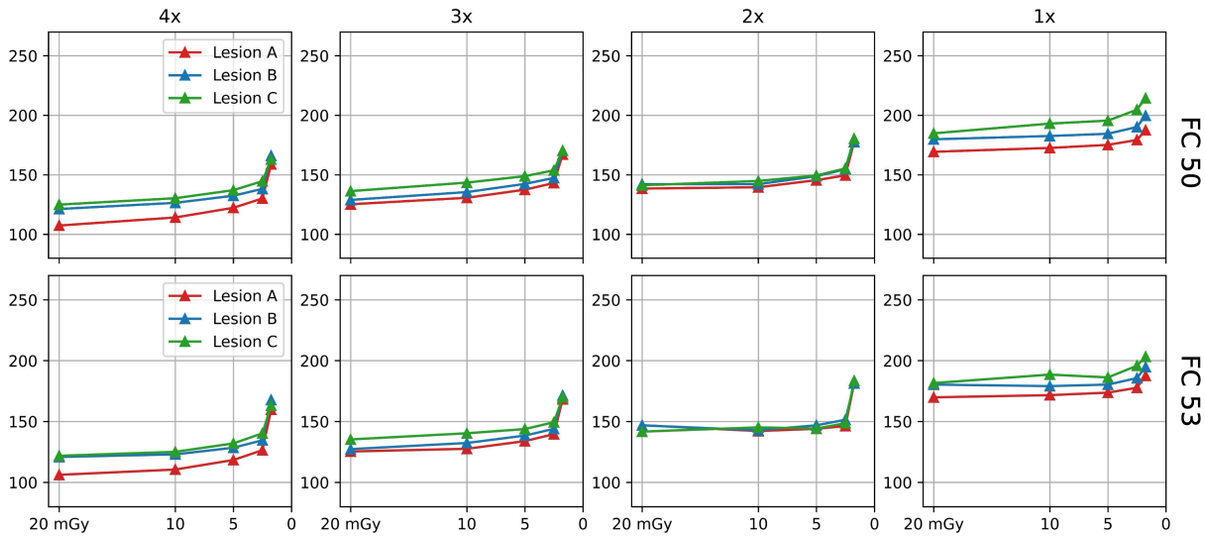


Fig. 5. Root mean squared errors [HU] of the lesion images compared to the ground truth in different scales (from left to right 4x, 3x, 2x and 1x) and under different dose (y-axis, sampled at 20, 10, 5, 2.5, and 1.7 mGy). The first row is standard lung reconstruction filter (FC50) and the second row is the sharper lung reconstruction filter (FC53).

For this specific study, printing speed varied from 5.5 to 20 mm/s with a constant volumetric rate of 18 mm^3/min, producing line widths from 0.15 to 0.54 mm. The spacing of these filament lines is fixed at 0.6 mm, making infill ratios ranging from 25.6% to 90.2%. Lines from neighboring layers are altered in directions, either horizontal or vertical. The acceleration was 500 mm/s^2 and the jerk setting (or acceleration threshold) was 8 mm/s. The height of each printed layer is 0.1 mm.

C. PCD-CT scan

The printed phantom was placed in an extension ring to increase the diameter from 15 cm to 20 cm. The extension ring was manufactured via 3D printing utilizing PLA with a targeted density close to water. Both phantoms were placed in the 20 cm bore of the Multi-Energy CT Phantom for experimental scanning.

CT scans were performed on a PCD-CT prototype (Canon Medical Research USA) equipped with a CdZnTe-based photon counting detector. The scans were performed with tube voltage of 120 kVp and exposure levels of 200, 100, 50, 25 and 17.5 mAs. These exposure levels translated to approximately 20, 10, 5, 2.5 and 1.7 mGy in volumetric CT dose index (CTDIv). Images were reconstructed using Adaptive Iterative Dose Reduction (AIDR) algorithm with a standard (FC50) and a sharper (FC53) lung reconstruction kernel at 0.5 mm pixel spacing. Slice thickness was 0.6 mm. Standard strength of AIDR was used to minimize the noise and maximize image quality.

D. Image analysis

The images were exported, and the entire CT volume was aligned 3D-wise to the ground truth phantom image containing digital lung lesions. Only images from the highest dose scan underwent this rigid registration. We used the SimpleITK Python library. Key settings included a 50-bin histogram and a 0.5% metric sampling. The process utilized gradient descent optimization, with a learning rate of 0.1 and 500 iterations. This resulted in an affine transformation, and the result images were remeshed by SimpleITK.

Subsequently, three lesions at all four different sizes were extracted. Each was further realigned to their respective original images, maintaining original shapes and scales, using the same parameters as before. These adjusted images were then applied to calculate root mean squared errors (RMSE) for further analysis.

Lesion shapes were segmented from the background through basic thresholding. The threshold was -380 HU and

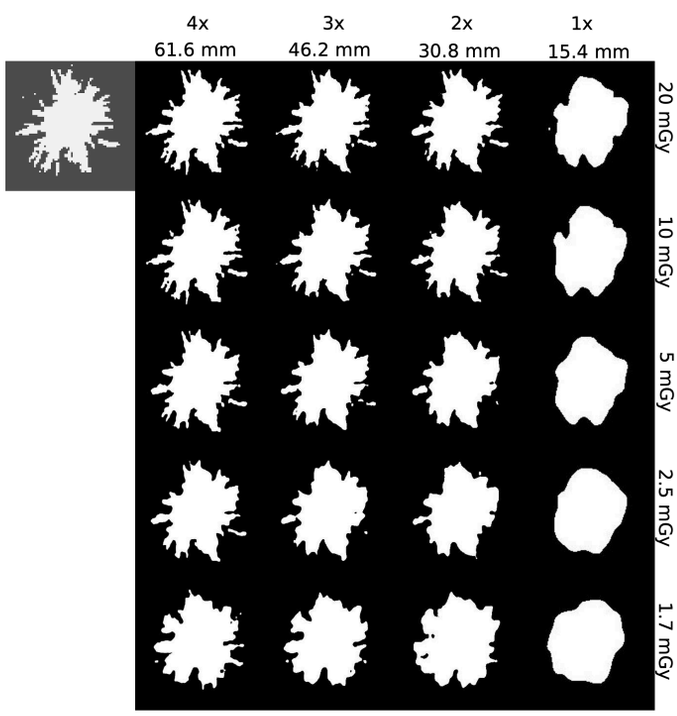


Fig. 6. Segmentation result of lesion A in different scales under different dose levels. 0.8, 0.6, 0.4 and 0.2 mm resolutions are noted as 4x, 3x, 2x and 1x. The length on the top represents the length of the edge of the images.

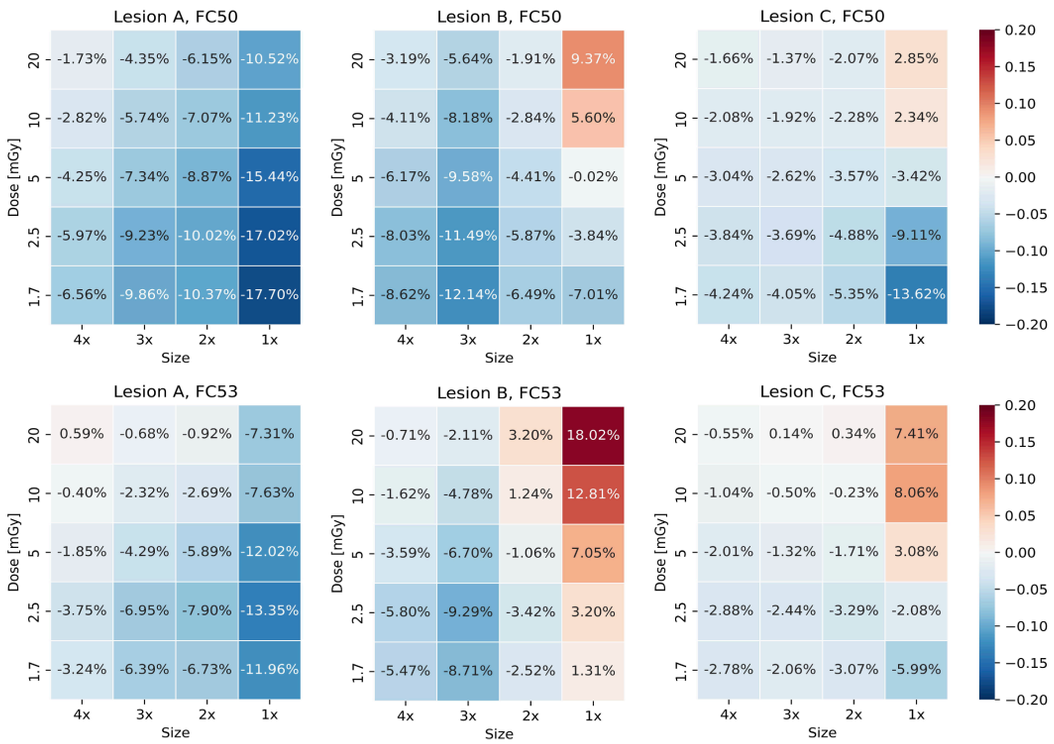


Fig. 7. Changes of volumetric size of the lesion segmentation, compared to the ground truth. Left, middle and right plot indicates Lesion A, B and C. Dose level (exposure) is noted in y-axis with unit mGy. Size of the lesion is noted in x-axis.

was set as the best match for the largest Lesion A in the highest dose in the sharper reconstruction. The volume size of all three differently spiculated lesions, in all four different scales, were computed and compared to the ground truth.

III. Results

The volume registrations (Fig. 2), along with the individual registrations of each lung lesion, were successful. Fig. 3 and Fig. 4 display the central axial slice of each lesion. It is evident that as the dose decreases, the lesion spiculations become less conspicuous. However, even at the lowest dose (1.6 mGy), lesions larger than approximately 15 mm remain visually identifiable. For lesions smaller than 15 mm, challenges arise in accurately recovering their shape. Concurrently, root mean squared errors (RMSEs) of the CT comparing to the ground truth images of the lesions increase with decreasing dose, showing a notable change around the lowest dose, 1.6 mGy (Fig. 5).

Fig. 6 illustrates the effect of lesion scale on RMSE. When segmenting the smaller lesions, where the spiculated surface becomes indiscernible, the surface of the segmentation appears smoother. Additionally, significant change in volume size is observed when the lesion is too small (Fig. 7).

IV. Conclusion

In this study, we explored the capabilities of pulmonary imaging using PCD-CT, specifically focusing on the characterization of aggressive feature from lung lesions. We designed and 3D-printed a phantom containing various types of digital lung lesions in different sizes. Our findings revealed that PCD-CT offers accurate evaluation of lung nodules, paving the way for finding the optimized imaging protocol to access lesion features, and for providing better diagnosis and treatment plans for patients. This approach showcased the ability to bridge digital image system evaluation with actual experimental validation. Such a combination holds great promise in expediting the clinical adoption of new CT technologies.

V. Acknowledgement

We acknowledge support through the National Institutes of Health (R01EB035092) and Canon Medical Systems. We also thank the Center for Virtual Imaging Trials (CVIT) at Duke University for providing the lung lesion models.

VI. References

[1] S. Hsieh, et al. "Photon counting CT: clinical applications and future developments." *IEEE transactions on radiation and plasma medical sciences* 5.4 (2020): 441-452.

[2] K. Mei, et al. "Three-dimensional printing of patient-specific lung phantoms for CT imaging: Emulating lung tissue with accurate attenuation profiles and textures." *Medical physics* 49.2 (2022): 825.

[3] N. Shapira, et al. "Three-dimensional printing of patient-specific computed tomography lung phantoms: a reader study." *PNAS Nexus* 2.3 (2023).

[4] K. Mei, et al. "Design and fabrication of 3D-printed patient-specific soft tissue and bone phantoms for CT imaging." *Sci. Rep.* 13, 17495 (2023)

[5] R. Sauer, et al. "Development and CT image-domain validation of a computational lung lesion model for use in virtual imaging trials." *Medical Physics* 50.7 (2023): 4366-4378

SinoFlow: Fluid flow estimation from sinograms using a physics informed neural network

Gurnoor Singh Khurana, Jinyuxuan Guo, Lauren Severance, Alejandro Gonzalo, Bahetihazi Maidu, Juan C. del Alamo, Francisco Contijoch

***Abstract*—Computed Tomography (CT) is clinically used to evaluate cardiovascular vessels' anatomy and size. CT has high spatial resolution and the ability to obtain volumetric data which makes it an appropriate choice for this task. Moreover, it is quite robust and can be done across a range of patient conditions. In terms of cardiovascular diseases, measuring accurate flow information is of great importance. However, this information is not typically obtained from CT images. This is because there does not exist a quantitative analysis tool to evaluate flow directly from images. As a result, flow estimation using CT relies on modeling approaches such as computational fluid dynamics (CFD). However, the results obtained using CFD can be sensitive to modeling choices (e.g., blood rheology) and boundary conditions (e.g., upstream/downstream vascular impedances).**

Physics-Informed Neural Networks (PINNs) can reconstruct flow fields from sequences of images depicting the dynamics of transported scalars like CT contrast agents. However, the potential of PINNs for CT imaging remains underexplored. One challenge associated with this reconstruction is that the scalar field changes during the interval of time involved in acquiring each CT image. Thus, an accurate flow reconstruction requires disentangling the fluid motion within each frame from the motion between different frames. In another submission to this conference, we describe how these flow motion artifacts limit the capability of PINNs to evaluate flows imaged with CT. In this work, we propose SinoFlow: a PINN that includes a model of the scalar transport during CT image acquisition to bypass image reconstruction and use the raw sinogram data directly for flow prediction.

***Index Terms*—Computed Tomography (CT), Cardiovascular CT Imaging, Physics Informed Neural Networks (PINNs), Deep Learning (DL)**

I. Introduction

Blood flow in heart vessels and chambers is an important indicator of cardiac health. Imaging modalities such as echocardiography and magnetic resonance imaging (MRI) are clinically used to measure flow since fluid velocity can be encoded in the imaging signal [1]. However, no equivalent modalities exist to assess flow in vessels directly using CT. As a result, clinical evaluation of CT images indirectly assess flow by evaluating changes in the size of vessels, looking to detect changes in vessel caliber or obstructions, or evaluation downstream consequences (lack of iodine enhancement in the parenchymal tissue).

CT has volumetric spatial coverage and high spatial resolution. Therefore, a method to evaluate flow using CT could have significant impact. However, a key limitation of CT in the setting of imaging cardiovascular flow is the presence of motion artifacts. [3] In some cases, artifacts arise due to voluntary movements which can be eliminated/reduced. However, in cases of involuntary movement, several methods have been developed to alleviate the impact of these artifacts. [10], [9], [5], [2], [6], [7] One such approach, outlined in [4] makes use of acquired sinogram to learn an implicit representation of the underlying object. This was shown to provide better results than the classical reconstruction methods, such as Filtered Back Projection (FBP), since FBP relies on the assumption that no motion occurs during scanning.

Contemporary work on flow field reconstruction [8] makes use of Physics Informed Neural Networks (PINN). In this approach, movies of the flowing concentration field are used as the supervision to train a neural network that predicts the flow field. In this abstract, we will refer to this framework as ImageFlow. Although this approach has shown promise to infer flow metrics associated with adverse events (e.g., atrial blood stasis an stroke [11]), its application to full flow reconstruction from CT imaging critically depends on image reconstruction accuracy. In particular, we have observed that the quality of ImageFlow is limited when the flow significantly deforms the contrast agent over a timescale comparable to the interval required to acquire an image.

In this work, we propose SinoFlow, a novel PINN that incorporates fluid motion into the image rendering, thus accurately reconstructing flow velocities and eliminating motion artifacts. In SinoFlow, the acquired sinogram data is used to learn an implicit representation of the contrast evolution movie and its CT acquisition process. To this end, SinoFlow optimizes losses from equations modeling contrast transport, fluid mass conservation, and fluid momentum balance, to learn the flow field in vessels. SinoFlow is trained using the raw sinogram data instead of FBP reconstructed data, leading to more precise contrast and flow field reconstructions.

II. Method

For both ImageFlow and SinoFlow, a multi-layer perceptron network was used to generate an implicit representation of the concentration field in the contrast evolution movie as well as to learn velocity and pressure fields by optimizing over

GSK is from the Department of Computer Science and Engineering, UC San Diego, La Jolla CA USA. GJ and LS are from the Shu Chien-Gene Lay Department of Bioengineering, UC San Diego, La Jolla, CA USA. AGG, BM, are from the Department of Mechanical Engineering, University of Washington, Seattle, WA, USA, JCdA is from the Department of Mechanical Engineering and Division of Cardiology, University of Washington, Seattle, WA, USA,FC is from the Shu-Chien-Gene Lay Department of Bioengineering and Department of Radiology, UC San Diego, La Jolla, CA, USA and Rady Children's Hospital San Diego, San Diego, CA

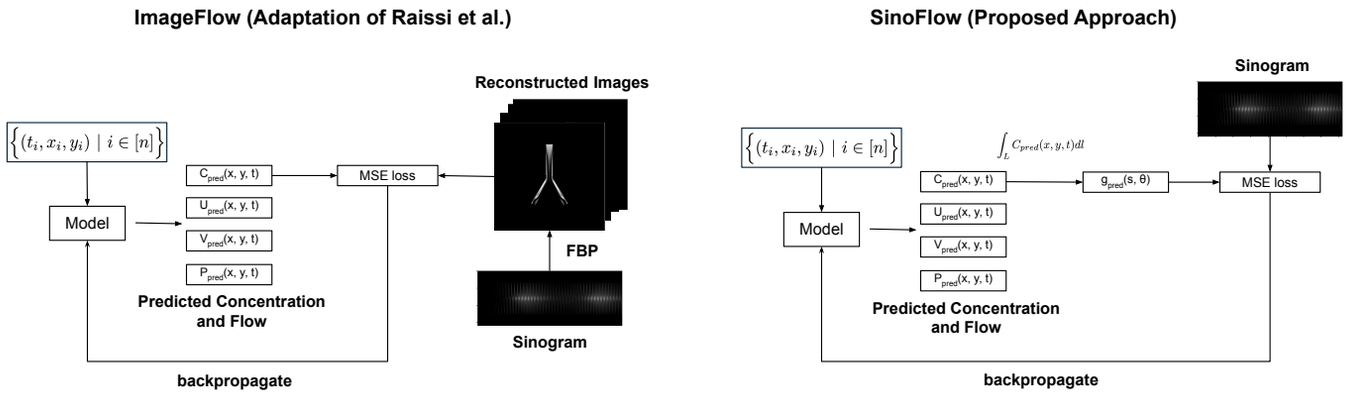


Fig. 1: Figure depicting differences between the initial and proposed approaches (ImageFlow, left and SinoFlow, right). In both cases, the same image domain NN is used to predict concentration and flow field movies. However, SinoFlow uses sinogram for supervision, while ImageFlow uses FBP concentration values for supervision

partial differential equations modeling the contrast transport, fluid mass conservation, fluid momentum balance (i.e., the Navier-Stokes equations). This can be seen in Figure 1. The input to the network is of the form (t, x, y) where t represents the time elapsed in contrast evolution movie, and x, y are the spatial locations of a point in space. In case of ImageFlow, these points are randomly sampled from the spatial domain of simulation. For SinoFlow, a set of collinear points is generated corresponding to randomly sampled locations on the sinogram. This set of points is used as the input to the network. The network outputs a quadruplet of values (c, u, v, p) which represent the concentration, magnitude of velocity along x direction, magnitude of velocity along y direction and the pressure value at the specified spatiotemporal position.

For SinoFlow, the sinogram of the contrast evolution movie was used for supervision instead of the artifact-prone, reconstructed concentration field, with the model jointly optimizing over the mean squared error of sinogram and the flow model equations losses. More details on loss function are discussed in section II-B.

A. Network Architecture

A Multi Layer Perceptron network with 10 hidden layers and 200 neurons in each layer was used for both approaches. For the activation function, we use the Sigmoid Linear Unit (SiLU). This model was implemented in Python using the PyTorch library.

B. Loss function

The loss function is a combination of loss over sinogram and physics based terms. Mathematically,

$$\mathcal{L} = \lambda_1 \mathcal{L}_{sino} + \lambda_2 \mathcal{L}_{flow}$$

where λ_1 and λ_2 are hyperparameters. In our experiments, we set $\lambda_1 = \frac{1}{n_p}$, where n_p is the number of spatial points used for calculating sinogram value at a particular location, and $\lambda_2 = 1$. In our experiments n_p was equal to 1600. These values were experimentally tuned to ensure that both sinogram and flow information are given adequate weight in model supervision.

Mathematically,

$$\hat{g}(s, \theta) = \int_{L(\theta)} c[x(s, \theta), y(s, \theta), t] dl$$

$$\theta = \theta_0 + \omega t$$

$$\mathcal{L}_{sino} = \frac{1}{n} \sum_{n} \left(g(s, \theta) - \hat{g}(s, \theta) \right)^2$$

$$\mathcal{L}_{flow} = e_1^2 + e_2^2 + e_3^2 + e_4^2$$

$$e_1 = c_t + (u * c_x + v * c_y)$$
$$e_2 = u_t + (u * u_x + v * u_y) + p_x - \frac{1}{\text{Re}}(u_{xx} + u_{yy})$$
$$e_3 = v_t + (u * v_x + v * v_y) + p_y - \frac{1}{\text{Re}}(v_{xx} + v_{yy})$$
$$e_4 = u_x + v_y$$

where $[x(s, \theta), y(s, \theta)]$ represents the projection ray for gantry angle θ, which varies along time as $\theta_0 + \omega t$, with ω being the gantry rotation speed. In the flow equations, Re is the Reynolds number, which indicates the ratio of inertial forces vs. viscous forces. In this work, we used a fixed value $Re = 900$ to match our ground-truth CFD simulations, although this parameter is also trainable. This Reynolds number is representative of mid-size arteries like the coronary or cerebral arteries.

The terms e_1 to e_4 correspond to the Navier-Stokes equations. Specifically, e_1 represents conservation of mass, e_2 and e_3 represent conservation of momentum and e_4 represents incompressibility of fluid.

C. Data generation and Training

Flow was simulated for a 2D Y-shaped vessel with pulsatile flow (with a simulated heartrate of 60 bpm). The vessel was 1 cm in diameter and 7 cm in length. Contrast was injected

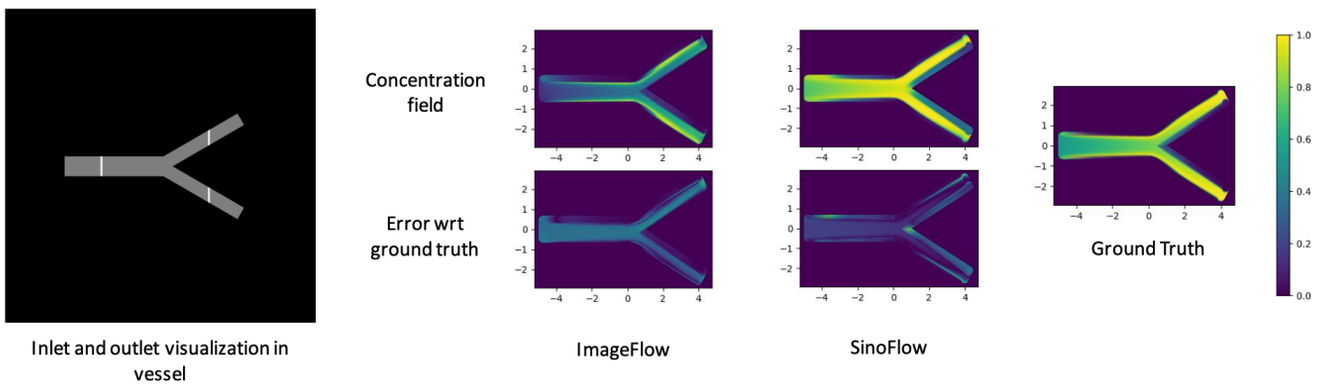


Fig. 2: Figure (left) depicting the inlet and outlet locations in the Y shaped vessel and a figure (right) visualizing the concentration field in ImageFlow and SinoFlow (top) and the error wrt ground truth (bottom)

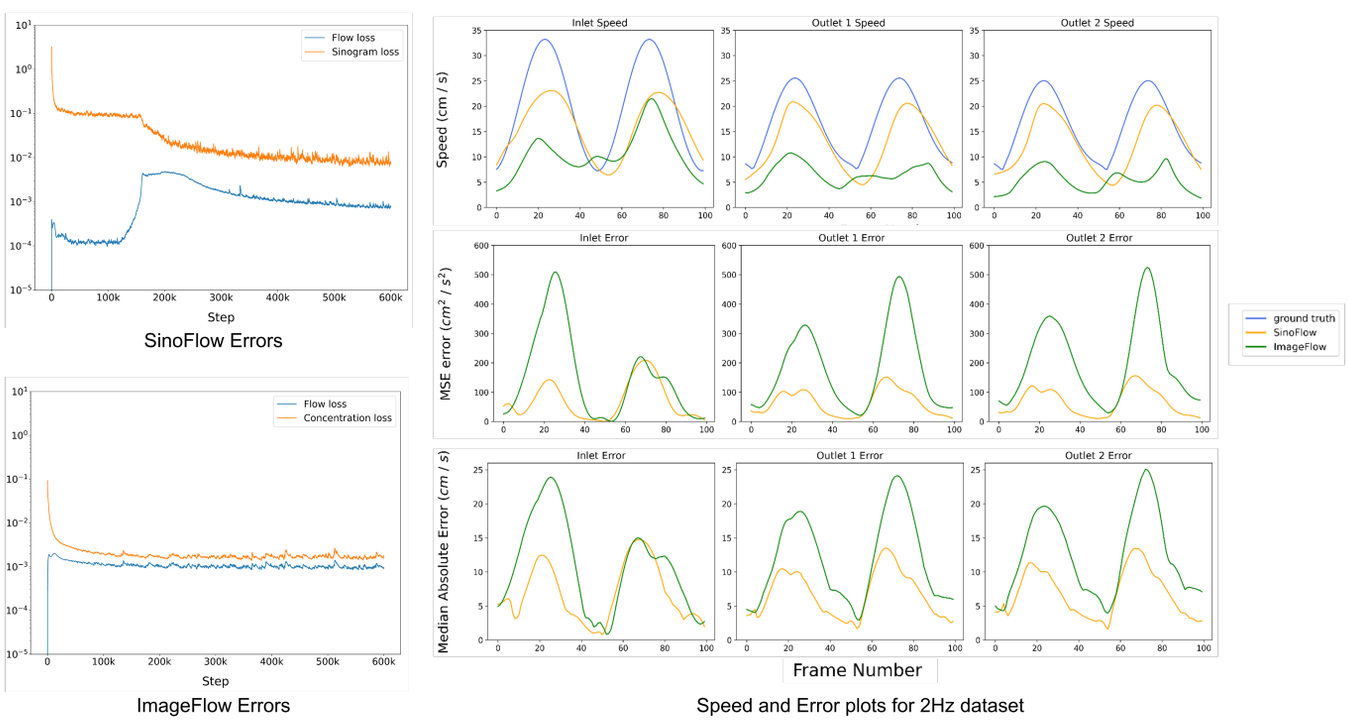


Fig. 3: Figure (left) depicting the loss plots obtained while training SinoFlow and ImageFlow and a figure (right) depicting the comparison of fluid speed predicted by ImageFlow and SinoFlow

sinusoidally into the vessel to generate a gradient of contrast The Y-shaped vessel can be seen in Figure 2 with the location of inlet and outlet measurements shown in white.

We simulated CT imaging by creating a sinogram using a parallel beam geometry and gantry rotation speed of 2Hz.

SinoFlow was trained on the generated data for 600k iterations. In each iteration we sampled 256 locations on the sinogram. Corresponding to each location, a set of 1600 spatiotemporal points, i.e. (t, x, y), were generated, such that the summation of concentration across these points represents the sinogram value.

The model was trained using the Adam optimizer with a learning rate of 0.001. The training was done using an Nvidia RTX 3090 GPU, and it took approximately 48 hours to complete. Since SinoFlow and ImageFlow use the same underlying architecture, the per iteration training time of both approaches is similar. However, SinoFlow takes significantly more time for convergence.

This is expected, since SinoFlow relies on indirect supervision and significantly smaller training data (i.e. line integral of concentration values) whereas ImageFlow makes use of direct supervision (i.e. full concentration value images).

D. Data Sampling

As we see in Figure 2, most of the image consists of zeros. Therefore, if the whole domain was sampled uniformly, most of the components of the line integral would have zero

concentration and flow. This would lead to highly inefficient training.

To correct this situation, we filtered the input to the network, to ensure that only the points that fall inside the vessel are used for model training. This improved the training.

III. Experiments and Results

We compared ImageFlow and SinoFlow at three different levels. First, we evaluated differences during training in terms of losses. Second, we compared the concentration fields relative to the ground truth. Third, we evaluated differences in the resulting flow estimates at the inlet and outlet locations. Our overall hypothesis is that SinoFlow will improve flow estimation accuracy at the 3 locations of interest (relative to ImageFlow) by creating a better estimate of the concentration field which in turn, leads to a better estimate of the flow fields.

A. Experiment 1: Comparison of Training

Figure 3 shows the plot of concentration and flow loss terms over the number of steps elapsed in training for ImageFlow and SinoFlow. ImageFlow is extremely quick in achieving convergence on all losses. In less than 100k iterations, the values of the training losses converges.

The training plot for SinoFlow shows that the model begins with a low error in the flow field. Presumably, the network starts by predicting a zero flow field which is incorrect but consistent with the Navier-Stokes terms. After approximately 150k iterations, the flow errors increase and then start to decrease again. Our interpretation is that when the model begins training, it has a large error in terms of concentration. Reducing the concentration error introduces errors in the flow field in terms of deviations from Navier-Stokes terms. However, once the concentration error decreases, then the network aims to lower its loss by improving its estimate of the flow errors.

The delayed learning of SinoFlow is justified, since the supervision using sinogram is an indirect supervision, compared to ImageFlow, where the model has a simpler objective function to match a given concentration field.

B. Experiment 2: Comparison of concetration images generated with SinoFlow and ImageFlow

ImageFlow aims to learn or represent the concentration field obtained using Filtered Back Projection (FBP). However, in SinoFlow, the sinogram is used directly as the supervision, and the optimization generates a prediction of the concentration field along with the flow fields.

Figure 2 shows a visual comparison of the concentration field predicted by ImageFlow and SinoFlow. SinoFlow is able to learn a much better representation of the concentration field than ImageFlow.

C. Experiment 3: Impact of reconstruction accuracy on flow field

The flow fields generated by SinoFlow and ImageFlow were evaluated at three locations in the vessel: the inlet, the outlet 1 and the outlet 2. Figure 3 shows the comparison of speed predicted by SinoFlow and ImageFlow over time. SinoFlow performs considerably better than ImageFlow at the outlets. At the inlet, SinoFlow performs better than ImageFlow for the most part, except for a small section towards the end.

In summary, our results indicate that SinoFlow will be useful in the scenarios where CT scanning is done using low frequency.

IV. Conclusion

Our experiments show that the accuracy of flow field prediction can be improved by using a better estimate of the concentration field. However, this comes at the cost of increased training time. With the gantry rotation speed being the key limitation in modern day CT scans, we believe that our study will have a profound impact in improving the quality of flow estimation.

References

[1] Javier Bermejo, Pablo Martínez-Legazpi, and Juan C del Álamo. The clinical assessment of intraventricular flows. *Annual Review of Fluid Mechanics*, 47(1):315–342, 2015.

[2] Roshni Bhagalia, Jed D Pack, James V Miller, and Maria Iatrou. Nonrigid registration-based coronary artery motion correction for cardiac computed tomography. *Medical physics*, 39(7Part1):4245–4254, 2012.

[3] Francisco Contijoch, J. Webster Stayman, and Elliot R. McVeigh. The impact of small motion on the visualization of coronary vessels and lesions in cardiac ct: A simulation study. *Medical Physics*, 44(7):3512–3524, 2017.

[4] Kunal Gupta, Brendan Colvert, and Francisco Contijoch. Neural computed tomography, 2022.

[5] Maria Iatrou, Jed D Pack, Roshni Bhagalia, Dirk Beque, and John Seamans. Coronary artery motion estimation and compensation: a feasibility study. In *IEEE Nuclear Science Symposuim & Medical Imaging Conference*, pages 2819–2821. IEEE, 2010.

[6] JD Pack and B Claus. An analysis of motion artifacts in ct and implications for motion compensation. In *Proc. 2nd Int. Conf. Image Formation X-Ray Comput. Tomograph.*, pages 322–325, 2012.

[7] Jed D. Pack, Ashish Manohar, Sathish Ramani, Bernhard Claus, Zhye Yin, Francisco J. Contijoch, Andrew J. Schluchter, and Elliot R. McVeigh. Four-dimensional computed tomography of the left ventricle, part i: Motion artifact reduction. *Medical Physics*, 49(7):4404–4418, 2022.

[8] Maziar Raissi, Alireza Yazdani, and George E. Karniadakis. Hidden fluid mechanics: A navier-stokes informed deep learning framework for assimilating flow visualization data. *CoRR*, abs/1808.04327, 2018.

[9] Cameron J Ritchie, Carl R Crawford, J David Godwin, Kevin F King, and Yongmin Kim. Correction of computed tomography motion artifacts using pixel-specific back-projection. *IEEE Transactions on Medical Imaging*, 15(3):333–342, 1996.

[10] D Schäfer, M Grass, et al. Motion compensated cone beam filtered back-projection for 3d rotational x-ray angiography: A simulation study. In *Proc. of the Conference on Fully 3D Reconstruction in Radiology and Nuclear Medicine, Fort Douglas/Olympic Village, Salt Lake City, Utah, USA*, 2005.

[11] Åshild Telle, Clarissa Bargellini, Yaacoub Chahine, Juan C Del Álamo, Nazem Akoum, and Patrick M Boyle. Personalized biomechanical insights in atrial fibrillation: opportunities & challenges. *Expert Review of Cardiovascular Therapy*, pages 1–21, 2023.

3D printed anthropomorphic phantoms for CT imaging – Are their attenuation properties consistent for different image reconstructions and CT manufacturers?

Irene Hernández-Girón, PhD.

***Abstract*—Image quality in Computed Tomography is frequently assessed with relatively simplistic test objects (phantoms), neither representative of the patients' anatomy or morphometry, nor linked to the diagnostic tasks performed by radiologists and clinicians. There is a need to incorporate anthropomorphic phantoms in quality control and optimization processes of imaging systems. Such phantoms should be somewhat representative of patients and designed with relevant clinical goals in mind. The use of commercially available anthropomorphic phantoms is not broadly extended and they have limited capabilities to represent the widely diverse cohorts of patients.**

We propose 3D printing as a manufacturing technique to create customized and affordable anthropomorphic phantoms tailored to different diagnostic tasks, in particular for thorax CT imaging. The generalizability of the properties of 3D printed anthropomorphic phantoms among different reconstruction methods and CT manufacturers has not been answered yet. In this paper, anthropomorphic lung vessel inserts and parenchyma disease were mathematically modelled and manufactured with tissue equivalent 3D printing materials. Their X-ray attenuation, (Hounsfield Units, histograms) and visual evaluation was characterized. Clinical thorax protocols (same kV) were selected from two CT manufacturers, iterative reconstruction and two reconstruction filters The results were compared with a state of the art commercial anthropomorphic thorax phantom. Differences in attenuation and pixel value distributions were observed, with a shift towards higher values when sharper kernels were selected in both manufacturers for 3d printed and commercial phantoms. Though the selected clinical CT protocols were set at the same kV, differences in behavior were observed for both the 3D printed and commercial phantoms. 3D printed phantoms for diagnostic imaging require a thorough evaluation for the relevant range of energies and specific imaging systems, to guarantee that the materials used are representative of the patients scanned with equivalent protocols.

***Keywords*—Computed Tomography, 3D printing, phantoms, anthropomorphic, modelling, image quality, image reconstruction**

Irene Hernandez-Giron (IHG) is an Assistant Professor, Ad astra fellow at the School of Physics in University College Dublin (UCD), Ireland (email: irene.hernandezgiron@ucd.ie)

Funding: IHG has an Ad astra fellowship (UCD) (2023 -)

Another funding relevant for this paper (projects already finished):

IHG started their research on 3D printing anthropomorphic phantoms with a NWO-Veni personal grant (2019-2022). Pr. Nr. 17378 – Through the eyes of AI. Held at the Leiden University Medical Center (LUMC)

The initial lung vessel code version, seed of the one used in this paper was created by Chiel den Harder and Irene Hernandez-Giron within the NWO funded project: CLUES (Clinical Image Quality Assessment), led by Dr. Veldkamp (2015-end 2019). Project number: 13592. Held at the Leiden University Medical Center (LUMC)

.

I. INTRODUCTION

THE framework for image quality assessment in Computed Tomography (CT) in guidelines published by medical physics societies and other professionals needs to be updated to account for the latest technological developments in the field [1]. The implementation of iterative reconstruction and AI-based reconstruction, both with the promise of reducing the patient doses, has introduced challenges and exposed the limitations of the test objects (phantoms) and image quality metrics traditionally accepted [2]. Such objects do not usually represent the overall attenuation, morphometry or shape of the patient. The classical technical image quality metrics cannot capture the diagnostic image quality required for specific clinical tasks [3,4]. Anthropomorphic phantoms are commercially available but they represent a very limited cohort of patients and disease inclusion, if any. Their relative high cost together with the budget restrictions in healthcare have slowed down the inclusion of such phantoms in our hospitals and very few have access to them.

3D printing has shaken the manufacturing process in many fields, including Medicine, ranging from customized patient prosthesis to radiotherapy bolus and applications for diagnostic imaging. Among them, the generation of affordable customized test objects or phantoms for the evaluation of diagnostic imaging and radiotherapy systems is expanding fast [5]. The actual composition of the 3D printing materials is usually protected (partially) for intellectual property reasons and it can also be altered to adapt to general physical, chemical or other requirements related to the main uses of such material. This can affect greatly the material X-ray attenuation properties (for instance if metal doping is added in a new version of the same polymer) and needs to be addressed for the application of image quality phantoms in Radiology.

The goal of this study is to create customized

anthropomorphic lung vessel for CT thorax evaluation and to evaluate the influence of kV, CT systems and selected image reconstruction on the 3D printing selected materials attenuation and tissue equivalence potential.

II. METHODS

A mathematical model of the human lung vessel tree was generated based on the work by Weibel [6]. The code, developed in Matlab, was modified based on previous work [7] to generate a 6 cm long test object and the vessels being encased in an ellipse of 10x15 cm to help with stability during 3D printing, easiness of transportation and protect the vessels from external damage (Fig.1). The lung vessels (with diameters between 10mm and 1 mm) started growing on one side of the ellipse decreasing in diameter and length towards the elliptical container and the phantom length in z. Different expansion factors in the XY plane and the Z plane were applied to account for the longitudinal vessel growth in the z-direction. Gaps were left on purpose in the model to include lung parenchyma or other lung sections inserts in the future. By design two pockets of vessels were generated as "closed" to enable trapping the printing powder material to evaluate its potential as lung parenchyma surrogate.

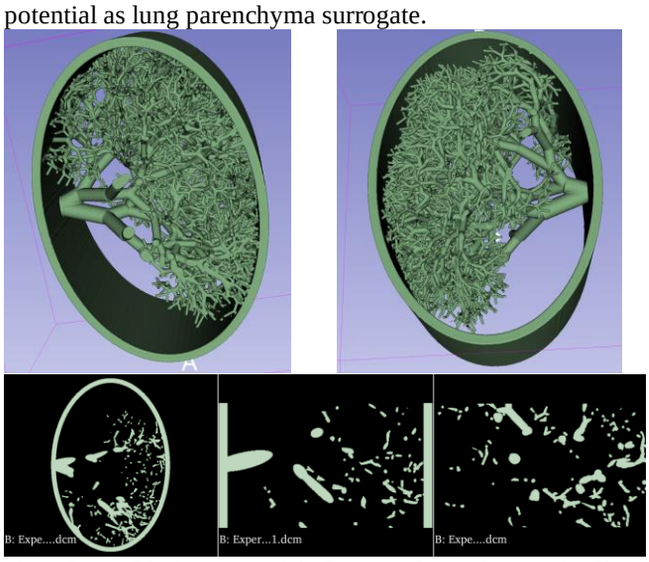


Fig. 1. 3D model of a section of the lung vessel tree, shown in 3D Slicer. Top row shows the volume from both phantom sides. Bottom row shows the central part of the phantom in axial, coronal and saggital views

An elliptical model of the thorax was designed (FreeCAD) (200x300x60 mm) with two identical insert gaps two fit the lung inserts and a 3cm diameter hole to include a spine surrogate. It was manufactured with a CNC machine, in-house at the workshop. The spine insert was made with a section of a Teflon rod, cut and milled to fit the insert.

After performing a screening of 3D printing materials, based on scanning small cubic samples (5x5x5cm) for a range of kV values, it was decided to explore TPU (ρ=1.1g/cm^3) and different variations of PA (PA11, ρ=1.05g/cm^3 and PA12, ρ=1.01g/cm^3) as potential tissue equivalent for lung vessels. Solid and hollow samples (filled with the "raw" printing materials), were evaluated. The selected technique was MJF from HP, which allowed for the resolution required for the designed vessels, to remove easily the "support" material and reduce potential deformations during the printing process of the vessels as the powder acts effectively as scaffolding of the printed structures. The attenuation of the 3D printing materials was evaluated taking histograms and ROIs inside images of the cubic samples (hollow and solid) for a range of kV in a Canon Aquilion ONE Genesis.

After this pilot, TPU was selected as a viable candidate material due to its attenuation properties and also flexibility compared to the other materials, which is desired to avoid small vessels breakage and allow for the future insertion of lesion-like objects in the phantom. Two identical TPU lung vessel inserts were 3D printed (mirrored versions of each other) with an external provider (detail in Figure 2). The phantoms were ordered at different points in time to evaluate in the future the accuracy and reproducibility of the phantoms printing and selected materials, when the process is externalized. The total costs were below 400 euros.



Fig. 2. Left: one of the 3D printed lung vessel inserts made of TPU (MJF, HP). Right: the insert placed inside the thorax PMMA holder.

The thorax 3D printed phantom was scanned in two commercial CT systems, selecting high resolution clinical protocols. The kV was kept the same (120kV), as well as slice thickness and spacing (1mm) and automated tube current modulation was allowed. The systems were a Canon Aquilion ONE Genesis and a Siemens Somatom Edge. For image reconstruction the default patient protocol filters (one soft, one sharp) and iterative reconstruction levels were selected. The CTDIvol was recorded. The CT manufacturers will be called system 1 and system 2 now on, as the goal of this study is not two compare the systems between them, but to study potential differences observed in the phantom images and 3D printed materials attenuation.

A commercial state of the art thorax phantom (Lungman, Kyoto Kagaku) was scanned in a Siemens Somatom Definition Flash with the clinical high resolution thorax protocol and 120kV based on the same image reconstruction setup as above [REFERENCE KYOTO Kagaku]. This phantom represents a small/average male patient with a realistic lung vessel distribution and contains different organs and bones created with uniform solid materials. Lung vessels and soft tissue are made of polyurethane (density 1.06 g/cm^3) and bones of a mixture of epoxy resin and calcium carbonate according to the manufacturer.

The dose (effective mAs and $CTDI_{vol}$) were recorded for the

3D printed in-house and commercial phantoms. The attenuation of the 3D printed and commercial phantom relevant tissues and its dependence with reconstruction method and kV, was evaluated placing regions of interest and measuring the mean pixel value and with histograms. When different CT image reconstructions were available the same locations to measure the histograms and ROIs were selected. The volumes included in the analysis in both phantoms were kept as similar as possible, considering that the ROIs were in some cases drawn freehand to account for the changing size of the lungs in the z-direction in the Lungman phantom.

III. RESULTS

The HU values of the screened materials as a function of kV can be seen in Fig. 3. It is interesting to notice that the materials in powdered (compacted) form shown quite low attenuations that could be compatible with consolidations in the lung and other lung parenchyma diseases. TPU in solid form had the higher attenuation and the lowest in powder form. PA11 was the closest to water in solid form and PA12 could be a good surrogate for fatty tissue. All the materials increased their attenuation with increasing kV.

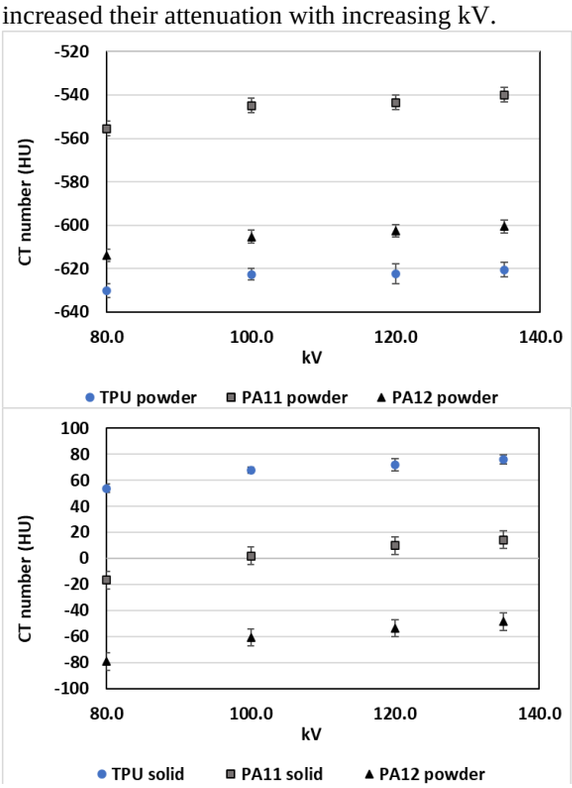


Fig. 3. CT values (HU) of the screened 3D printed materials in the original powder form (top) and solid (bottom) for a range of kV values in CT system 1.

A selection of CT images of the thorax PMMA phantom and the 3D printed lung inserts in both CT systems is shown in Figure 4 (120kV).

The CTDIvol selected by system 2 with AEC, for the PMMA thorax phantom and two 3D printed lung inserts was 2.75 mGy. For the same CT manufacturer and the commercial thorax phantom, it was 3.58 mGy. It is to be noted that the scanned length of the commercial phantom was the full thorax, and for the customized 3D printed phantom just 6 cm. The dimensions of the thorax PMMA holder are similar to the commercial phantom at its largest point (~210x308 mm).

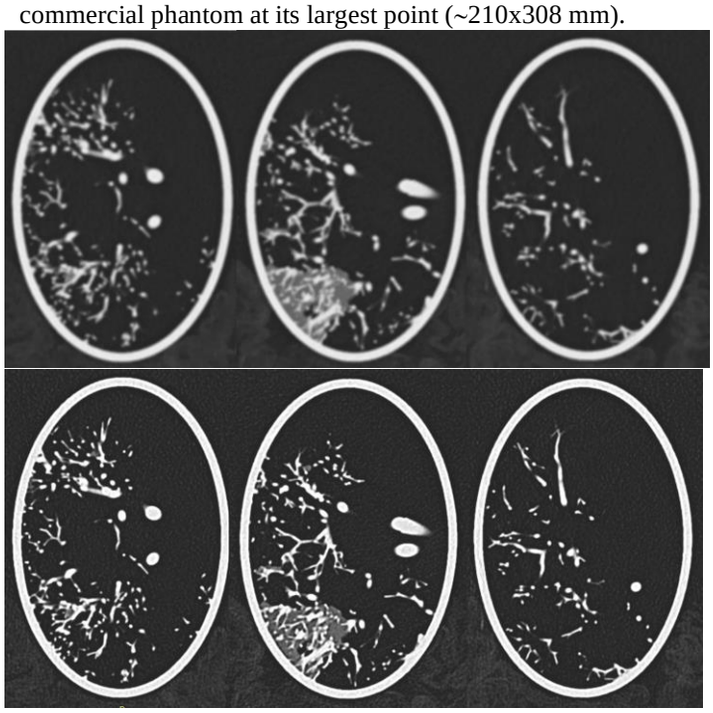


Fig. 4A. Sample CT images of one of the lung inserts from system 2, top row with a soft filter and bottom row with a sharp filter. The central slice shows at the bottom left, the powder trapped to mimic lung parenchyma disease.

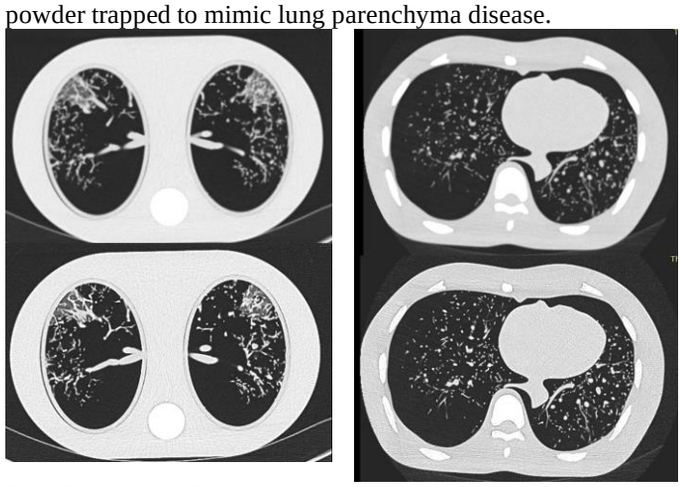


Fig. 4B. Sample CT images in system 2 of the full thorax phantom with 3D printed inserts (left) and the commercial thorax phantom (right) reconstructed with a soft (top row) and sharp (bottom) filter.

The histogram results for the 3D printed lung phantoms are displayed in Fig. 5, comparing the values measured on equivalent ROIs inside one of the 3D printed lung inserts in systems 1 and 2, selecting a soft reconstruction filter (Fig. 5 top) and comparing both reconstruction filters (soft and sharp) for system 2 (bottom). The peak related to air has been omitted. The first peak to the left is related to the areas designed to trap printing powder and simulate lung parenchyma disease, the second peak is related to the vessels (TPU in solid form). There is a clear change in the shape of the histograms between both systems, selecting similar clinical

protocols and filters. The peaks of related to the 3D printed material in both forms are reached for similar HU values (TPU powder-lung parenchyma at ~ -640 HU-system 1 and ~ -660 HU –system 2; TPU solid – lung vessels at ~ 65 HU – system 1 and ~88 HU -system 2). The values obtained in the pilot study with the cubes in system 1 were ~ -620 HU (TPU powder) and ~70HU (TPU solid).

The influence of the selected kernel on the histogram distribution is shown in Fig.5 (bottom) observing a flattening trend for the sharp kernel, hampering the peaks identification related to the lung vessels and diseased lung parenchyma.

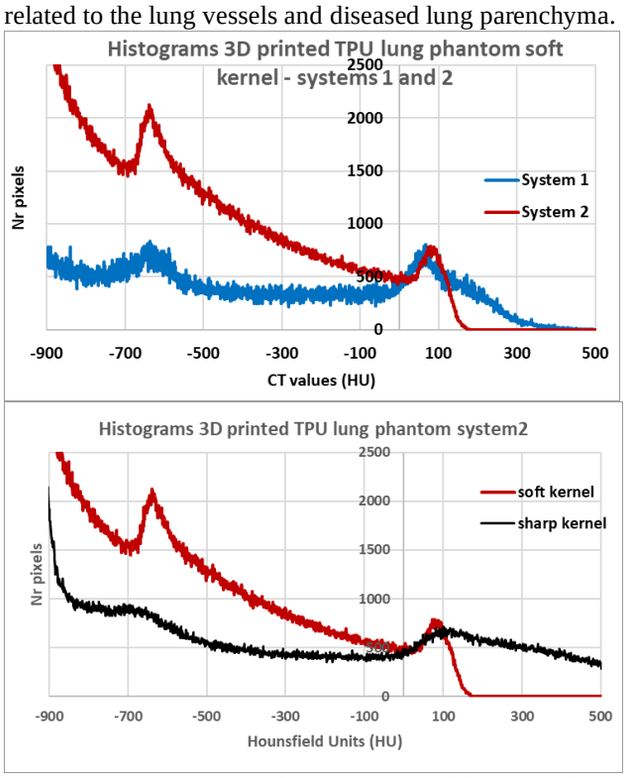


Fig. 5. Histogram analysis for the 3D printed lung vessel insert in both CT systems for a soft kernel (top) and comparing a soft and sharp kernel for system 2 (bottom).

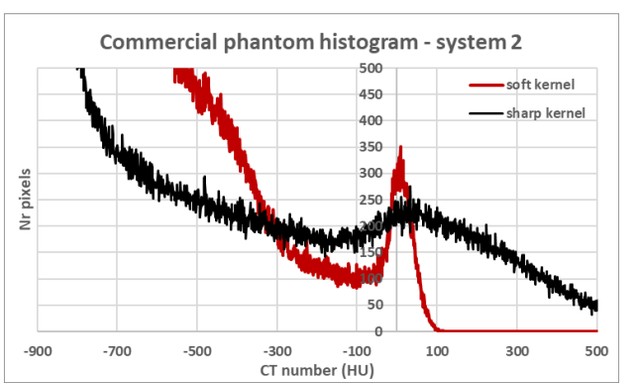


Fig. 6. Histogram analysis for the 3D printed lung vessel insert in both CT systems for a soft kernel (top) and comparing a soft and sharp kernel for system 2 (bottom).

The histogram analysis for the commercial thorax phantom and system 2 with both reconstruction kernels is shown in Fig. 6. Note that it does not exist a peak for lung parenchyma as there is not any in the phantom, but a continuous decay of CT values can be observed from the peak for air due to partial volume effect. The attenuation of lung vessels is around ~10HU for the soft kernel images and follows a very flat distribution for the sharp filter, slightly shifted towards higher CT values compared to the latter.

IV. DISCUSSION AND CONCLUSIONS

An affordable thorax phantom, including lung vessel inserts and lung disease was designed and 3D printed to evaluate image quality in CT. Its properties, in terms of the delivered dose for equivalent clinical protocols dimensions and materials attenuation were compared with a commercial phantom. The materials Hounsfield Units varied to different extents when scanning the commercial and 3D printed phantom in CT systems of different manufacturers and for different reconstruction options. More research is needed to determine the equivalence of commercial and customized 3D printed phantoms to patients and their generalizability. CT technology is advancing towards more complex AI-based image reconstruction and quantitative imaging, especially with spectral and photon counting CT, which will require phantoms designed to evaluate such advances. The phantom was limited in size, especially in the z-direction, which can be expanded without increasing costs. Other 3D printed materials will be explored in the future and more CT manufacturers included in the study. The accuracy and reproducibility of 3D printed phantoms will be evaluated using micro-CT imaging as an independent imaging modality.

V. AKNOWLEDGEMENTS

We would like to thank the following collaborators for their contributions to this study: Barbara Snow, Senior Radiographer at Saint Vincent's University Hospital (SVUH) for her help with CT scanning. Marcel Greuter, Medical Physicist at the University Medical Center (UMCG) in The Netherlands for providing the scans of the commercial phantom. Wouter Veldkamp (LUMC) for leading the seminal project for phantom development a decade ago. Chiel den Harder, Medical Physicist at the Amsterdam Medical Center (AMC) for the initial version of the lung vessel software in 2014. John Egan (Technical Officer) and Richard Byrne (Senior Technical Officer) of the School of Physics Workshop (UCD) for their assistance with the thorax PMMA phantom.

VI. REFERENCES

1. Performance Evaluation of Computed Tomography Systems. The Report of AAPM Task Group 233 (2019)
2. Verdun et al. Image quality in CT: From physical measurements to model observers. Physica Medica (2015)
3. Willemink MJ, Noël PB. The evolution of image reconstruction for CT-from filtered back projection to artificial intelligence. Eur Radiol (2019)
4. Filippou V, Tsoumpas C. Recent advances on the development of phantoms using 3D printing for imaging. Med Phys (2018)
5. Weibel ISBN: 978-3-642-87553-3 Morphometry of the human lung (1963)
6. Hernandez-Giron et al. Development of a 3D printed anthropomorphic lung phantom for image quality assessment in CT Physica Medica (2016)
7. Kyoto Kagaku website

Evaluation of the detectability index of CT thorax AI-based reconstruction protocols

Elsa B. Pimenta, Luuk J. Oostveen and Paulo R. Costa

***Abstract*—Diagnostic image quality can be quantified using the detectability index, d'. In this investigation, the image quality resulting from thorax computed tomography (CT) protocols was studied employing a custom-made Mercury phantom. Phantom images were acquired using the Aquilion One Prism CT scanner, and the image reconstruction algorithm employed was AiCE, a deep learning-based technique. Two distinct CT protocols, namely CT thorax and thorax low dose, were implemented. The AiCE algorithm was integrated with Body Sharp and Lung reconstruction kernels, along with aluminum-copper and aluminum-silver filters. This comprehensive approach facilitated the quantification of CT image quality under diverse acquisition parameters. The spatial resolution, as indicated by the Task Transfer Function (TTF), and the noise characteristics, assessed using the Noise Power Spectrum (NPS), were systematically examined. The detectability index was subsequently determined through the analysis of these functions across various acquisition parameters. The impact of $CTDI_{vol}$ on these parameters was also studied.**

***Index Terms* — Detectability index, Image quality, Computed tomography, lung CT**

I. Introduction

The application of the As Low As Reasonably Achievable (ALARA) principle in medical imaging requires the development of tools to balance radiation dose and image quality by managing the trade-off between noise and clinical performance. Nevertheless, the adequacy of a medical image in discerning clinically important tasks depends not solely upon the chosen technology and acceptable dose levels, but also on the observer's performance and expertise. The interrelations among these aspects are of particular significance when the clinical task combines the need for detection of low contrast pathological structures with low doses constraints, such as in lung cancer screening programs [1, 2].

In the last decades, screening trials have been performed in different countries aiming to detect lung cancer early in specific higher-risk populations. More than a decade ago, trials such as the US National Lung Screening Trial (NLST) demonstrated that using CT scans, up to 95% of the detected lung nodules were erroneously classified as malignant [3]. Practical consequences of these false identifications are the addition of costs to the health system, need of additional and sometimes invasive investigations, and unnecessary patient's anxiety. However, recent data from De Koning et al [4] reported that in the evaluation of a cohort of male patients submitted to the Dutch–Belgian lung-cancer screening trial (Nederlands–Leuvens Longkanker Screenings Onderzoek [NELSON]) only 1.2% of the participants presented false positive tests. This demonstrates the substantial improvement on the efficacy of this kind of screening program using low-dose CT (LDCT) and its value for early detection of lung cancer.

The investigation of lung nodule occurrence in high-risk populations through CT images allows the establishment of preventive interventions and treatments in the early stage upon cancer confirmation. This kind of large population CT studies is only possible when balancing low doses with adequate image quality. Therefore, LDCT protocols were introduced in the clinical routine in order to allow adequate lung nodule detectability considering radiation doses as low as possible in order to justify its application in screening programs.

These advances in LDCT have impacted the methods for qualification of CT images, since most of the adopted reconstruction algorithms do not present the linearity and stationarity attributes required in traditional image quality evaluation methods. Methods for assessing the quality of medical images consistent with these reconstruction methods and observer characteristics have lately included the detectability index d' [5], which was first proposed in the 1980s [6, 7]. The detectability index was initially developed to assess how well radars identified objects or detected targets [8]. The metric was then adapted to include human observers and imaging systems [6]. In addition, the detectability of lung nodules is also influenced by the limitations of the observer's performance and radiation dose [9].

This work presents results of the application of a physical phantom with different diameters designed to estimate d'. The task-transfer function and noise power spectrum were also evaluated. These metrics were determined and compared across different CT thorax protocols (regular dose and LDCT) and beam filter combinations and reconstruction kernels.

This paper was submitted for review in 22 January 2024. This work was supported in part by 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).

Paulo R. Costa and Elsa B. Pimenta are with Department of Nuclear Physics, Institute of Physics, University of São Paulo, Brazil (e-mail: pcosta@if.usp.br)

Paulo R. Costa is a guest researcher of the Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands

L. Oostveen is with the Department of Medical Imaging, Radboudumc, Nijmegen, The Netherlands (e-mail: Luuk.Oostveen@radboudumc.nl)

II. MATERIAL AND METHODS

The following sections present (A) a brief overview of the phantom that was employed, (B) the image acquisition parameters and reconstruction techniques, (C) detectability index estimation, and (D) the image quality analysis using the estimated metrics.

A. The custom-made phantom

To perform the detectability index analysis, a Duke University Mercury-based phantom was built at the Institute of Physics Central Workshop [10, 11]. The phantom was manufactured using ultra-high molecular weight polyethylene (UHMW). It has sections of different diameters (120, 185, 230, 300, and 370 mm) with five entries for cylindrical rods (inserts) of different materials (Fig. 1). This design allows to evaluate image quality metrics associated to the performance of the system´s automatic exposure control. In this study, polyurethane inserts were used to evaluate the detectability index for each section diameter and CT chest protocol considering different beam filter combinations and reconstruction kernels.

B. Image Acquisition

The phantom images were acquired with an Aquilion One PRISM CT scanner (Canon Medical Systems, Otawara, Japan) (Fig. 2). Two CT protocols for thorax were used: CT thorax and thorax low dose, both using automatic tube current modulation. The CT thorax protocol used 100 kV and 120 kV, with tube currents ranging from 80 to 700 mA. The thorax low dose protocol, on the other hand, only used 120 kV and currents from 80 to 260 mA. The image reconstruction algorithm was AiCE (Advanced intelligent Clear-IQ Engine). AiCE is a deep learning-based reconstruction technique for CT images that enhances image quality by utilizing a deep convolutional neural network. The AiCE algorithm reconstruction kernels adopted Body Sharp and Lung options combined with acquisitions using aluminum-copper (Al/Cu) and aluminum-silver (Al/Ag) beam filters.

The following acquisition settings were maintained constant for both protocols: slice thickness, 0.5 mm; pitch, 0.813; rotation time, 0.275 s; beam collimation, 0.5x80 mm.

C. Detectability index estimation

The non-prewhitening model observer with eye filter (NPWE) was employed in this paper [12, 13]. The model considers the characteristics of the human visual system and measures image quality in terms of human perception. It is not necessary to use the pre-whitening process in this observer model, which removes correlations between noisy pixels. The detectability index of the NPWE observer in the frequency domain is given by [14-16]:

$$d'^2_{NPWE} = \frac{\left[\int W(f)^2 TTF^2(f) E^2(f) df\right]^2}{\int W(f)^2 TTF^2(f) E^4(f) NPS(f) df} \quad (1)$$

where f is the radial spatial frequency. The $W(f)$ is the task function and it is a representation of the signal to be detected in the Fourier domain:

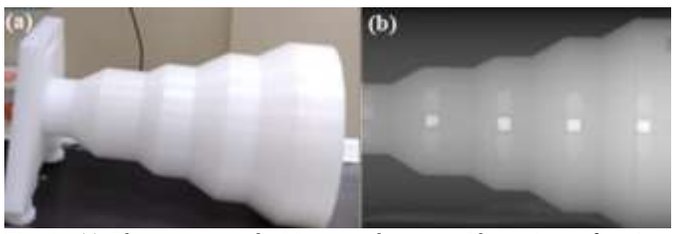


Fig. 1. (a) The custom-made Mercury phantom with sections of various diameters (120, 185, 230, 300, and 370 mm). (b) CT localizer radiographic image of phantom with inserts on the sagittal plane.



Fig. 2. The phantom positioned on the table during the z-axis alignment on the of Aquilion One CT scanner.

$$W = |FT\{h_1(x,y) - h_2(x,y)\}| \quad (2)$$

where FT is the Fourier transformation, In this equation,, $h_1(x,y)$ relates to the hypotheses of the object present and $h_2(x,y)$ of the object not present.

The definition of the task function by the ImQuest software requires the definition of a contras profile function, $c(r)$. As the present study involves lung protocol, the contrast profile used was designed (used to simulate a non-speculated lung nodule), given by [17-19]

$$c(r) = C\left[1 - \left(\frac{2r}{D}\right)^2\right]^n \quad (3)$$

where $c(r)$ is the signal value at a radial distance r, C denotes the peak contrast of the signal with respect to the background in HU, D is the signal diameter, and n is a constant dictating the sharpness of the edge of the contrast profile.

The eye filter function, $E(f)$, in equation (1) used in this investigation was modeled by Eckstein et al [20] and modified by Saunders et al. [21] as an ocular filter model as

$$E(f) = \left|\eta \rho^{a_1} e^{-a_2 \rho^{a_3}}\right|^2 \quad (4)$$

where η is a normalization factor, the parameters a_1, a_2 and a_3 are equal to 1.50, 3.22 and 0.68, respectively, and ρ is the radial spatial frequency (in cycles/degree).

D. Image Analysis

The phantom images were processed using imQuest software (Duke University, EUA), and the detectability indexes, TTF, and NPS were calculated.

TABLE I – Detectability indexes measured using the custom-made Mercury phantom considering different diameters and combinations of protocols and reconstruction kernels. The resulting $CTDI_{vol}$ of each setup are also presented

Protocol	Tube voltage [kV]	kernel	filter	Detectability Index d'				$CTDI_{vol}$ [mGy]
				370 mm	300 mm	230 mm	185 mm	
CT Thorax	100	Lung	Al/Cu	6.4	8.4	9.6	10.0	2.7
	120		Al/Cu	4.9	7.1	9.6	12.8	2.7
	120		Al/Ag	5.7	6.6	7.3	9.1	1.9
	100	Body Sharp	Al/Cu	10.7	11.8	14.3	17.1	2.7
	120		Al/Cu	9.2	11.2	15.1	18.7	2.7
	120		Al/Ag	8.2	10.1	11.2	13.9	1.9
Thorax low dose	120	Lung	Al/Cu	4.9	5.5	6.2	6.2	0.5
	120		Al/Ag	3.3	4.1	5.2	5.8	0.4
	120	Body Sharp	Al/Cu	6.9	7.5	8.7	9.5	0.5
	120		Al/Ag	6.1	6.9	7.8	8.1	0.4

III. RESULTS

Fig. 3 presents examples of the influence of clinical protocols and beam filters on spatial resolution and image noise. Fig 4 demonstrates the impact of the reconstruction kernels on the spatial resolution and noise. Fig. 3 and Fig. 4 were obtained for the largest section of the phantom (370 mm). Table I presents the detectability indexes measured considering different phantom diameters and combinations of protocols and reconstruction kernels. The resulting $CTDI_{vol}$ of each setup are also presented

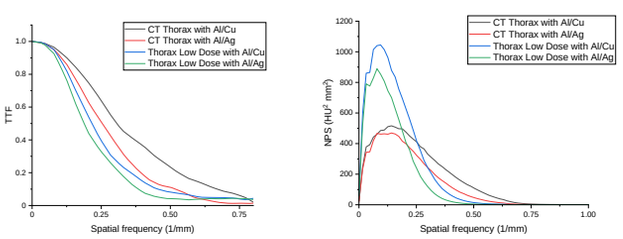


Fig. 3. TTF and NPS curves resulting from evaluations using Body Sharp kernel at 120 kV for the CT thorax and thorax low dose protocols with Al/Cu and Al/Ag filters.

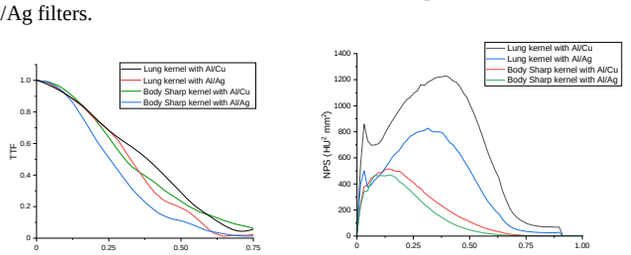


Fig. 4. TTF and NPS curves for the Body Sharp and Lung kernels with Al/Cu and Al/Ag filters for the CT Thorax protocol at 120 kVp.

IV. DISCUSSION AND CONCLUSIONS

The TTF curves in Fig. 3 showed a notable variation between the lung protocols. The TTF curves for the Thorax Low Dose protocols shows a sharper drop towards higher frequencies compared to the CT Thorax protocol. In both cases, the use of Al/Cu beam filters impact positively on the spatial resolution. Also, the NPS curves (Fig. 3) for the CT Thorax protocol present lower areas than Thorax Low Dose. Additionally, the peak of the NPS curve for Thorax Low Dose shows a shift to lower frequencies when compared to the CT Thorax protocol. The use of the Al/Ag beam filter combination decreases the area under the NPS curves, in addition to shifting the curves further to lower frequencies. In other words, as expected, the CT Thorax protocol generated less noise in the image than the Thorax Low Dose protocol. For the same exposure, the Al/Ag filter suppressed noise more than the Al/Cu filter. The impact of the reconstruction kernels (Fig. 4) on the spatial resolution showed to be small. However, the NPS curves with lung kernel demonstrated an increase in the area covered by NPS. On the other hand, body sharp kernels cause the region covered by NPS curves to decrease and show that noise at high frequencies is more efficiently suppressed. The lung kernel yielded images with a small increase on spatial resolution but more noise than a body sharp kernel.

The behaviors of the TTF and NPS curves explain the values of the detectability index presented in Table 1. The performance of the TTF and NPS curves is supported by the approximately 80% of reduction in $CTDI_{vol}$ values seen with the Thorax Low Dose protocol in comparison to CT Thorax. The detectability index increased with the use of the CT Thorax protocol, kernel body sharpness, tube potential, and $CTDI_{vol}$. Furthermore, the detectability index decreased with the size of the phantom section due to the increase in noise associated with the diameter of the phantom.

Comparing Thorax Low Dose protocol with the CT Thorax protocol considering each kernel/beam filters investigated, the reduction on the detectability index varied 29-42% for lung kernel and Al/Ag beam filter and 1-52% for lung kernel combined with Al/Cu beam filter. Considering the Body Sharp kernel the variations was 26-42% for Al/Ag beam filter and 25-49% for Al/Cu beam filter.

V. CONCLUSIONS

The implementation of lung screening programs at a broader scale using CT images introduces a paradigm into the imaging chain. Given the increased number of individuals undergoing these programs in larger populations, it is imperative to develop solutions for dose optimization, concurrently with sustainable image quality in order to keep elevated the detectability rates.

These challenges have been addressed through the incorporation of technological advancements, notably iterative reconstruction and the use of deep-learning based reconstruction algorithms.

The advent of these new technologies, however, does not allow for the continued use of the traditional metrics for evaluating image quality. In this context, the more general quantities investigated in the present work have demonstrated consistency with the prevailing reconstruction algorithms. Acknowledging this scenario, it is expected that the increasing routine use of these metrics in protocol optimization, and the incorporation of experimental methods to interpret their results will become an integral aspect of the regular tasks undertaken by medical physics experts in the near future.

Moreover, these metrics, particularly the d', must be correlated with the observer's detectability, especially when the clinical task is characterized by heightened sensitivity to size and contrast, as in lung cancer investigation. Consequently, the presented work focused on enhancing empirical techniques and on the application of mathematical formulation for evaluation of AI-based reconstruction techniques and protocol variations. Subsequently, future investigation will be undertaken to identify combinations that yield reduced doses while maintaining high observer's lung nodules detectability.

VI. 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). Special acknowledgements to Ioannis Sechopoulos, PhD, for his technical support for this work and comments on the manuscript.

References

[1] D. R. Aberle, C. D. Berg, W. C. Black, T. R. Church, R. M. Fagerstrom, B. Galen, I. F. Gareen, C. Gatsonis, J. Goldin, J. K. Gohagan, B. Hillman, C. Jaffe, B. S. Kramer, D. Lynch, P. M. Marcus, M. Schnall, D. C. Sullivan, D. Sullivan, and C. J. Zylak, "The National Lung Screening Trial: overview and study design," *Radiology*, vol. 258, no. 1, pp. 243-53, Jan, 2011.

[2] Y. Ru Zhao, X. Xie, H. J. de Koning, W. P. Mali, R. Vliegenthart, and M. Oudkerk, "NELSON lung cancer screening study," *Cancer Imaging*, vol. 11 Spec No A, no. 1a, pp. S79-84, Oct 3, 2011.

[3] "Reduced Lung-Cancer Mortality with Low-Dose Computed Tomographic Screening," *New England Journal of Medicine*, vol. 365, no. 5, pp. 395-409, 2011/08/04, 2011.

[4] H. J. de Koning, C. M. van der Aalst, P. A. de Jong, E. T. Scholten, K. Nackaerts, M. A. Heuvelmans, J.-W. J. Lammers, C. Weenink, U. Yousaf-Khan, N. Horeweg, S. van 't Westeinde, M. Prokop, W. P. Mali, F. A. A. Mohamed Hoesein, P. M. A. van Ooijen, J. G. J. V. Aerts, M. A. den Bakker, E. Thunnissen, J. Verschakelen, R. Vliegenthart, J. E. Walter, K. ten Haaf, H. J. M. Groen, and M. Oudkerk, "Reduced Lung-Cancer Mortality with Volume CT Screening in a Randomized Trial," *New England Journal of Medicine*, vol. 382, no. 6, pp. 503-513, 2020/02/06, 2020.

[5] A. Ba, Abbey C. K., Baek J., Han M., Bouwman R. W., Balta C., Brankov J., Massanes F., Gifford H. C., Hernandez-Giron I., Veldkamp W. J. H., Petrov D., Marshall N., Samuelson F. W., Zeng R. P., Solomon J. B., Samei E., Timberg P., Fornvik H., Reiser I., Yu L. F., Gong H., Bochud F. O., "Inter-laboratory comparison of channelized hotelling observer computation," *Medical Physics*, vol. 45, no. 7, pp. 3019-3030, 2018.

[6] H. H. Barrett, Yao J., Rolland J. P.,Myers K. J., "Model observers for assessment of image quality," *Proc Natl Acad Sci U S A*, vol. 90, no. 21, pp. 9758-65, 1993.

[7] P. Sharp, Barber D. C., Brown D. G., Burgess A. E., Metz C. E., Myers K. J., Taylor C. J., Wagner R. F., Brooks R., Hill C. R., Kuhl D. E., Smith M. A., Wells P., Worthington B., "Report 54," *Journal of the International Commission on Radiation Units and Measurements*, vol. os28, no. 1, pp. NP-NP, 1996.

[8] W. Peterson, T. Birdsall, and W. Fox, "The theory of signal detectability," *Transactions of the IRE professional group on information theory*, vol. 4, no. 4, pp. 171-212, 1954.

[9] J. G. Fletcher, D. L. Levin, A.-M. G. Sykes, R. M. Lindell, D. B. White, R. S. Kuzo, V. Suresh, L. Yu, S. Leng, D. R. Holmes, A. Inoue, M. P. Johnson, R. E. Carter, and C. H. McCollough, "Observer Performance for Detection of Pulmonary Nodules at Chest CT over a Large Range of Radiation Dose Levels," *Radiology*, vol. 297, no. 3, pp. 699-707, 2020/12/01, 2020.

[10] J. Solomon, Mileto A., Ramirez-Giraldo J. C., Samei E., "Diagnostic Performance of an Advanced Modeled Iterative Reconstruction Algorithm for Low-Contrast Detectability with a Third-Generation Dual-Source Multidetector CT Scanner: Potential for Radiation Dose Reduction in a Multireader Study," *Radiology*, vol. 275, no. 3, pp. 735-45, 2015.

[11] J. M. Wilson, Christianson O. I., Richard S., Samei E., "A methodology for image quality evaluation of advanced CT systems," *Med Phys*, vol. 40, no. 3, pp. 031908, 2013.

[12] P. Monnin, Marshall, N. W., Bosmans, H., Bochud, F. O., Verdun, F. R., "Image quality assessment in digital mammography: part II. NPWE as a validated alternative for contrast detail analysis," *Physics in Medicine & Biology*, vol. 56, no. 14, pp. 4221, 2011/06/23, 2011.

[13] E. Đedović, Faj D., Bokulić T., "Computed tomography protocols optimization using non-prewhitening model observer," *Journal of Physics: Conference Series*, vol. 1814, no. 1, pp. 012009, 2021.

[14] M. Ishida, K. Doi, L. N. Loo, C. E. Metz, and J. L. Lehr, "Digital image processing: effect on detectability of simulated low-contrast radiographic patterns," *Radiology*, vol. 150, no. 2, pp. 569-75, Feb, 1984.

[15] E. Abadi, Harrawood B., Sharma S., Kapadia A., Segars W. P., Samei E., "DukeSim: A Realistic, Rapid, and Scanner-Specific Simulation Framework in Computed Tomography," *IEEE Trans Med Imaging*, vol. 38, no. 6, pp. 1457-1465, 2019.

[16] F. Ria, Solomon J. B., Wilson J. M., Samei E., "Technical Note: Validation of TG 233 phantom methodology to characterize noise and dose in patient CT data," *Med Phys*, vol. 47, no. 4, pp. 1633-1639, 2020.

[17] D. Cester, M. Eberhard, H. Alkadhi, and A. Euler, "Virtual monoenergetic images from dual-energy CT: systematic assessment of task-based image quality performance," *Quantitative Imaging in Medicine and Surgery*, vol. 12, no. 1, pp. 726-741, 2021.

[18] H. Choi, W. Chang, J. H. Kim, C. Ahn, H. Lee, H. Y. Kim, J. Cho, Y. J. Lee, and Y. H. Kim, "Dose reduction potential of vendor-agnostic deep learning model in comparison with deep learning–based image reconstruction algorithm on CT: a phantom study," *European Radiology*, vol. 32, no. 2, pp. 1247-1255, 2022/02/01, 2022.

[19] J. Greffier, S. Si-Mohamed, J. Frandon, M. Loisy, F. de Oliveira, J. P. Beregi, and D. Dabli, "Impact of an artificial intelligence deep-learning reconstruction algorithm for CT on image quality and potential dose reduction: A phantom study," *Med Phys*, vol. 49, no. 8, pp. 5052-5063, Aug, 2022.

[20] M. P. Eckstein, J. L. Bartroff, C. K. Abbey, J. S. Whiting, and F. O. Bochud, "Automated computer evaluation and optimization of image compression of x-ray coronary angiograms for signal known exactly detection tasks," *Optics Express*, vol. 11, no. 5, pp. 460-475, 2003/03/10, 2003.

[21] R. S. Saunders Jr, and E. Samei, "Resolution and noise measurements of five CRT and LCD medical displays," *Medical Physics*, vol. 33, no. 2, pp. 308-319, 2006/02/01, 2006.

Dark-field X-ray Imaging of Anisotropic Sub-del Microstructures: Rod-shaped Air-bubbles

Xiangyang Tang and Yi Yang

***Abstract*—Dark-field (D-F) contrast is one of the major mechanisms that may make the grating interferometry-based X-ray imaging more capable in preclinical and clinical applications in comparison to conventional attenuation-based X-ray imaging. With resort to the Fresnel-Kirchhoff analysis under paraxial approximation, we derived the formulation that analytically characterizes the process of signal formation in the grating interferometry-based X-ray imaging. Based on our findings, we extended the definition of D-F contrast in this implementation of X-ray imaging from a real metrics to a complex one. In general, both the real and imaginary parts of the complex D-F contrast can be the foundation for grating interferometry-based X-ray Imaging. In this work, we explore the potentials of utilizing the real part of complex D-F contrast for X-ray imaging of sub-del anisotropic rod-shaped air-bubble towards preclinical and clinical applications. The preliminary data show that there should be a good chance for the D-F X-ray imaging, especially its concerting with novel and bio-safe contrast agents such as spherical and rod-shaped air-bubbles, to play significant roles in innovating or extending X-ray's capability–one of the most classical medical imaging modalities that save life or improve the quality of life worldwide–to meet the ever-increasing challenges imposed by advanced preclinical/clinical applications.**

***Index Terms*—X-ray dark-field contrast, complex dark-field contrast, X-ray dark-field imaging, grating interferometry-based X-ray imaging, X-ray phase-sensitive imaging.**

I. Introduction

The research and development in X-ray dark-field (D-F) imaging via grating interferometry is gaining momentum to establish its utility in clinical applications [1]. To solidify the foundation in physics and mathematics for the grating interferometry-based D-F imaging to fulfill its potentials in clinical applications and to explore its potentials towards more clinical and/or preclinical applications, we study the signal corresponding to the D-F contrast generated by the micro-structures at scales smaller than detector element's dimension (namely sub-del microstructures), especially in the case wherein the sub-del microstructures are anisotropic particles (namely anisotropic sub-del particles), in this work.

Manuscript submitted Jan. 23, 2024, revised May 1, 2024, and published Aug. 9, 2024.

X. Tang is 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).

In material and biological sciences, small angle scattering (SAS) has been recognized as the mechanism based on which X-ray can be utilized to study the property of microstructures that are not directly accessible due to the limitation in spatial resolution [2]. Diffraction-enhanced imaging is one of such methods in which, with resort to the radiation transfer theory, a line integral relation can be established between the second central moment of the sub-del microstructure's SAS angular distribution and its cross section [3]. Following their reports on grating interferometry's application for differential phase contrast imaging, Pfeiffer *et al* present their feasibility study of radiographic D-F imaging using X-ray as the source by heuristically connecting the D-F contrast, which is quantified as degradation in the visibility contrast, to sub-del microstructures' SAS [4]. The study of D-F computed tomography using neutron source is also reported [5], in which a squared quantity that characterizes the width of sub-del structure's SAS angular distribution is heuristically perceived as Radon transform of its SAS cross section. Later, it is analytically derived that the degradation in D-F contrast can be connected to the second central moment of sub-del microstructure's SAS angular distribution, which in turn can be connected to the SAS cross section through a line integral [6]. Also, it has been reported that such a fundamental and essential line integral can be established via Fourier analysis [7] and other approaches, based on scattering analysis [8], probability density function analysis [9] and linear diffusion [10], by assuming that the medium passed by X-ray is consisted of stratified layers. Notably, in all those approaches mentioned above, the D-F contrast is heuristically assumed independent of the phase contrast.

Through autocorrelation function analysis in which the paraxial approximation is made, Yashiro *et al* [11] connect the D-F contrast to the variance, correlation length and Hurst exponent that are the three major parameters characterizing the fluctuation in sub-del structure's SAS wave-front. They show that the D-F contrast is in fact not independent of the phase contrast, providing a more accurate characterization of the relation between the D-F contrast and the phase contrast. Though their derivation is more rigorous in terms of mathematical treatment, they still assume that is shared by all previous works–the sub-del microstructure's SAS is random (implying isotropic) and observes Gaussian distribution that is symmetric. In general, the assumption on the randomness and Gaussian distribution holds in practice but may be invalid in scenarios wherein the sub-del microstructure's SAS angular

distribution is anisotropic and/or asymmetric, as shown in the cases reported in [12–14]. Especially, of note are the cases wherein the sub-del microstructures exist in highly ordered layouts naturally or artificially.

To be more general, we relaxed the assumption that the angular distribution of sub-del microstructures is Gaussian and re-derived the analytic formulae of 0th- and 1st-order Fourier components that define the D-F contrast and phase contrast, via a mathematical treatment under the paraxial Fresnel-Kirchhoff theory [15]. We found that the sub-del microstructures may not only generate strong D-F contrast but also contribute to phase contrast. Then, we introduced the concept of complex D-F contrast to fully characterize the sub-del microstructures' effects on the D-F contrast and phase contrast. In this study, we review the concept of complex D-F contrast and report our recent work in exploring the potential of X-ray D-F imaging of sub-del microstructures towards preclinical and clinical applications.

II. Materials and Methods

The study's design of experiment is presented herein.

A. Characterization of dark-field contrast as complex metrics

With resort to paraxial Fresnel-Kirchhoff analysis, we went through a lengthy derivation in [15] and arrived at defining the D-F contrast as a complex quantity

$$B^c(m,n) = B_r^c(m,n) + iB_i^c(m,n) = -ln\big(V(m,n) + iV'(m,n)\big) \quad (1)$$

$$B_r^c(m,n) = B(m,n) = -ln\left(\sqrt{\big(V(m,n)\big)^2 + \big(V'(m,n)\big)^2}\right) = -\ln\left(\frac{a_1^s(m,n)/a_0^s(m,n)}{a_1^r(m,n)/a_0^r(m,n)}\right) \quad (2)$$

$$B_i^c(m,n) = B'(m,n) = \begin{cases} -arctan\left(\frac{V'(m,n)}{V(m,n)}\right) & if \quad V(m,n) > 0 \\ -arctan\left(\frac{V'(m,n)}{V(m,n)}\right) - sign(V'(m,n))\pi & if \quad V(m,n) < 0 \\ -\frac{\pi}{2} sign\big(V'(m,n)\big) & if\ V(m,n) = 0 \end{cases} \quad (3)$$

$$B_i^c(m,n) = d_c \partial_x \bar{\Phi}(x_m, y_n) + \phi_1^r(m,n) - \phi_1^s(m,n) \quad (4)$$

where (m,n) indexes the detector element in data acquisition. It was shown that the imaginary part vanishes and the complex quantity degenerates into a real one that is the same as that has been derived by others in the literature [4, 5, 7, 8], if the angular distribution of sub-del microstructures' SAS is random and Gaussian. Due to space limitation, the derivation is omitted here and the readers who are interested in the detail are referred to [15].

B. Modeling of grating-based phase-sensitive X-ray imaging

As has been evaluated and verified in our previous work [15], modeling of X-ray refraction-based imaging via Talbot interferometry is based on Fresnel-Kirchhoff analysis. In the simulation study, the detector cell dimension is set at 256 μm, and the period of gratings G_1 and G_2 are 4 and 2 μm. The X-ray source is assumed monochromatic at 35 or 80 keV, and the distance between the gratings G_1 and G_2 is 48.4 or 129.1 mm, corresponding to the respective 1st fractional Talbot distance. 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, and one of the gratings shifts 100 steps to acquire the data for each sampling.

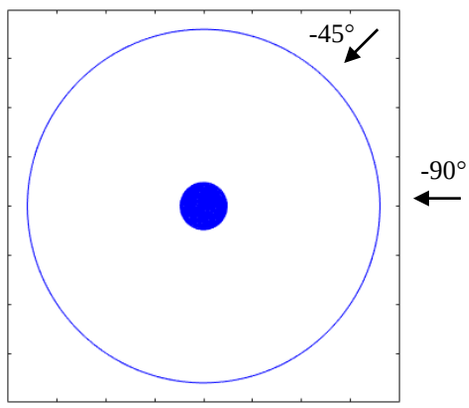


Fig. 1. Trans-axial view of the cylindrical numerical phantom (36.864 mm diameter) with a cylindrical target (2.5 mm diameter) consisting of micro-bubbles at the center and the azimuthal angles to show how the projection profiles are generated in the simulation study.

C. Phantom studies

A two-dimensional (2-D) phantom, which is a section of a water cylinder embedded with a target consisting of sub-del particles that are either spherical or rod-shaped air-bubbles, is designed for the simulation study to assess the accuracy of the real part of D-F contrast over the sub-del particle's refractive index, size, shape, distribution, and orientation. As shown in Fig. 1, the diameter of the water cylinder is 36.864 mm, while that of the central target is 2.5 mm. The refractive index of water is set at 1 - 1.8813×10^{-7} + i8.8402×10^{-11} for X-ray at 35 keV and 1 - 3.598×10^{-8} + i2.266×10^{-11} at 80 keV. At view angles -90°, -45° and 0°, the profile corresponding to the attenuation contrast and the real part of D-F contrast are generated and plotted for quantitative analysis. In the simulation study, the diameter of spherical air-bubbles are set at 2.5, 5.0 and 7.5 *u*m, respectively, while that of rod-shaped ones are 2.5×2.5, 2.5×5.0 and 2.5×7.5 μm^2, respectively. The location of each spherical or rod-shaped air-bubble is random, while the orientation of the rod-shaped air-bubbles (parallel to their long axis) is set as random or vertical, i.e., in parallel to the line connecting the 6–12 o'clock.

III. Results

The phantom study is carried out at 35 keV and 80 keV, respectively, but only the selected results corresponding to the former are presented due to space limitation.

A. Real D-F signal generated by spherical air-bubbles

In the case that the air-bubbles within the central target are isotropic spheres and distributed randomly in location, the projection of the 2-D phantom corresponding to the real D-F

signal and its variation over the view angles are plotted in Fig. 2. It is observed that the attenuation contrast is very weak and varies little over the view angle, while the D-F contrast is very strong and varies moderately over the view angles. Note that the isotropy in morphology implies the isotropy in orientation.

B. Real D-F signal generated by rad-shaped air-bubbles

In the case that the air-bubbles within the central target are anisotropic rods and distribute randomly in location and orientation, the projection of the 2-D phantom corresponding to the real D-F signal and its variation over the view angles are plotted in Fig. 3. It is observed that the attenuation contrast is still very weak and varies little over the view angle, while the D-F contrast is very strong and varies moderately over the view angles.

In the case that the air-bubbles within the central target are anisotropic rods and distribute randomly in location but consistently in orientation (long axis in parallel to the line connecting the 6 and 12 o'clock), the projection of the 2-D phantom corresponding to the real D-F signal and its variation over the view angles are plotted in Fig. 4. It is observed that the attenuation contrast is still very weak and varies little over the view angle, but it is important to note that the D-F contrast is very strong and varies radically over the view angles.

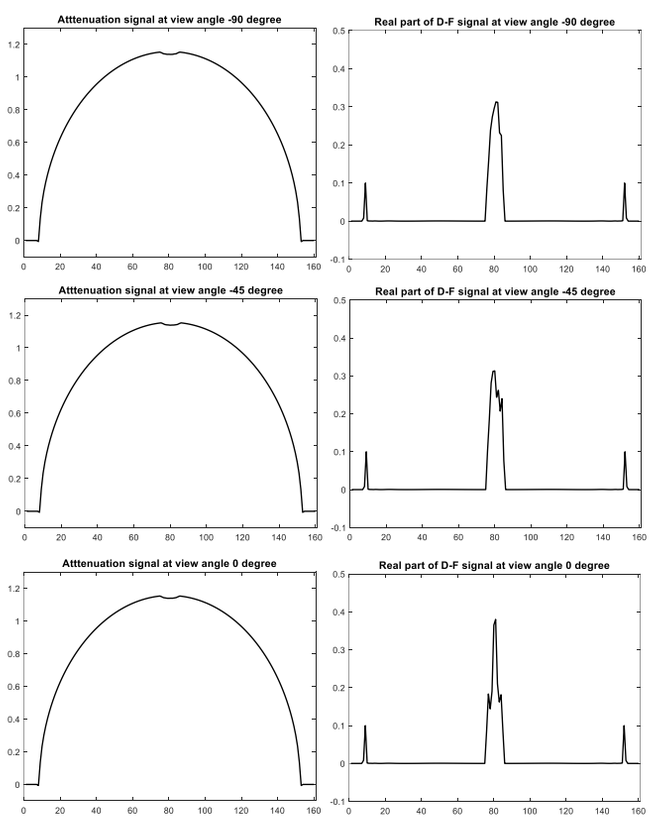


Fig. 2. The profiles of attenuation (left) and D-K (right) signal projected at view angles -90° (top), -45° (mid) and 0° (bottom) of the phantom with its central target consisting of isotropic spherical air-bubbles (7.5 μm diameter) distributed randomly in location.

IV. DISCUSSION

The complex definition of D-F contrast brings in challenges to its applications in tomographic imaging, as the foundation for the essential line integral relation is in principle undermined. The root cause underlying the fact that the D-F contrast has to be defined as a complex metrics is due to that X-ray's SAS is not independent of X-ray's refraction, as illustrated in our derivation based on the paraxial Fresnel-Kirchhoff theory in [11, 15].

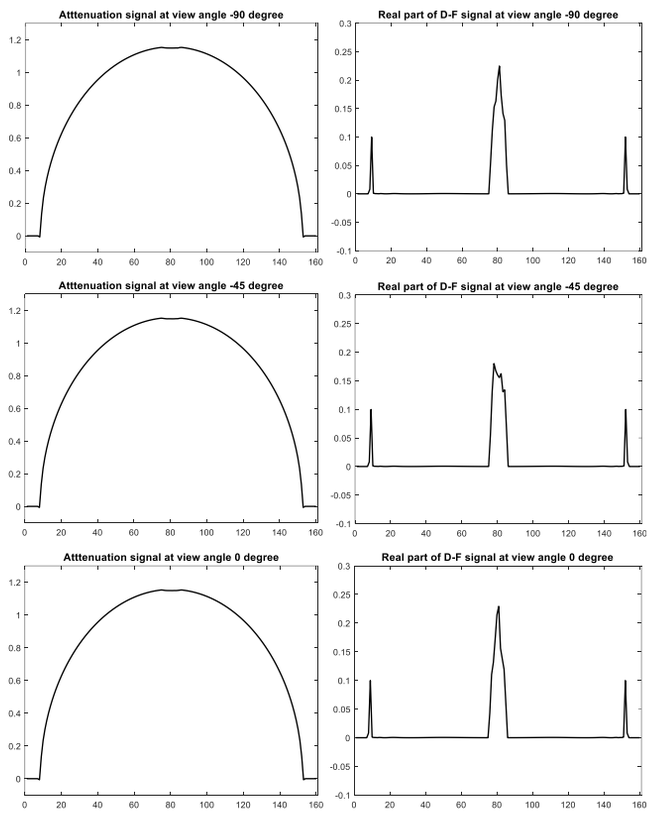


Fig. 3. The profiles of attenuation (left) and D-K (right) signal projected at angles -90° (top), -45° (mid) and 0° (bottom) of the phantom with its central target consisting of anisotropic rod-shaped air-bubbles (2.5×7.5 μm^2) distributed randomly in location and orientation.

In human perception, an observer discerns an interested target against its background. The subject contrast of an interested target is defined as a relative metric $C = (I_{target} - I_{background})/I_{background}$, where I_{target} and $I_{background}$ are the intensities corresponding to the target and background, respectively. One of the most effective ways to increase the subject contrast in medical imaging is the aid by contrast agent, as exemplified by the small molecular biomarker targeted radiolabels being used in SPECT/PET, and the blood compartment-based gadolinium contrast agents in MRI. The blood compartment-based iodinated contrast agent is essential to the success of conventional X-ray imaging modalities in clinics, as the iodine generates the largest attenuation at its *K*-edge (33.3*keV*) that is approximately equal to the effective energy at which the data are acquired. In analogue, there should be a material or structured materials, which can generate strong small-angle scattering, to be the contrast agent for X-ray D-F imaging. As demonstrated in Figs. 2–4, no matter if they are spheres or rods, the sub-del air-bubbles generate super strong D-F contrast and thus should be able to fit the role as contrast agent

for X-ray D-F imaging.

Filled with gases of low diffusivity, such as sulfur hexafluoride and perflurocarbons, and stabilized by shell structures made of lipid, protein or polymer, the half-life of microbubbles (air bubbles) in human and animal circulation systems can be up to ~10 minutes. The microbubbles have been approved by the FDA as ultrasonographic contrast agent for cardiovascular imaging (2014) and liver imaging (2016), respectively, making us optimistic about the air-bubbles' clinical safety profile while they are being adopted as the contrast agent for X-ray D-F imaging.

If the air-bubbles are anisotropic in morphology, e.g., the rod-shaped air-bubbles, and line up in their orientation but still distribute randomly in location (see Fig. 4), the D-F contrast generated by them is super strong and varies radically over the view angles. We are making use of this phenomenon to implement instantaneous digital subtraction angiography (iDSA) for diagnosis and therapeutic intervention of cerebrovascular and cardiovascular diseases.

We hope the preliminary data presented by us in this work may encourage the community to invest more effort on the research and development of D-F X-ray for preclinical and clinical applications. More investigations along this avenue are under way in our lab and we will promptly update the community with more information or findings soon.

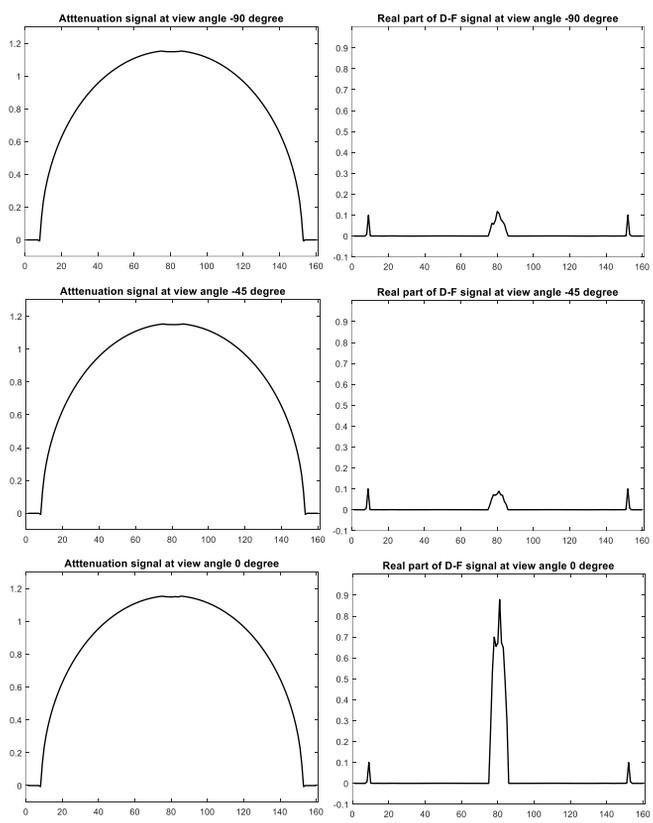


Fig. 4. The profiles of attenuation (left) and D-K (right) signal projected at view angles -90° (top), -45° (mid) and 0° (bottom) of the phantom with its central target consisting of anisotropic rod-shaped microbubbles (2.5×7.5 μm^2) distributed randomly in location but oriented vertically (in parallel to the line connecting the 6–12 o'clock (see Fig. 1)).

V. Conclusions

The preliminary data in our investigation show that no matter if they are spheres or rod-shaped, the anisotropic sub-del air-bubbles can generate super strong D-F contrast and thus they are candidate of being the contrast agent for X-ray D-F imaging. Moreover, the rod-shaped or any other air-bubbles that are anisotropic in morphology may be utilized as the contrast agent to implement the instantaneous digital subtraction angiography (iDSA) that can overcome the challenges existing in conventional DSA for diagnosis, therapeutic intervention, and therapeutic efficacy assessment in cerebrovascular and cardiovascular diseases.

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*, vol. 119, no. 8, e2118799119, Feb. 2022, 10.1073/pnas. 2118799119.

[2] O. Glatter and O. Kratky, eds., *Small Angle X-ray Scattering.* New York, NY, USA: Academic Press, 1982.

[3] M. N. Wernick, O. Wirjadi, D. Chapman, Z. Zhong, N. P. Galatsano, Y. Yang, J. G. Brankov, O. Oltulu, M. A. Anastasio and C. Muehleman, "Multi-image radiography," *Phys. Med. Biol.*, vol. 48, no. 23, pp. 3875–3895, Dec. 2003.

[4] F. Pfeiffer, M. Bech, O. Bunk, P. Kraft, E. F. Eikenberry, C. H. Broennimann, C. Greunzweig and C. David, "Hard-X-ray dark-field imaging using a grating interferometer," *Nat. Mater.*, vol. 7, no. 2, pp. 134–137, Feb. 2008.

[5] M. Strobl, C. Crituenzweig, A. Hilger, I. Manke, N. Kardjilov, C. David and F. Pfeiffer, "Neutron dark-field tomography," *Phys. Rev.* Lett., vol. 101, no. 12, 123902, 4 pages, Dec. 2008.

[6] G. Khelashvili, J. G. Brankov, D. Chapman, M. A. Anastasio, Y. Yang, Z. Zhong and M. N. Wernick, "A physical model of multiple-image radiography," *Phys. Med. Biol.*, vol. 51, no. 1, pp. 221–236, Jan. 2006.

[7] Z. Wang, K. Kang, Z Huang and Z Chen, "Quantitative grating-based x-ray dark-field computed tomography," *Appl. Phys.* Lett., vol. 95, 123902, 4 pages, Sept. 2009.

[8] G. Chen, N. Bevins, J. Zambelli and Z. Qi, "Small-angle scattering computed tomography (SAS-CT) using a Talbot-Lau interferometer and a rotating anode X-ray tube: Theory and experiments," *Optics Express*, vol. 18, no. 12, 12963, 11 pages, Jun. 2010.

[9] L. Rigon, A. Astolfo, F. Arfelli and R. Menk, "Generalized diffraction enhanced imaging: Application to tomography," *Euro. J. Ridio.*, vol. 68S, pp. S3–S7, Apr. 2008.

[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.*, vol. 55, no. 18, pp. 5529–5539, Aug. 2010.

[11] W. Yahiro, Y. Terui, K. Kawabata and A. Momose, "On the origin of visibility contrast in X-ray Talbot interferometry," *Optics Express*, vol. 18, no. 16, 16890, 12 pages, Jul. 2010.

[12] Y. I. Nesterets, "On the origins of decoherence and extinction contrast in phase-contrast imaging," *Opt. Commun.*, vol. 281, no. 4, pp. 533–542, Feb. 2008.

[13] S. K. Lynch, V. Pai, J. Auxier, A. F. Stein, E. E. Bennett, C. K. Kemble, X. Xiao, W. Lee, N. Y. Morgan and H. H. Wen, "Interpretation of dark-field contrast and particle-size selectivity in grating interferometers," *Appl. Opt.*, vol. 50, no. 22, pp. 4310–4319, Aug. 2011.

[14] P. Modregger, F. Scattarella, B. R. Pinzer, C. David, R. Bellotti and M. Stampanoni, "Imaging the ultrasmall-angle X-ray scattering distribution with grating interferometry," *Phys. Rev*. Lett., vol. 108, no. 4, 048101, 4 pages, Jan. 2012.

[15] Y. Yang and X. Tang, "Complex dark-field contrast and its retrieval in X-ray phase contrast imaging implemented with Talbot interferometry," *Med. Phys.*, vol. 41, no. 10, 101914, 19 pages, Oct. 2014.

Multiple Focal Spots for High-Resolution Photon-Counting CT

Grace J. Gang, Yiqun Q. Ma, Leening Liu, Peter Noël, J. Webster Stayman

***Abstract*—High-resolution CT has been investigated for a large number of clinical applications. Such systems have, to a large extent, been driven by detectors with small pixel sizes. However, the need for reduced source blur to maintain high spatial resolution directly conflicts with the need for higher fluence to maintain acceptable image noise. These trade-offs pose a fundamental limitation in imaging performance. We have previously proposed a multi-focal-spot imaging protocol and data processing to transcend such limitation. In this work, we realized a dual-focal-spot protocol on an experimental photon counting CT test bench. Data were acquired using a small and large focal spot to emulate a dual source system using a different focal spot on each source. Exposures were chosen to emulate fluence limitations according to the area of each focal spot, with the large focal spot able to produce 4X exposure as compared to the small. The dual-focal-spot protocol reconstructs projection data from both the small and large focal spots at a total fluence that matches that of the large focal spot. A model-based reconstruction technique that accounts for source blur was adopted. The bias-variance trade-offs were compared between the dual, large-only, and small-only protocols. Reconstructions that considered the source blur were also compared with those that assumed an ideal point source (for the single focal spot protocols). Reconstructions that modeled the source blur have improved spatial resolution compared to those that did not, albeit at increased noise. The dual protocol was found to outperform all other protocols, achieving high spatial resolution matching that of the small-only protocol at reduced noise. Findings from this work support the theoretical advantages of multi-focal-spot protocols and form an important step towards the translation of such protocols to high-resolution CT systems.**

I. Introduction

HIGH-RESOLUTION CT has found growing applications in medical imaging. A number of studies have been reported for imaging of the lung,[1], [2] bone,[3], [4] breast,[5], [6] and cardiovascular diseases[7], [8]. The increased availability of high-resolution CT has been driven, in a large part, by the production of scanners with finer detector pixels. This is particularly true for photon counting CT systems where small pixel sizes are desirable to limit count rates, direct conversion of x-rays permits minimal light spread/detector blur, and the elimination of readout noise allows improved signal-to-noise ratio when photon counts are low.

However, the detector is only one limiting factor in the overall spatial resolution capabilities of a system. Another major contributor to spatial resolution losses is the x-ray source blur as a result of finite focal spot size. The focal spot imparts a blur kernel to the overall system resolution equal to the shape of the focal spot (magnified to the image plane). The larger the focal spot, the more source blur. Unfortunately, focal spots can not be made arbitrarily small due to the excessive heat produced during x-rays generation at a clinically relevant fluence level. Using a larger focal spot spreads heat production over a larger area allowing higher x-ray fluence; however, this also increases focal spot blur. Various strategies are used to increase heat capacity and hence the maximum fluence including rotating anodes and anode targets made of materials with high melting points (e.g. tungsten). It is common for the anode to be angled, allowing an extended linear focal spot that, when oriented obliquely, has the appearance of a smaller focal spot. A large focal spot may nullify the resolution advantage of small detector pixel size.

Furthermore, high-resolution systems with smaller pixel sizes suffer from increased noise and thus requires greater fluence to match the noise levels in systems with larger pixel sizes. The number of photons detected per pixel is proportional to the area of the pixel. Thus, to maintain similar noise levels in images, x-ray fluence needs to be scaled according to the ratio of the pixel areas. While it is possible to increase the integration time to allow for more x-rays to be detected, this increases overall scan time and the potential for other spatial resolution degradation like patient motion. The conflicting goals of higher fluence and small focal spot present a fundamental challenge for high-resolution CT.

The authors have previously proposed a strategy [9] to overcome this challenge by combining information from multiple focal spots. For example, projection data from a small focal spot (higher resolution, low fluence and higher noise) can be jointly reconstructed with data from a large focal spot (lower resolution, higher fluence and lower noise) to obtain high resolution reconstructions at reduced noise. Collection of such multi-resolution data is technically feasible on dual-source CT systems where each x-ray source can operate at a different focal spot setting. We note that other focal spot structures and combinations are also possible to achieve this purpose.

In this work, we have sought to translate the theoretical

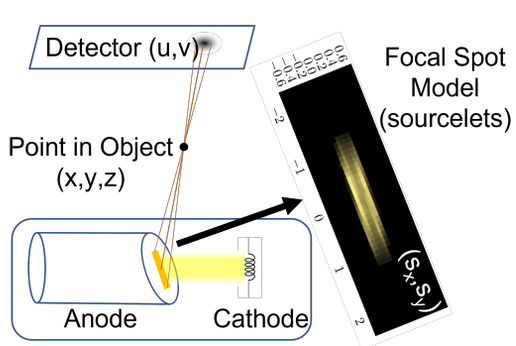


Fig. 1. Illustration of the focal spot model. A 2D distribution of "sourcelet" emitters on the anode are simulated. Each sourcelet has its own system geometry associated with its position on the anode and in 3D space. The projections from all sourcelets are weighted by their relative intensities and integrated to model both the depth-dependent blur and shift-variance associated with the apparent focal spot at different detector positions.

Grace J. Gang, Leening Liu, and Peter Noël are in the Department of Radiology at the University of Pennsylvania. Yiqun Q. Ma and J. Webster Stayman are in Biomedical Engineering at Johns Hopkins University, e-mail: web.stayman@jhu.edu



Fig. 2. Illustration of the x-ray test-bench used in these studies. The source-object distance was 685 mm and the object-detector distance was 340 mm. A 3D-printed resin target object with high-resolution details was used for performance investigations.

development and simulations reported in [9] to physical experiments. Specifically, we have used a prototype photon-counting CT test-bench with a dual-focal-spot x-ray source to demonstrate the high-resolution multi-focal-spot imaging. The following sections detail the methodology for these physical experiments, the data processing, and an evaluation of performance of the multi-focal-spot technique relative to single focal spot methods and variations in reconstruction techniques.

II. Methods

A. Focal Spot Modeling

In order to take advantage of the multi-resolution data from a multi-focal-spot data collection, one must model the varied information content in the measurements. For example, the model must quantify the relative spatial resolution losses in each set of projection data as well as the relative noise levels for effective combination of information from the two datasets.

Towards this end, we adopt the relatively sophisticated model of x-ray source blur from Tilley et al.[10] This model is illustrated in Figure 3 and includes a x-ray tube with an angled anode. The x-ray intensities emitted from the anode are modeled as a two-dimensional distribution on the anode. Each position on the anode is modeled as an individual "sourcelet" with its own system geometry. Projections from every sourcelet are formed, attenuated by the object, and then summed at the detector with a scaling based on the relative intensity of each sourcelet. Such a model will inherently incorporate both the intrinsic depth-dependence of source blur as well as the shift-variance associated with the changing apparent focal spot size for different regions of the detector. Mathematically, the mean measurement model is given by:

$$\bar{y}(\mu) = \mathbf{B}\exp\left(-\mathbf{A}\mu\right) = \left[s_1\mathbf{I}\cdots s_K\mathbf{I}\right]\exp\left(-\begin{bmatrix}\mathbf{A}_1\\ \vdots \\ \mathbf{A}_K\end{bmatrix}\mu\right) \tag{1}$$

where $\mathbf{A}_k$ denotes the projection matrix associated with the kth sourcelet, s_k is the relative intensity of that sourcelet, and μ represents the volume of attenuation coefficients. A compact form of this forward model can be described with the weighted sum matrix $\mathbf{B}$ and the stacked projection matrix $\mathbf{A}$.

To accommodate dual focal spot measurements, we adopt the above model for each set of projection data, stacking the measurement vectors such that:

$$\bar{y}_{\text{dual}}(\mu) = \begin{bmatrix}\bar{y}_1(\mu)\\ \bar{y}_2(\mu)\end{bmatrix} = \begin{bmatrix}\mathbf{B}_1 & \mathbf{0}\\ 0 & \mathbf{B}_2\end{bmatrix}\exp\left(-\begin{bmatrix}\mathbf{G}_1\\ \mathbf{G}_2\end{bmatrix}\mu\right) \tag{2}$$

$$\mathbf{G}_f^T = \begin{bmatrix}\mathbf{A}_{(f,1)}^T & \cdots & \mathbf{A}_{(f,K)}^T\end{bmatrix} \tag{3}$$

where $\mathbf{B}_f$ is the summing matrix for the fth focal spot and $\mathbf{G}_f$ contains the stack of projection matrices for the sourcelets in each spot. We may then solve following the nonlinear weighted least-squares problem to estimate the object:

$$\hat{\mu} = \arg\min_{\mu}\left[y-\bar{y}(\mu)\right]^{\mathrm{T}}\mathbf{W}\left[y-\bar{y}(\mu)\right] + \beta\mathbf{R}(\mu). \tag{4}$$

We solve the above iteratively using the separable paraboloidal surrogates (SPS) approach presented in Tilley et al.[10] We note that (4) may also be used to estimate images produced from single focal spot data with a blur model, and traditional no-source-blur models when a single sourcelet is used. All reconstructions in this work used 50 iterations of the SPS technique with 10 ordered-subsets and Nesterov acceleration[11]. Simple quadratic regularization on voxel differences in a 3D first-order neighborhood was used.

B. Physical Experiments

For the experiments in this work, the x-ray testbench and geometry in Figure 2 was used. The x-ray source was a Rad-94 x-ray tube (Varex, Salt Lake City, UT) that has two focal spots: nominally 0.4 mm and 0.8 mm with a tungsten anode tilted at 14°. An object stage (Physik Instrumente, Auburn, MA) includes a rotation stage for precison step-and-shoot acquisition. The detector is a XC-ACTAEON CdTe photon counter (Varex/Direct Conversion, Salt Lake City, UT) that is 51.2 mm by 25 mm with 100×100 μm pixels.

The imaging target (Figure 2) was a 3D-printed resin cylinder with a randomized interior with spherical voids. This object contains high-resolution details and is appropriately sized to avoid truncation with the photon counting detector. Projection data were acquired using each focal spot. An x-ray technique of 90 kVp and 0.15 mAs/frame was selected to avoid detector saturation. Small focal spot data had 360 frames over 360°; large focal spot data followed the same geometry with

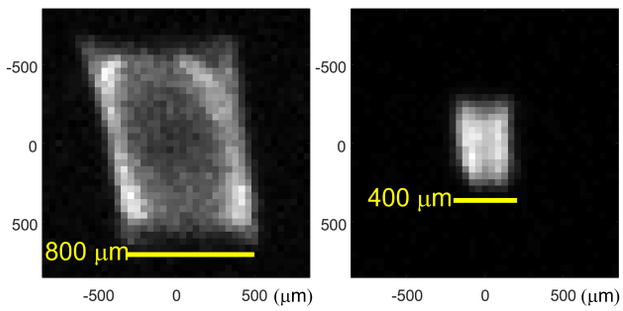


Fig. 3. Pinhole images showing the apparent intensity distribution at the center of the detector for the large (left) and small (right) focal spots. Nominal focal spot sizes provided.

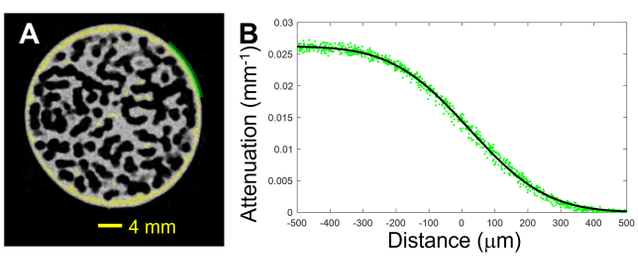


Fig. 4. Illustration of how spatial resolution and noise were quantified in the resin phantom. An edge spread function is estimated using samples along an arc - shown in green. Samples across the edge are fit with an error function to estimate the equivalent full-width half-maximum Gaussian blur in the reconstructed image. For noise estimation, a mask (shown in yellow) selecting voxels that are well within uniform regions is used to form an ensemble of voxels over which to compute standard deviation.

4 projections collected at each angle, giving a total of 1440 projections. This allows for the following emulated protocols:

1) *Small Focal Spot Data*: 1 projection/angle, emulating a limited maximum current level
2) *Large Focal Spot Data*: sum of 4 projections per angle, emulating a maximum current level based on the larger (approximately 4X area) focal spot
3) *Dual Focal Spot Data*: 1 projection/angle for the small focal spot and sum of 3 projections/angle for the large focal spot. The total fluence is matched with 2) using just the large focal spot)

C. Focal Spot Characterization

To construct the focal spot models for the above system, a 10 μm pinhole (Fluke Biomedical, Cleveland, OH) was placed approximately 30 cm from the source and imaged onto the photon counting detector using the same technique as the tomographic data collections above. The projection data was demagnified (by 2.4X) to obtain the apparent focal spot images as viewed from the detector center and sampled at the x-ray source (see Figure 3). These data were downsampled to reduce computation time and projected onto the 14° tilted anode to obtain the location and the intensity (s_k) of each sourcelet. A 6×6 sampling grid was used for each focal spot ($K = 36$).

D. Comparisons and Evaluation Metrics

Each of the three emulated protocols (Small, Large, Dual) were reconstructed using a $320 \times 320 \times 50$ volume with

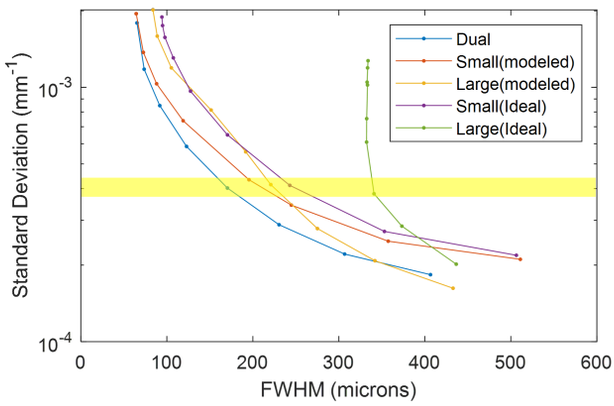


Fig. 5. FWHM spatial resolution versus noise standard deviation for the different acquisition protocols and reconstruction models. A band of approximately noise matched reconstructions (around 0.004 mm^{-1}) is highlighted in yellow. These reconstructions are shown in Figure 6.

100 μm cubic voxels. The Small and Large focal spot datasets were reconstructed using the estimator in (4) with both the sourcelets (blur) model as well as a, more typical, ideal (no blur) single point source model for comparison. The regularization parameter, β, was swept across a range of values to explore performance of each protocol and reconstruction model.

For evaluation, we approximated spatial resolution in reconstructed images using a $\sim 40^\circ$ sector at the edge of the object. Since the object is uniform resin with voids, a thin region at the edge of the phantom represents a step function between the nominal attenuation of resin (~ 0.026 mm^{-1}) and air (0 mm^{-1}). See Figure 4A for sample positions in the central slice (in green). Computing the distance to the edge forms an oversampled edge response (Figure 4B) which is fit using a parameterized error function to obtain a full-width, half-maximum resolution (FWHM) based on an assumption of Gaussian blur.

To compute noise in the reconstruction, a binary mask was formed by averaging slices over the central 2 mm of the phantom and thresholding values above 0.0254 mm^{-1} for the Large (ideal model) reconstruction with the largest regularization strength. This mask (shown in yellow in Figure 4) represents uniform resin regions of the phantom that are unlikely to be impacted by spatial resolution effects. The standard deviation of voxels within this mask was computed to quantify the level of noise in the reconstructed volume.

III. Results

Results of the performance investigation are shown in Figure 5. We make a number of observations. When the Large focal spot is used with an idealized reconstruction model, the spatial resolution is limited to around 330 μm. Additional decreases in β only result in additional reconstruction noise, not improved spatial resolution. A similar trend is evident for the Small focal spot; however, the limit is closer to 100 μm.

Applying the focal spot model in the reconstruction improves both the Large and Small focal spot protocols. Between these two protocols, the approach that yields lower noise depends on the limiting spatial resolution. The transition

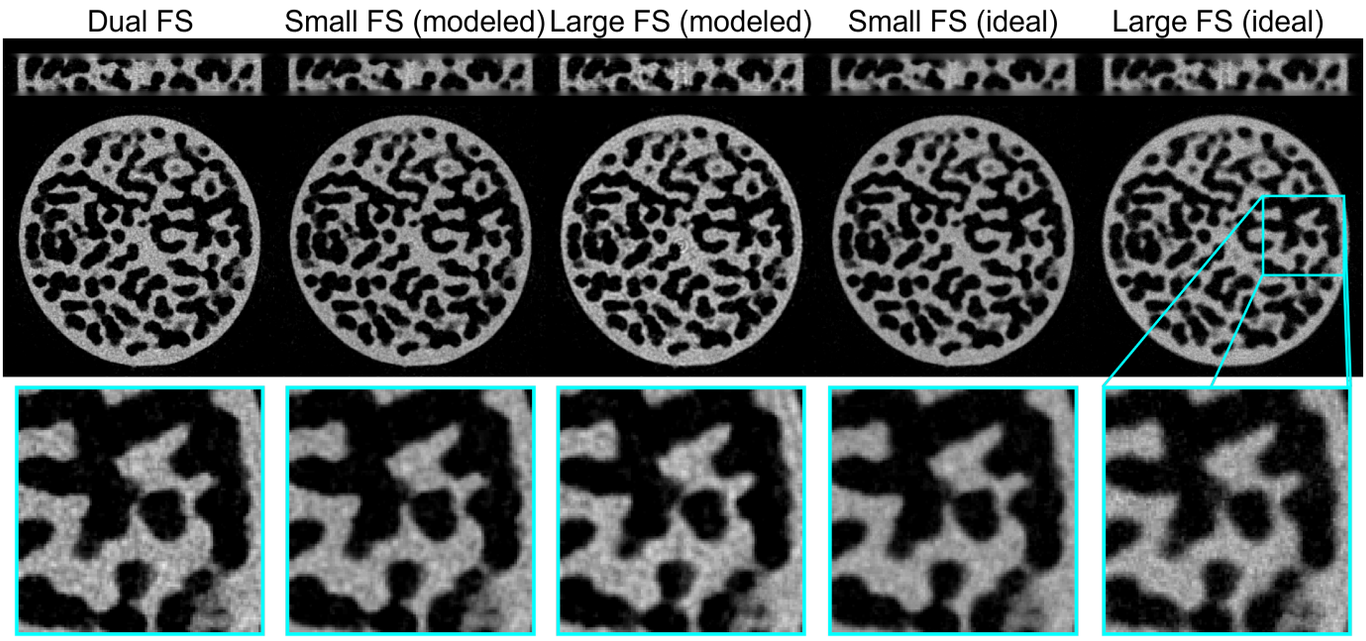


Fig. 6. Illustration of approximately noise-matched reconstruction across acquisition protocols and reconstruction methods. Methods are organized from left-to-right according to increasing coarseness in spatial resolution. Coronal, axial, and axial zoom images are shown. Window is 0.04 mm^{-1} and level is 0.02 mm^{-1}. Voxels are 100 μm cubes.

occurs around 250 μm, where above 250 μm the Large focal spot with modeled blur outperforms the Small focal spot with modeled blur. Below 250 μm, the relative performance of these two approaches is reversed.

We find that the Dual focal spot approach provides the best performance over almost the entire range, achieving high spatial resolution comparable to the Small focal spot protocol at reduced noise. Exceptions are at the lowest spatial resolutions (above 400 μm) and at the very finest spatial resolution (around 60 μm) where all techniques are resolution limited by the reconstruction voxel size.

Reconstructions for each approach at an approximately noise-matched (around a standard deviation of 0.004 mm^{-1}) level of regularization are shown in Figure 6. The approaches are organized according to decreasing spatial resolution from left to right. The spatial resolution differences are most apparent in a zoomed region of the central axial slice. The improved spatial resolution of the approaches that use source blur modeling is readily evident over the idealized (no blur) processing. Among modeled blur approaches, the Dual focal spot approach resolves the finest details.

One interesting feature of these reconstruction is the change in noise character. For example, even though the Dual and Small (modeled) reconstructions are approximately noise matched, the appearance of the noise is different. Specifically, the Dual approach has more high-frequency noise while the Small (modeled) reconstruction has lower frequency noise. This, perhaps, is logical since the Dual protocol may use the large focal spot data to reduce noise, but much of that noise reduction is likely to be at lower spatial frequencies. Thus, in noise-matched images, the Small (modeled) volume will have the appearance of lower frequency noise.

IV. Discussion and Conclusion

This work presents the potential of dual focal spot acquisition for improvements to the noise resolution trade-off in CT with particular application to high-resolution image using a photon-counting detector. Previous simulation studies described in Stayman et al.[9] have been implemented and demonstrated in physical experiments.

These studies are limited in some aspects, focusing on a relatively small target based on the small area photon-counting detector. This small area detector and small object make the depth-dependent and shift-variant effects relatively small. In future studies, we seek to extend this work to larger photon-counting systems where those effects are more significant. This work represents an important initial step in translating the dual focal spot technique through a demonstration of the improved image quality tradeoffs in physical experiments.

References

[1] A. Inoue *et al.*, *Investigative radiology*, vol. 57, no. 11, pp. 734–741, 2022.
[2] Y. Xu *et al.*, *International Journal of Chronic Obstructive Pulmonary Disease*, pp. 2283–2290, 2019.
[3] G. Shi *et al.*, in *Medical Imaging 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 11317. SPIE, 2020, pp. 356–362.
[4] K. Rajendran *et al.*, *Skeletal Radiology*, vol. 52, no. 1, pp. 23–29, 2023.
[5] N. Berger *et al.*, *Investigative radiology*, vol. 54, no. 7, pp. 409–418, 2019.
[6] W. A. Kalender *et al.*, *European radiology*, vol. 27, pp. 1081–1086, 2017.
[7] V. Mergen *et al.*, *Investigative Radiology*, vol. 57, no. 12, p. 780, 2022.
[8] S. A. Si-Mohamed *et al.*, *Radiology*, vol. 303, no. 2, pp. 303–313, 2022.
[9] J. W. Stayman *et al.*, in *Medical Imaging 2023: Physics of Medical Imaging*, L. Yu *et al.*, Eds., vol. 12463, International Society for Optics and Photonics. SPIE, 2023, p. 124630U. [Online]. Available: https://doi.org/10.1117/12.2654455
[10] S. Tilley *et al.*, *IEEE Transactions on Medical Imaging*, vol. 37, no. 4, pp. 988–999, 2018.
[11] A. S. Wang *et al.*, *Medical Physics*, vol. 42, no. 5, pp. 2699–2708, 2015. [Online]. Available: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.4914378

Hybrid Spectral CT: Combining rapid kVp-switching and dual-layer detectors for high sensitivity iodine imaging

Olivia F. Sandvold[1,2], Roland Proksa[1], Heiner Daerr[3], Kevin M. Brown[4], Thomas Koehler[3], Amy E. Perkins[4], Grace J. Gang[1], J. Webster Stayman[5], Ravindra M. Manjeshwar[4], Peter B. Noël[1]

***Abstract*—Over the past two decades, spectral computed tomography (CT) has undergone significant advancements, particularly in the realm of diagnostic accuracy, prompting a surge in clinical studies. This research examines the development of a new hybrid spectral CT system that combines a clinical-grade rapid kVp-switching X-ray tube with a dual-layer detector, aiming to boost quantitative spectral imaging performance in different clinical applications. The performance of the system was evaluated using varying tube voltages, duty cycles, and rotation times to enhance spectral outcomes. This experimental setup allowed for adjustments in parameters such as tube voltage pairs (140/80 and 120/70 kVp), duty cycles (ranging from 15/85 to 75/25), and tube currents (300 and 570 mA), which were tested on both large and small phantoms. The quantitative analysis was conducted using a two-input projection-based material decomposition approach. The study highlighted the impact of spectral weighting schemes on noise reduction and quantification bias in iodine density images. Results indicated that the maximum spectral separation scheme presented the least bias, suggesting its potential for improved clinical applications and outcomes. In conclusion, the research underscores the potential of integrating dual-energy technologies in hybrid spectral CT systems.**

***Index Terms*— Multi-energy CT, quantitative imaging, spectral CT, instrumentation**

I. Introduction

Over the last two decades, advancements in the implementation of spectral computed tomography (CT) have significantly enhanced diagnostic accuracy and spurred a variety of clinical studies. These studies underscore the benefits of quantitative spectral CT imaging across diverse fields, including cardiovascular, oncologic, and interventional procedures [1, 2]. This has propelled the development of specialized reconstruction and material decomposition approaches to minimize quantitative bias to the limits of underlying hardware constraints. Improved spectral performance has been theorized for novel CT systems equipped with more than one implementation for spectral acquisition [3]. These systems overcome challenges of individual technologies by mitigating tradeoffs in spatial and temporal resolution, spectral separation of datasets, and alignment of projection line integrals. Despite these proposed improvements, further optimization of system parameters and material decomposition, as well as the implementation with clinically compatible hardware, has yet to be fully investigated. Towards this end, we present quantitative results of an experimental hybrid spectral CT system built with clinical-grade rapid kVp-switching (kVp-s) x-ray tube and dual-layer detector.

Given the prominence of iodine contrast for diagnostic workup of perfusion cases as well as detection and differentiation of lesions, we illustrate spectral performance for the quantification of iodine at varying concentrations. Two phantom sizes are used to demonstrate application of hybrid spectral CT in both adult and pediatric patient populations. This research showcases the application of a weighting technique in the processing of multi-channel data in combination with established two-input (2D) projection-based material decomposition algorithms.

II. Methods

A. CT benchtop components and scan parameters

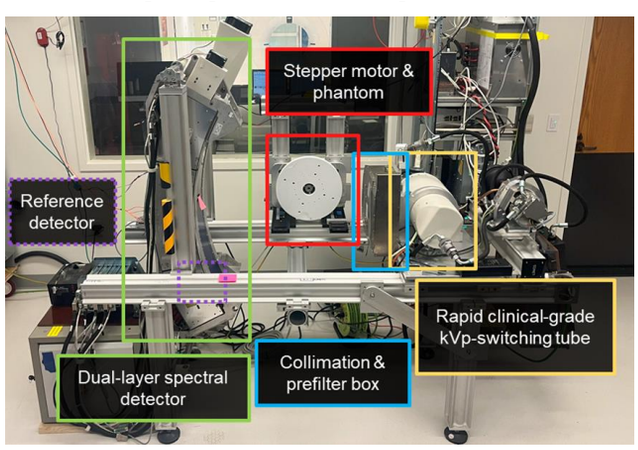


Fig. 1. Photograph of the experimental hybrid spectral CT system equipped with clinical-grade rapid kVp-switching x-ray tube, stepper motor and polyamide phantom, and clinical dual-layer energy-integrating detector.

The rapid kVp-switching x-ray tube (Philips Healthcare) of the hybrid spectral CT system was capable of alternating tube

This study was funded by the National Institutes of Health (R01EB030494) and supported by Philips Healthcare. O.F. Sandvold, R. Proska, G.J. Gang, and P.B. Noël are with the [1]Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. O.F. Sandvold is also with the [2]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA. H. Daerr and T. Koehler are with [3]Philips Innovative Technologies, Hamburg, Germany. A.E. Perkins, K.M. Brown, and R.M. Manjeshwar are with [4]Philips Healthcare, Cleveland, OH, USA. J.W. Stayman is with the [5]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, Maryland, USA. (e-mail: sandvold@upenn.edu).

voltages up to 50 kV difference with a minimum voltage of 60 kVp and a maximum voltage of 150 kVp. For the large phantom, switching tube voltages were set to 140/80 whereas the small phantom switching tube voltages used both 140/80 and 120/70 kVp pairs. The minimum cycle time (cT), containing both high and low voltage integration periods (IP), was approximately 200 µs with a minimum IP per kV of 100 µs. The duty cycle, which refers to the ratio of cycle time spent with the high kVp tube voltage compared to low kVp, was adjustable from 15/85 to 85/15. Three duty cycles, 15/85, 33/67, 75/25, were chosen for this experiment to investigate the impact of the total fraction of high kVp relative to low kVp in spectral results. Tube currents 300 and 570 mA were used as moderate dose and high dose acquisitions, respectively. The large phantom used both tube currents; only the moderate dose was used for the small phantom size. The system is equipped with a clinical dual-layer detector (Philips Healthcare). Incident beam collimation was set to a width of 16 slices (10 mm) and detector scatter was reduced using a 2D anti-scatter grid.

B. Phantom configuration

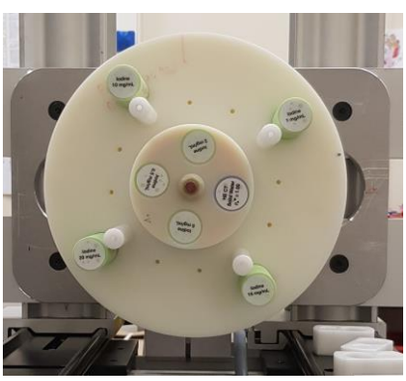


Fig. 2. Large phantom mounted on stepper motor containing smaller pediatric phantom and eight specified tissue-mimicking rods.

Two polyamide (PA6) cylindrical phantoms containing tissue-mimicking inserts (Sun Nuclear) were examined (**Fig. 2**). The larger 270 mm diameter phantom represented an average adult-sized patient and contained eight rods: iodine 0.5, 1.0, 2.0, 5.0, 10.0, 15.0, 20.0 mg/mL and solid water [1 g/mL]. The 100 mm diameter phantom approximated a pediatric patient circumference at the 95th percentile of 2- to 5-year-olds [4] and contained four inserts: iodine 0.5, 2.0, 5.0 mg/mL and solid water [1 g/mL]. The expected iodine concentration for each rod was based on the manufacturer-specified composition and tolerances after a spectral decomposition was applied.

Each phantom was mounted to a rotating stage controlled by a stepper motor operated using three total rotation times—333, 500, and 1000 ms. For all rotation times, the total number of projections per rotation was fixed at 2000. Due to IP limitations, the 15/85 duty cycle was only obtainable for 1000 ms rotation time, the 75/25 duty cycle was obtainable for 500 and 1000 ms rotation times, and the 33/67 duty cycle was obtainable for all rotation times.

C. Spectral decomposition methodology

For each scan, a total of four spectral channels were collected. The origin of each of these channels is depicted in **Fig. 3**. For practical reasons, we were interested in using a two-input projection-based material decomposition. Therefore, reduction of the four spectral channels into two input channels was required. kVp-s results were produced by combining the upper- and lower-layer scintillation signals for each kVp. Two additional combinations, "all 80/highE 140" and "max separation", were generated by the same summation process but selectively excluded some layer information (**Fig. 4**). Although these combinations are not as dose efficient as the kVp-s solution, they gain spectral separation as there is a reduction of overlapping photon energies captured in the channels included in the decomposition process.

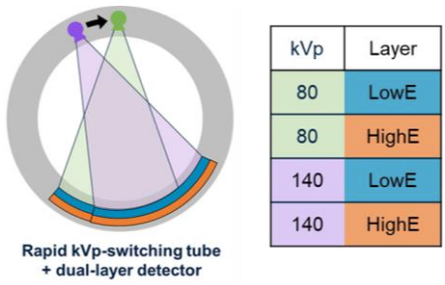


Fig. 3. Schematic of combination kVp-s and dual-layer CT acquisition with four defined spectral channels generated from kVp and layer pairs.

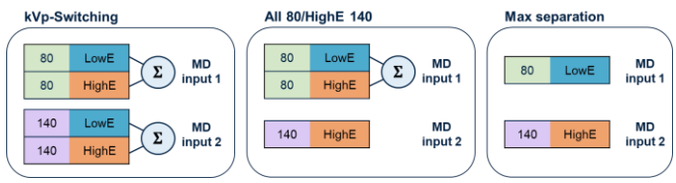


Fig. 4. Three different combination weighting schemes for production of two material decomposition (MD) inputs. Note the exclusion of some channel data in the "all 80/highE 140" and "max separation" schemes.

For each scan configuration of tube current, tube voltage pair, duty cycle, and rotation time, an air scan was acquired where a fast reference detector was used to estimate each kVp spectrum. The spectrum per kVp per layer was calculated by incorporation of the detector layer responses provided by the manufacturer at each photon energy. Line integrals were calculated for each channel, and then specific channels and spectra were combined by previously described weighting schemes to form two input spectra. Using these spectra, a 2D material decomposition was performed. Photoelectric effect, scatter, iodine density, and water basis images were created. No additional denoising or scattered radiation correction were used. For the large phantom, a field-of-view (FOV) of 300 mm was used; 150 mm FOV was used for the small phantom.

D. Quantitative image analysis

Circular regions of interest (ROI) with a physical diameter of 12.5 mm, corresponding to half the insert diameter, were placed inside all inserts as well as in one phantom background location (**Fig. 5**). The mean and standard deviation (noise) of each insert ROI was measured in each image slice. The average mean insert value was calculated by taking the average over 16 slices, and the average noise for each insert was computed as the average standard deviation over the 16 slices. Bias (error) in the iodine density images was calculated for each rod as the

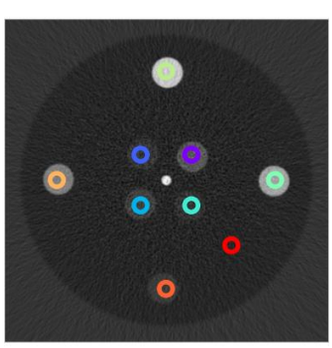


Fig. 5. Large phantom iodine density image with ROI for each insert and in PA6 background. W/L 40/10 mg/mL.

difference between measured ROI value in each slice and the expected iodine concentration in mg/mL. The standard deviation of bias was calculated over the bias values measured in each slice. For large phantom analysis, noise and bias were averaged over the 0.5, 1.0, 2.0, 10.0 mg/mL iodine and solid water inserts. In the small phantom, all four inserts were analyzed.

III. Results

A. Rotation time and weighting scheme effect on noise

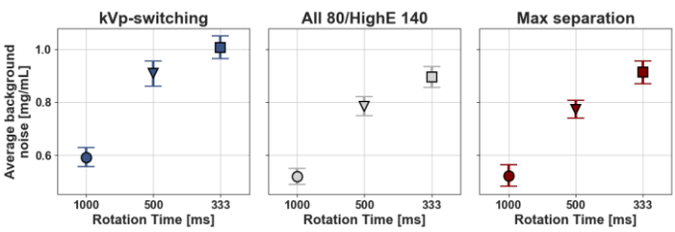


Fig. 6. Mean measured background ROI noise values for the three weighting schemes in varying rotation time acquisitions. Specifically, in the large phantom, with 140/80 kVp, 570 mA, and 33/67 duty cycle ratio iodine density images.

To examine the effect of exclusion of multi-channel data and rotation time on spectral reconstruction, the average noise in the background ROI was measured in the 140/80 kVp, large phantom 33/67 duty cycle ratio at 570 mA (**Fig. 6**). For the background noise, decreasing the rotation time from 1000 ms to 333 ms increased noise by 72% on average between the three weighting schemes.

TABLE I
AVERAGE NOISE [MG/ML] ACROSS ALL INSERTS FOR LARGE PHANTOM, 140/80 KVP, 570 MA, 33/67 DUTY CYCLE IODINE DENSITY IMAGES

	Rotation Time [ms]		
	1000	**500**	**333**
kVp-s	0.70 ± 0.11	0.99 ± 0.15	1.26 ± 0.21
All 80/ HighE 140	0.62 ± 0.10	0.88 ± 0.15	1.13 ± 0.20
Max separation	0.61 ± 0.10	0.88 ± 0.15	1.15 ± 0.19

Noise was additionally calculated across all iodine inserts for the same scan parameters (**Table I**). The maximum separation weighting scheme and all 80/highE 140 scheme show similar noise across the rotation times, with both containing less noise on average compared to kVp-s.

B. Bias with varying rotation time and spectral weighting

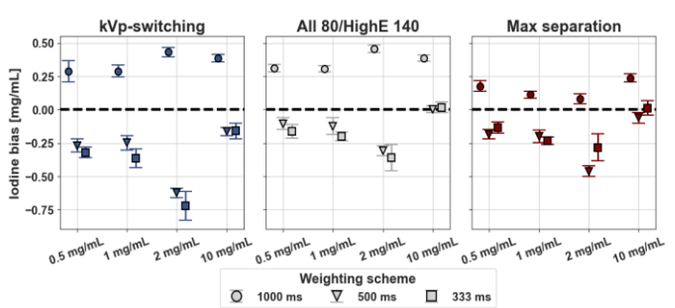


Fig. 7. Average bias in iodine density images for 140/80 kVp and dual-layer acquisitions comparing the effects of the inclusion of spectral channel data and rotation time for selected inserts. Maximum separation results in the smallest average bias in the 1000 ms and 333 ms rotation time acquisitions.

Fig. 7 demonstrates a clear relationship between spectral multi-channel inclusion scheme and measured iodine error. In the longest rotation time (1000 ms), the maximum separation scheme has the smallest average absolute bias over selected inserts of 0.15 ± 0.07 mg/mL compared to 0.37 ± 0.07 mg/mL for the all 80/highE 140 scheme and 0.35 ± 0.08 mg/mL error for kVp-s. Comparing the direction of bias, the 1000 ms rotation time acquisition led to overestimation of iodine concentrations compared to underestimation observed in 500 and 333 ms rotation time scans across the three weighting schemes.

TABLE II
ABSOLUTE BIAS MEAN AND STD [MG/ML] ACROSS ALL INSERTS FOR LARGE PHANTOM, 140/80 KVP, 570 MA, 33/67 DUTY CYCLE IODINE DENSITY IMAGES

	Rotation Time [ms]		
	1000	**500**	**333**
kVp-s	0.43 ± 0.16	0.50 ± 0.46	0.57 ± 0.51
All 80/ HighE 140	0.47 ± 0.15	0.32 ± 0.29	0.35 ± 0.30
Max separation	0.23 ± 0.17	0.39 ± 0.36	0.29 ± 0.23

Table II shows the average absolute bias in all measured iodine inserts for the large phantom at 570 mA, 33/67 duty cycle ratio using three rotation times. In the shortest rotation time, the maximum separation scheme contained the lowest average bias over all inserts compared to kVp-s and all 80/highE 140 schemes.

IV. Discussion

It is well understood that additional spectral diversity through acquisition of more than two spectral channels can both improve material decomposition accuracy and reduce noise [5, 6]. Producing a low bias in spectral CT is necessary for high sensitivity clinical scenarios where detection of small amounts of iodine is paramount for diagnosis of disease and treatment outcomes [7]. These types of vignettes include difficult to visualize pancreatic cancer lesions, patients with low tolerance to bolus volumes, and even pediatric patients whose small vascular structures and radiation exposure concerns can diminish diagnostic benefits of spectral CT [8-10]. Ultimately, this experimental hybrid spectral CT bench enables further investigation in high accuracy quantification of low contrast iodine leveraging implementation advantages of a rapid kVp-switching x-ray tube and dual-layer detector. In this abstract, we demonstrate several approaches to adapt multi-channel spectral data into a traditional 2D material decomposition pipeline in an adult-sized phantom. Specifically, we show that an average bias less than 0.5 mg/mL is obtainable at clinically feasible rotation times using only the maximally energy separated channels collected on the system.

The reduction in dimensionality of spectral channels plays a crucial role in the clinical translation of any multi-channel design. We have proposed alternative methods to optimize this dimensionality reduction challenge using the acquired spectral data [11]. Our work showcases the benefits of integrating various spectral and dual-energy technologies. The enhancements demonstrated in noise reduction and bias have

significant implications for photon-counting detectors [12]. This leads to the conclusion that combining technologies, such as rapid kVp-switching with photon-counting detectors, holds promise for improving future system performances.

V. Conclusion

The findings of this study demonstrate the feasibility of establishing a hybrid spectral CT system using clinical-grade components, and the translation of theoretically derived potential improvements into actual instrumentation. Further research is expected to yield greater insights into the clinical advantages of hybrid spectral CT systems.

References

[1] P. Rajiah, S. Abbara, and S. S. Halliburton, "Spectral detector CT for cardiovascular applications," *Diagn Interv Radiol*, vol. 23, no. 3, pp. 187-193, May-Jun 2017, doi: 10.5152/dir.2016.16255.

[2] D. Deniffel, A. Sauter, J. Dangelmaier, A. Fingerle, E. J. Rummeny, and D. Pfeiffer, "Differentiating intrapulmonary metastases from different primary tumors via quantitative dual-energy CT based iodine concentration and conventional CT attenuation," *Eur J Radiol*, vol. 111, pp. 6-13, Feb 2019, doi: 10.1016/j.ejrad.2018.12.015.

[3] M. Tivnan, W. Wang, G. J. Gang, E. Liapi, P. B. Noël, and J. W. Stayman, "Combining spectral CT acquisition methods for high-sensitivity material decomposition," in *Medical Imaging 2020: Physics of Medical Imaging*, 2020-03-16 2020: SPIE, doi: 10.1117/12.2550025.

[4] J. R. Fernández *et al.*, "Changes in pediatric waist circumference percentiles despite reported pediatric weight stabilization in the United States," (in eng), *Pediatr Obes*, vol. 12, no. 5, pp. 347-355, Oct 2017, doi: 10.1111/ijpo.12150.

[5] C. H. McCollough *et al.*, "Principles and applications of multienergy CT: Report of AAPM Task Group 291," *Medical Physics*, vol. 47, no. 7, 2020-07-01 2020, doi: 10.1002/mp.14157.

[6] 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, pp. 4031-47, Aug 7 2008, doi: 10.1088/0031-9155/53/15/002.

[7] J. G. Zarzour *et al.*, "Quantitative iodine content threshold for discrimination of renal cell carcinomas using rapid kV-switching dual-energy CT," *Abdominal Radiology*, vol. 42, no. 3, pp. 727-734, 2017-03-01 2017, doi: 10.1007/s00261-016-0967-5.

[8] D. A. Mankoff, L. K. Dunnwald, S. C. Partridge, and J. M. Specht, "Blood flow-metabolism mismatch: good for the tumor, bad for the patient," (in eng), *Clin Cancer Res*, vol. 15, no. 17, pp. 5294-6, Sep 1 2009, doi: 10.1158/1078-0432.Ccr-09-1448.

[9] B. Quiney, A. Harris, P. McLaughlin, and S. Nicolaou, "Dual-energy CT increases reader confidence in the detection and diagnosis of hypoattenuating pancreatic lesions," *Abdom Imaging*, vol. 40, no. 4, pp. 859-64, Apr 2015, doi: 10.1007/s00261-014-0254-2.

[10] A. Z. Liu, A. J. Winant, L. K. Lu, V. Rameh, K. Byun, and E. Y. Lee, "Computed tomography and magnetic resonance imaging for pulmonary embolus evaluation in children: up-to-date review on practical imaging protocols," *Pediatric Radiology*, 2022-07-22 2022, doi: 10.1007/s00247-022-05451-2.

[11] O. F. Sandvold *et al.*, "A computationally efficient method for dimensionality reduction in multi-channel spectral CT," *medRxiv*, p. 2024.02.18.24302905, 2024, doi: 10.1101/2024.02.18.24302905.

[12] S. Wang *et al.*, "Fast kV switching for improved material decomposition with photon counting x-ray detectors," presented at the Medical Imaging 2022: Physics of Medical Imaging, 2022.

Multi-energy data acquisition using a stationary spatial spectral encoder (S3E)

Jingyan Xu and Frédéric Noo

***Abstract*—We introduce the stationary spatial spectral encoder (S3E), a simple hardware addition to energy integrating CT systems to enable multi-energy data acquisition. The S3E is made of a ring of spectral filters that can be affixed to the CT gantry and provides continuous modulation of the x-ray source spectrum. The S3E can emulate many existing multi-energy CT solutions; moreover, the expanded surface area of spectral materials enables new spectral CT methods. We discuss the design concept of the S3E, and use a few examples to illustrate its potential applications.**

***Index Terms*—multi-energy data acquisition, material decomposition (MD), spectral imaging, energy integrating detectors**

I. Introduction

Spectral CT acquisition on energy-integrating CT systems takes a variety of approaches, such as dual source [1], fast (or slow) kVp switching [2], [3], or dual-layer detectors [4]. These approaches often compromise on features such as cost, spectral separation, or complexity in data processing. The quest to capture more desirable features drives the continual effort for improving spectral CT.

We introduce the stationary spatial spectral encoder (S3E), a ring-shaped device made of spectrally diverse materials for spectral data acquisition. The S3E is stationary – not with respect to the x-ray source – but by itself as an outside field-of-view (FOV) object (Fig. 1). During data acquisition, as the x-ray source moves, different parts of the S3E are exposed to the source, producing spectrally diverse data.

Compared to existing solutions, the S3E is easy to implement and has minimal requirements on the CT systems. The S3E is separate from both the source and the detector; it can work either as a standalone device or in combination with existing spectral CT solutions. These advantages make the S3E a feasible option to enable spectral acquisition on a large number of CT systems.

Next we present the design concept of S3E and provide a few examples of how S3E can emulate existing spectral CT solutions. Our discussion also uncovers new spectral data acquisitions that allow projection domain MD.

II. The stationary spatial spectral encoder – S3E

The S3E is a large ring located concentric to the iso-center of a CT scanner. Depending on a CT system's geometric parameters – namely, the source to iso-center distance (s2o), the detector to iso-center distance[1], and the size of the field-of-view (FOV) – there can be a few variations in the positioning of the S3E that affects its imaging properties.

If the detector to iso-center distance is smaller than the source distance, the S3E can be placed in the clearance between the source trajectory and the detector (Fig. 1). During data acquisition, while the source-detector rotates, the S3E remains stationary and provides spectral modulation. In this setup, the x-rays pass through the encoder only once before reaching the detector, making S3E a source side filter.

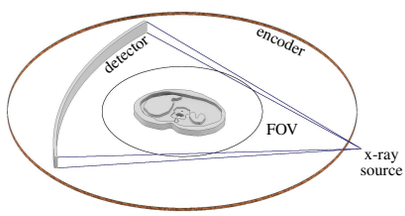


Fig. 1. The stationary spatial spectral encoder (S3E) consists of a ring of spectral elements that modulates the x-ray spectrum continuously. Shown here is the S3E-S variation, where the encoder is located behind the detector but in front of the source. The S3E-S variation works like a source-side filter. Two other variations are possible depending on the CT geometry.

There can be two other variations in the positioning of the S3E: (2) if s2o is much smaller than the detector distance, then the S3E can be placed in front of the detector but behind the source – making S3E act like a detector-side filter; (3) when there is not much difference between s2o and the detector distance, the S3E can be placed in front of both the source and the detector (but still outside the FOV), making it both a source-side and a detector-side filter. These design variations have different properties such as dose efficiency, spectral diversity, and ease of implementation.

In this work, we focus on the first variation (Fig. 1), which we refer to as S3E-S, where the last "S" is short for source-filter. In the following, we consider a few design choices of S3E-S and discuss their spectral imaging properties.

A. Spatial spectral filters without motion

The spatial spectral filter (SSF) [5] is a novel source-side filter proposed for multi-energy ($\geq$ 3) CT. During data acquisition, the SSF is translated with respect to the x-ray source, which by itself is a challenging engineering task.

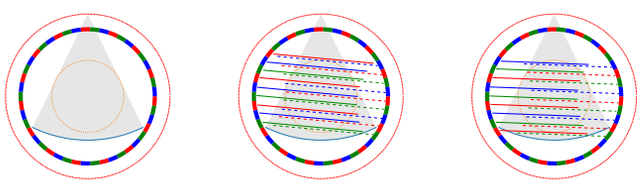
The stationary design of S3E provides a solution to the motion issue in SSF. By populating spectral elements around the S3E ring, we obtain the design in Fig. 2(a) that is capable of data acquisition similar to SSF. As the source rotates, the x-ray beam pass through different parts of the S3E, emulating a translation motion of the encoder without actual moving parts.

Email: jxu@jhmi.edu; frederic.noo@hsc.utah.edu. Work was supported in part by NIH R21 EB034337.

[1]For reference, the two distances are 595/490.6 mm and 538/408.2 mm, for Siemens and GE scanners, respectively. Here the detector is closer to the iso-center than the source, making it possible to utilize the gap in between.

The stationary nature of S3E makes it easy to analyze its spatial-spectral sampling pattern. Below we use the design in Fig 2(a) with 360° fan-beam data acquisition as an example to describe this analysis. The same analysis will be applied to other designs later in the paper.

The sampling pattern is analyzed in a rebinned parallel beam geometry. We connect the center of the spectral blocks so as to form a set of parallel "spectral lines"[2], and repeat this for different view angles (Fig. 2). At each view angle, the two ends of the "spectral lines" may be of the same color (*i.e.*, the same kind of spectral elements), or different colors (different kinds of spectral elements). As the S3E-S works purely as a source side filter, each spectral line is modulated by one of the blocks only. If the spectral lines connect same-colored blocks (*i.e.*, r-r, b-b, g-g), then these lines are doubly measured with the same spectral profile; if the spectral lines connect different-colored blocks (*e.g.*, r-b, b-g, g-r), then these lines are doubly measured with different spectral profiles.



(a) Emulating SSF (b) view 1 (c) view 2

Fig. 2. (a) The S3E can be designed to acquire data similar to SSF without filter motion. Here three spectrally distinct filters, represented by r, g, b blocks, are arranged in the S3E-S variant. (b)-(c) Spatial spectral sampling pattern of (a) with 2π acquisition. Spectral lines connecting different-colored blocks are doubly measured with different spectra; spectral lines connecting the same-colored blocks are doubly measured with the same spectra. The doubly measured lines are distinguished by the solid or dashed line types.

At each (rebinned) view angle, all spectral lines can be grouped into two kinds: (a) about 1/3 of the lines are doubly measured with the same spectrum (*e.g.*, g-g in Fig. 2(b)), and (b) the remaining 2/3 of the lines are doubly measured with different spectra (*e.g.*, r-b in Fig. 2(b)). These two groups of lines are interlaced spatially to cover the entire FOV. At the next view angle, the same pattern repeats but with different spectral information. For example, group (a) is now measured with spectrum b-b, and group (b) is now measured with r-g.

We may consider feasibility for material decomposition (MD) in either image domain or projection domain as they are more efficient than the one-step inversion. To apply image domain MD (iMD), we could potentially extract those lines that are measured with a single spectrum. These lines form a radially sparse and nonuniform set due to the presence of other spectral materials. Reconstruction of such irregular single spectrum data may be challenging for analytical algorithms. On the other hand, projection domain MD (pMD) can be applied to those rays that are doubly-measured with different spectra and obtain material density line integrals. These material density line integrals again form a radially nonuniform and sparse set that is challenging to work with.

[2]The "spectral lines" are not the measured line integrals. Assuming that the spectral blocks are larger than the detector elements, then each "spectral line" may correspond to a group of measured lines with the same spectral profile.

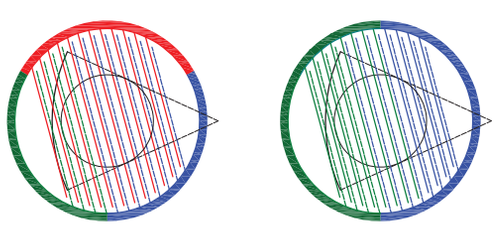
Our sampling pattern analysis shows that conventional pMD or iMD may face difficulty with the data in Fig. 2(a), but the one-step inversion or deep learning (DL) methods remain viable options. Promising one-step model-based MD results have indeed been reported for similarly challenging data [5].

In SSF, the spectral diversity is confined to the small area that the x-ray source can see at one view angle. Increasing spectral diversity comes at the price of compromised spatial sampling. The same compromise is also present in the fast or slow kVp switching approach to multi-energy CT. If we increase spectral diversity by utilizing more kVp settings, then each individual kVp data set will have fewer samples that makes MD challenging.

Unlike these existing solutions, the encoder ring in S3E provides an expanded surface area for spectral modulation. Spectral diversity can come from far-apart view angles to allow some decoupling between spatial and spectral acquisition, and to satisfy their individual requirements without necessarily trading off each other.

B. Alternative solution to kVp switching

Fig. 3 shows an alternative spectral data acquisition method that provides some decoupling between spatial and spectral sampling. They can be obtained by rearranging the spectral elements in Fig. 2(a) to form contiguous segments. For example, three spectral materials would form three spectral segments each spanning 120°. In general, N_{seg} spectral materials would form N segments each covering $360°/N_{\text{seg}}$.



(a) Three spectral blocks (b) Two spectral blocks

Fig. 3. New design of S3E-S by re-arranging spectral elements to form contiguous spectral segments. (a) With three segments, all doubly measured lines are measured by different spectra. (b) With two segments, some lines are doubly measured by the same spectral profile.

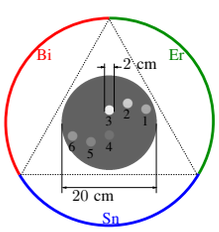
Similar to the SSF in Sec. II-A, we can analyze the spatial-spectral sampling pattern of Fig. 3 in a rebinned parallel beam geometry. With only two spectral segments each covering 180° (Fig. 3(b)), there always exist doubly measured rays from a single segment crossing the FOV. With a minimal of three spectral segments (Fig. 3(a)), those lines that cross a single segment twice will be too oblique and will be outside the FOV. Then all rays crossing a central region are measured twice with two spectra, making pMD a possibility. More distinct spectral segments can be added to further enlarge the FOV and improve spectral diversity without trading off spatial sampling.

It is often implicitly assumed that to enable pMD, there must be multiple single-spectrum data sets that each supports a full reconstruction. This is true for existing spectral CT solutions, *e.g.*, dual-layer detectors or photon counting CT. But the data acquisition in Fig. 3(a) clearly contradicts it. Each single-spectrum data set has an angular range of 120° and is not very useful; but the composition of all three enables pMD.

Fig. 3 also suggests an alternative to fast (or slow) kVp switching, in which the tube voltage is changed to a new setting every 120°. This data acquisition scheme, named SegMECT, was investigated in [6] using post (limited-angle) reconstruction iMD. Here we make the observation that SegMECT data, with at least three spectral settings, support pMD into *two* materials. Compared to the more frequent kVp switching, this alternative may reduce requirements on the source/generator, the detector, and spectral calibration. However, an advantage of fast kVp switching is the near-zero temporal delay in the spectral sinogram. Discussion related to the temporal aspect, for the S3E and the existing solutions, is deferred to Sec IV.

III. Numerical examples

We simulated spectral CT data from the S3E-S setup similar to Fig. 3 (redrawn in Fig. 4), where the three spectral materials are Bismuth of thickness 380 µm, Erbium of thickness 320 µm, Sn of thickness 600 µm. The analytic phantom, which is a modified version of the QRM dual energy CT phantom, consisted of a water cylinder of diameter 20 cm, and six cylindrical inserts of diameter 2 cm. Cylinders 1, 2, 3 were split into top and bottom halves with various concentration of Calcium and Iodine; cylinders 4, 5, 6 consisted of Iodine only.



Insert	concentration [mg/ml]
1 (top, bot)	Ca/20, I/2
2 (top, bot)	Ca/30, I/3
3 (top, bot)	Ca/40, I/4
4	I/0.5
5	I/1.0
6	I/1.5

Fig. 4. Our simulation setup (left) and the specification of the inserts (right). The spectral elements are Bismuth, Erbium, and Sn. Drawing is not to scale; the encoder covers $\geq$ 50 cm FOV, *i.e.*, larger than depicted. For data acquisition, the source starts at the 3 o'clock position and rotates clock-wise.

A. Data generation

The spectral CT data follows the polychromatic model

$$y_i = \int_{E_{\min}}^{E_{\max}} dE\, N_i(E)\, E\, e^{-L_i(E)} \qquad (1)$$

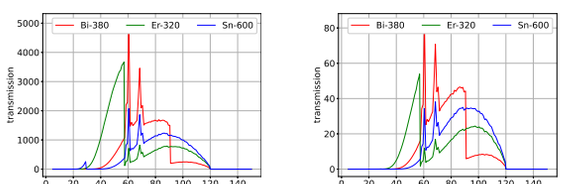
where i is the ray index, the incident spectrum $N_i(E) = T(E)B_i(E)S_i(E)$ models (1) the 120 kVp x-ray tube $T(E)$, (2) a body bowtie filter $B_i(E)$, and (3) the encoder spectrum $S_i(E)$ (see Fig. 5); and $L_i(E)$ is the line-integral of the linear attenuation coefficient at energy E.

The model (1) was discretized using an energy bin width ΔE of 0.5 keV, and $E_{\min} = 5$ keV, $E_{\max} = 150$ keV. We generated spectral CT data in the fanbeam geometry using Siemens CTSim package. Simulation parameters of interest are shown in Table I. They mostly follow the geometry of the Flash scanner, except that the detector was moved closer to the source (from 1085.6 mm to 950 mm) while maintaining the same FOV. This change led to a smaller detector pixel size. The modified parameters are also highlighted in Table I.

As a reference for the dose level, the airscan intensity, *i.e.*, y_i in (1) with $L_i \equiv 0$, was around 2.0e6 for all three

TABLE I
SIMULATION PARAMETERS. ALL SIZE/DISTANCE UNITS ARE MM. THE HIGHLIGHTED PARAMETERS DEVIATE FROM THE FLASH SCANNER.

s2o, source to iso-center distance	595.0
s2d, source to detector distance	950.0
R_e, radius of encoder	545.0
focal spot width, length	1.3, $1.5/\sin(7°)$
focal subsamples	3x , 2x
detector and trajectory subsamples	5x, 3x
fanbeam geometry	736 × 1152
detector width	0.067864°
fan angle	0.067864 × 736 = 49.947°
detector pixel size	rad (0.067864) × s2d = 1.125



(a) water - 0 cm (b) water - 20 cm

Fig. 5. Transmission spectrum $N_i(E)e^{-L_i(E)}$ (without energy weighting) through 0 cm water (a) and 20 cm water. The numbers following the element name in the label are the material thickness in µm. The percent transmission is 15.34%, 15.06%, 14.04%, for Bi, Er, Sn, respectively.

segments before Poisson noise perturbation was modeled in the transmission data.

We implemented pMD to obtain material line integrals. To do this, the airscan intensity, for each energy bin width of 0.5 keV, and the energy-integrated, noisy transmission data were rebinned to the parallel beam geometry to take advantage of the complementary rays. Afterwards, the rebinned airscan and transmission data were used in a maximum Poisson log-likelihood algorithm to perform pMD, assuming that the line integrals in (1) can be modeled as

$$L_i(E) = m_1(E)A_{1i} + m_2(E)A_{2i} \qquad (2)$$

where $m_k(E)$, $k = 1, 2$, are the linear attenuation coefficients of water and iodine, and A_{1i}, A_{2i} are their thicknesses. Basis material images of matrix size 512 ×512 were then obtained using FBP reconstruction of $\{A_{1i}\}$, $\{A_{2i}\}$, with a pixel size of 0.5 mm. These basis images were used to form virtual monoenergetic images (VMI) at different energy E.

B. Results

Our results for pMD from noise-free and noisy S3E projection and the reconstructed basis images are shown in Fig. 6. The iodine line-integrals in Fig. 6(a) show region-wise marked differences in noise properties. This regional difference is due to pMD from the pairwise combination of the spectral segments (cf. Fig 4) of Sn-Bi, Sn-Er, Bi-Er: the noise is the highest for the Bi-Er pair (smallest spectral separation), and the lowest for the Sn-Er pair (largest spectral separation).

The distinct noise properties in the line integrals also propagate to the reconstructed basis images. The noise streaks in the iodine and water images are oriented towards the Bismuth and Erbium segments (cf. Fig. 4), corresponding to the most noisy region in the basis line-integral.

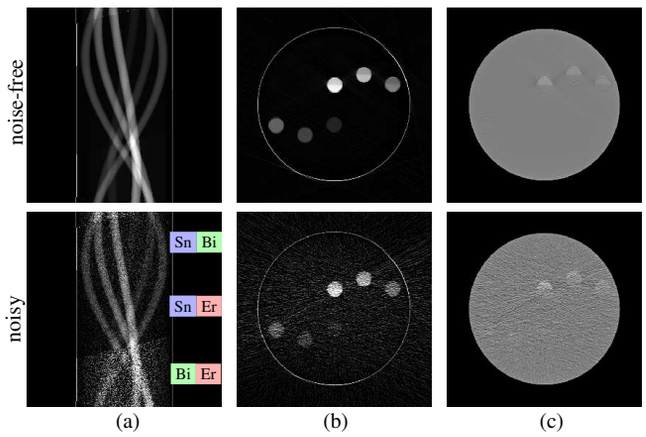


Fig. 6. Material decomposition. (a) iodine line-integral in parallel beam, (b) iodine map, [0-0.05, 4 + 0.05] mg/ml (c) water map, [1000-100, 1000 + 100] mg/ml. Water line-integrals are not shown as they are essentially that of a uniform 20 cm cylinder.

Mono-energetic images at energies = [60.0, 70.0, 80.0, 90.0] keV were synthesized from the noisy basis material maps of Fig. 6 and the results are shown in Fig. 7. These synthesized images exhibit less prominent streaks and noise due to the negatively correlated noise in the basis maps. Both the 60.0 and 90.0 keV images contain higher noise than the intermediate energies. The noise standard deviation in the central ROI (cyan) is 11.3, 3.6, 3.5, and 6.5 HU for the four energies, respectively. As for the signals, we observe that the contrast of the inserts decreases at higher energies.

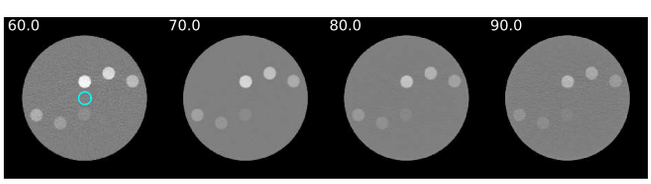


Fig. 7. Virtual mono-energy images. The noise (HU) in the ROI is 11.3, 3.6, 3.5, and 6.5, for the four energies, respectively. (WL, WH) = (-150, 150) HU.

IV. Discussion

A. Temporal delay

For applications affected by patient motion or fast contrast variation, an important consideration is the temporal offset between multi-energy scans. In Fig. 8 we plot the average time delay between coincident rays versus data acquisition for a few spectral CT solutions.

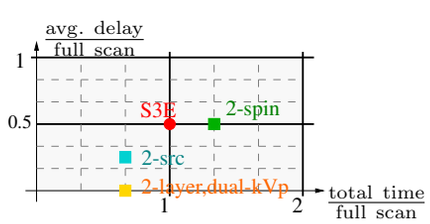


Fig. 8. The time delay in coincident rays versus total scan time.

Compared to the 2-spin approach, with two back-to-back short scans at two kVp settings, S3E has the same temporal delay and can achieve 30% saving in total acquisition time. S3E has a larger delay compared to the vendor-specific hardware solutions. These solutions, however, may involve compromises that are hidden in Fig. 8. For example, gantry rotation speed needs to be slowed down for dual-kVp switching while the S3E is free from such constraints. Depending on the applications, the simplicity of S3E may outweigh the temporal concerns, making S3E a viable alternative to spectral CT.

B. Other variations of S3E

This work has focused on S3E-S in 2D. As S3E-S behaves like a spatially expanded source-side filter, it can easily emulate 3D acquisition such as the TwinBeam. Alternative design choices may lead to new 3D multi-energy data acquisition and processing methods as well.

Between the two other variations of S3E, the 2nd variation, where the S3E is purely a detector-side filter, has spectral properties similar to S3E-S, although the detector-side filter may incur some dose penalty. The last variation, where the S3E is in front of both the source and the detector, has more interesting spectral properties. In the 2D plane, x-rays are blocked by the S3E twice; in the axial direction, most of the oblique x-rays are blocked by the S3E only once on the source side. Investigations on these variations are under way.

V. Conclusions

The stationary spatial spectral encoder (S3E) is a simple hardware addition to provide spectral CT data on energy-integrating CT systems. Delegating spectral acquisition to a separate device, *i.e.*, S3E, removes many constraints associated with modifying the source or the detector. The new freedom can lead to new spectral sampling patterns and data processing algorithms. As S3E is detached from both the source and the detector, it is complementary to existing spectral CT solutions. Combining them with S3E may improve dose balance, increase spectral diversity, and enable multi-material decomposition.

Our next step is to optimize S3E for different anatomy of interest (*e.g.*, head, chest, abdomen), different body sizes, and different basis materials. The ease of implementation of S3E makes it possible to select the right design from a collection of encoders to achieve task-specific data acquisition.

References

[1] M. Petersilka *et al.*, "Technical principles of dual source CT," *European journal of radiology*, vol. 68, no. 3, pp. 362–368, 2008.
[2] D. Xu *et al.*, "Dual energy CT via fast kVp switching spectrum estimation," in *Medical Imaging 2009: Physics of Medical Imaging*, vol. 7258, pp. 1198–1207, SPIE, 2009.
[3] K. Boedeker *et al.*, "Dual-energy: The Canon approach," in *Spectral Imaging: Dual-Energy, Multi-Energy and Photon-Counting CT*, pp. 63–72, Springer, 2022.
[4] R. Carmi *et al.*, "Material separation with dual-layer CT," in *IEEE Nuclear Science Symposium Conference Record, 2005*, vol. 4, pp. 1876–pp. 1878.
[5] M. Tivnan *et al.*, "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.
[6] L. Shen and Y. Xing, "Multienergy CT acquisition and reconstruction with a stepped tube potential scan," *Medical physics*, vol. 42, no. 1, pp. 282–296, 2015.

Three concepts to improve spatial resolution in next-generation CT scanners

Scott S. Hsieh

***Abstract*—Spatial resolution is a metric of fundamental importance in any imaging modality. Nonetheless, progress on in-plane spatial resolution was modest in diagnostic CT for the three decades between 1990 and 2020. The introduction of photon counting detector CT (PCD-CT) and a high-resolution energy-integrating detector CT (EID-CT) have re-energized interest in higher resolution systems. We discuss three new concepts to further advance the spatial resolution of CT platforms. (1) For energy-integrating detectors, reductions in pixel pitch decrease the fill factor: more space must be lost to interpixel reflective septa. We propose to load (or mix) a high-Z element into the interpixel septa, so that some x-rays incident on these septa have a chance to be fluoresce and be absorbed in an adjacent pixel. (2) For photon-counting detectors, detector resolution is excellent, but charge sharing reduces spectral separation. We previously proposed coincidence counting as a charge sharing mitigation technique. We observe now that these counters can be used as virtual pixels between two physical pixels to further improve sampling. (3) As detector resolutions increase, there is a need to improve focal spot characteristics. Small focal spots are thermally limited. Other groups have proposed complex source shapes, such as doublets, that improve high-frequency response. We conduct a thermal analysis and show that doublets improve total brilliance because cooling occurs in the rest period between the doublet. This strengthens the motivation for investigating complex source shapes. Simulations are provided for each concept to assess feasibility.**

***Index Terms*—spatial resolution, detector design, focal spots**

I. Introduction

SPATIAL resolution is a topic of fundamental importance in CT imaging. In other medical imaging modalities, spatial resolution has seen continuous improvement, but until recently, CT had plateaued with a typical detector pixel size of 1 mm (or 0.6 mm at isocenter), with some access to higher spatial frequencies through the use of oversampling. The modulation transfer function (MTF) of CT, a measure of resolution, can be modeled as a product of the source MTF and the detector MTF (disregarding for the moment other factors, such as apodization during the reconstruction and variation throughout the field of view). Loss of MTF at high frequencies from the source could be alleviated using small focal spots that compromise maximum brilliance. Loss of MTF from the detector could be recovered using comb filters (1), which reduce the penalty of the wide (~1 mm) rectangular aperture by increasing the width of the corresponding sinc pre-sampling MTF, although at great cost to dose efficiency.

Over the past five years, fresh progress has been made to improve spatial resolution. First, an energy-integrating detector (EID) system has appeared with small detector pixels (2). It is traditionally believed that smaller detector pixels cause a fill factor penalty and decrease dose efficiency, because each pixel is bordered by septa that channel visible light towards the photodiode. If the scintillator pixels decrease in size but the septa keep the same width, the fill factor must suffer. In keeping with this hypothesis, experiments suggest that this high-resolution EID suffers decreased dose efficiency (3), although engineering efforts in the detector may have reduced this penalty. Second, a photon counting detector (PCD) CT is now available that offers much better spatial resolution. The benefits of this resolution have been clearly demonstrated (4, 5).

In this work, we describe three new concepts to further improve spatial resolution in future CT scanners. The first concept provides a way of improving EIDs by reducing the effective fill factor. The second concept improves the already excellent spatial resolution of PCDs. The third concept will describe an improvement for the x-ray source.

Because this work is a bundle of three concepts, we have organized our work by concept. Within each concept, we provide separate subsections for methods and results.

II. Concept 1: X-ray Fluorescent Septa

Figure 1 provides a graphical summary of this concept.

A. Overview

The septa that separate scintillator pixels are typically composed of a white pigment such as titanium dioxide mixed into a binding material (6). The septa width is dictated by the amount of titanium dioxide to provide necessary reflectivity or by the width needed to provide mechanical strength. Therefore, the width cannot easily be decreased. We propose increasing the effective fill factor by loading (or mixing) a small amount of high-Z element into the binding material. The septa now would serve two functions: first, they would channel visible light from the scintillator towards the photodiode to reduce depth-dependent Swank noise (7); second, they would scatter incident x-rays through x-ray fluorescence. The emitted, fluorescent x-rays would be sent in a random direction and at

S. S. Hsieh (e-mail: Hsieh.Scott@mayo.edu) is with the Mayo Clinic, Rochester, MN 55905.

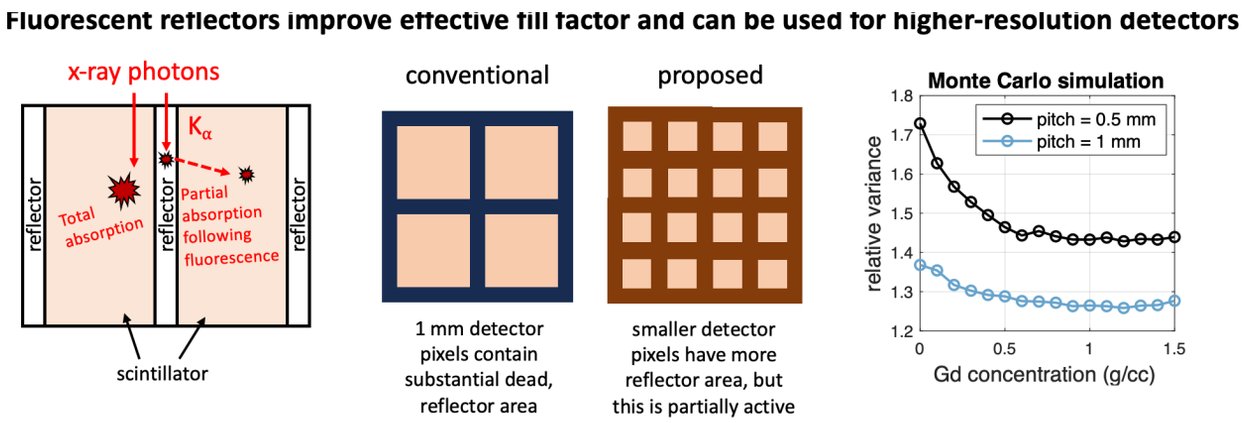


Figure 1. (Left) When a high-Z element is loaded into the reflector, incident x-rays can fluoresce and some of their energy can be restored. This is shown in the center ray. Only some of the energy is recovered. (Middle) Comparison of conventional 1 mm detector and proposed 0.5 mm detector with high-Z loading. Black areas are dead reflector areas, whereas dark brown signifies reflector areas that have a chance of recovering a photon. Although the fill factor is lower, these areas are partly useful, so the effective fill factor is improved. (Right) Monte Carlo simulations analyzing the relative zero-frequency variance at different concentrations of Gd, the selected high-Z element. For a pixel pitch of 0.5 mm, a septa width of 0.1 mm, and no anti-scatter grid, the variance decreases by 17% by adding Gd into the reflectors.

the energy of the characteristic x-ray, usually the K-alpha or K-beta line. Clearly, this second mechanism would be lossy, but it would be better than losing the photon entirely. The question is to quantify the size of this benefit.

B. Methods

We used GEANT4, a Monte Carlo simulation package, to assess the feasibility of our approach (8). We considered thin sheets of acrylic, which were a model of the reflective septa, with or without added high-Z material. We chose gadolinium as our high-Z material because it was a balance between having higher K-alpha emission (which is desired to keep more of the incident energy) with a lower K-edge (as photons below the K-edge cannot trigger fluorescence).

We assumed that the scintillator was gadolinium oxysulfide at nominal density (100% packing fraction) and 3 mm thick. We further assumed that any energy absorbed in the gadolinium oxysulfide would be fully collected without any loss, and we did not model the optical transport of the scintillated light.

The detector pixel was irradiated using a 120 kV spectrum filtered by 30 cm of water. The statistics of the energy absorbed in the scintillator were calculated over many noise realizations. Energy that was absorbed in the reflector region was disregarded. We did not differentiate by the location of absorbed region, so our results reflect zero-frequency quantum efficiency only.

C. Results

Figure 1 shows the variance as a function of gadolinium concentration, normalized to an ideal photon counting detector with a variance of 1.00. The septa width was 0.1 mm. At a pitch of 1.0 mm, the fill factor is 81% and the relative variance of a photon counting detector would be 1.23. In reality, it is higher because photons are weighted by their energy, introducing a form of Swank noise. As the concentration of Gd is increased, variance decreases because of the partial recovery of photons previously lost to the septa regions.

When the pixel pitch is reduced to 0.5 mm, the variance increases substantially, reflecting the fill factor penalty. This is largely restored by introducing Gd. The final variance, at 1 g/cc of Gd, is only about 6% higher than the original 1 mm pixel pitch. Note that adding an anti-scatter grid would diminish the value of this proposed concept, because some septa would be totally occluded and any loading of high-Z material underneath these septa would be useless.

III. Concept 2: Coincidence Counters for Sampling

A. Overview

The pixel pitch of PCD CT is typically 2 to 4 times smaller than energy integrating detectors, because there is no need for reflective septa and because this higher density of pixels ameliorates pulse pileup. On the other hand, very small pixels increase charge sharing. We previously proposed coincidence counting to mitigate this problem while retaining the ability to count quickly (9). If coincidence counts track the energy of the originating events, then further improvements in spectral performance may be obtained, particularly in K-edge imaging (10, 11).

Coincidence counts arise primarily from photons that arrive near a pixel border and which create a charge cloud that spans the two pixels. Therefore, coincidence counts can be considered to belong to a "virtual pixel" that straddles the boundary between two physical pixels. Figure 2 describes this idea.

B. Methods

We simulated a parallel beam scanner at an internal resolution that was 16 times better than the system resolution. We then gridded these counts into physical pixels or coincidence counters for three different systems: (1) an ideal PCD where all photons landing in the boundaries of a physical pixel are correctly assigned; (2) the same ideal PCD but with double sampling, such as what may be achieved with focal spot wobble; and (3) the same PCD, except that photons that are in the outer 12.5% of the pixel on each side are assumed to be placed into a coincidence counter.

Note that in a low-dose air scan (without pileup), the number of counts in the coincidence counter will be two times smaller

than the number of counts in the physical pixel after the coincidence counts on either side are subtracted. This creates an inefficiency in reconstruction, because the statistics are rapidly changing. To compensate, we used a frequency splitting technique (details omitted for brevity) where the low spatial frequencies are determined using a weighted sinogram, but the high spatial frequencies are determined using all pixels weighted by the air scan.

C. *Results*

Figure 2 shows an example reconstruction of a ring of cylinders phantom. The inner ring, surrounding the bright cylinder, is a model of the wires of a stent. The visualization of these wires is improved when using the coincidence counter.

IV. Concept 3: Brighter Source Doublets

A. *Overview*

With high resolution detectors, system MTF becomes limited by the source MTF. The source, in turn, is limited by thermal constraints: if the skin of the rotating anode track is brought too close to melting temperatures, it will deteriorate and the source will need to be replaced. In high-resolution applications where the focal spot is kept small, the primarily limitation is in the track itself, because the instantaneous power at a single point is very high. In the short time period (approximately 10 microseconds) in which a spot on the rotating anode is bombarded, temperatures can rise by more than 2000 degrees Celsius. On the other hand, when the focal spot is made large, the anode bulk may be the primary limitation, because power is not conducted outside the anode quickly enough. Some sources are designed for improved conduction of heat outside the bulk by the use of convection, making the use of large focal spots over long periods of time (or with shorter rest periods between scans) more practical (12). Thermal limits on the small focal spot, by contrast, are hard to overcome. Redesigns of the anode envelope are ineffective. The spot, after all, is only illuminated for a few microseconds. Any structure, even a hypothetical "magical" cooling element that is kept at absolute zero and which is situated only 1 mm away from the anode track, provides essentially no benefit to maximum power output because the heat from the focal spot cannot diffuse to the cooling element within a few microseconds. Faster rotation rates would be beneficial by reducing the time under bombardment, but further improvements are now mechanically challenging.

Recently, some groups have investigated focal spot shapes that can improve performance at high resolutions. Stayman et al. tested different focal spot configurations which come in the form of doublets or triplets rather than a single spot (13). Surprisingly, they showed that introducing a small gap inside the source spot (that is, splitting a single source spot into a doublet) improved performance, when evaluated for the goal of minimizing variance with an iterative reconstruction. Earlier, Pelc and Wang described a similar approach and evaluated reconstructed variance when using FBP (14). A more complete description would go beyond variance and analyze source MTF, and would show that by splitting a source spot into a doublet, certain high-frequency response can be accentuated at the cost of mid-frequency response.

We ask the question: how would a source doublet impact total source brilliance? Existing papers have assumed a power density constraint, assuming that the total power (W/mm^2) can be constrained. However, it is clear that real sources are not governed by a power density constraint. Doubling the source spot width does not double source brilliance, as would be expected of a power density constraint, but only increases it by the square root of two.

B. *The thermal constraint*

We will model the heat transfer in the tungsten anode using first principles only, acknowledging that a complete description requires finite element methods. This can include temperature-dependent changes in heat transfer (15). However, we argue that the one-dimensional heat equation provides a reasonably

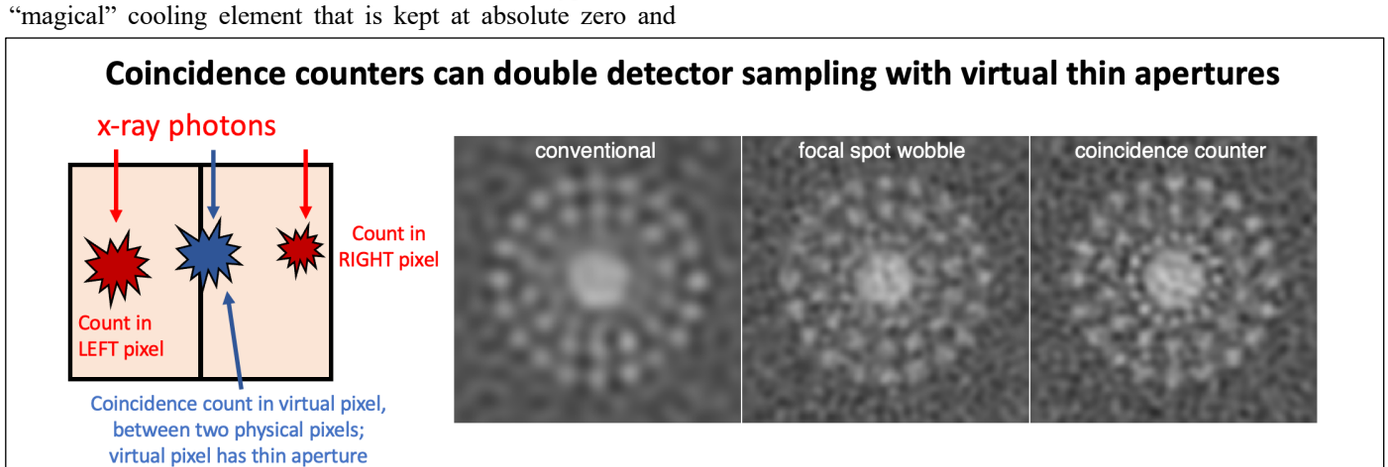


Figure 2. (Left) X-rays are incident on a photon counting detector with two pixels. The red photons deposit charge in one pixel only. The blue photon deposits charge in two pixels, and the detector can infer that it arrived near the pixel boundary, and can be counted accordingly. (Right) Simulations assuming a square pixel of a contrast phantom of many rings of circles. Without oversampling, the inner ring is hard to perceive. With detector oversampling (e.g., focal spot wobble), some ability to resolve the center wires is better, but the best images are with the proposed coincidence counter.

good model for the problem because of the time and spatial scales involved. The one-dimensional heat equation states that:

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2}$$

For tungsten, α is approximately 50 $\mu s^2/\mu m$.

Electrons originating from the cathode and accelerated by the voltage difference will undergo bremsstrahlung within the first five microns of tungsten (16), creating a thin skin of heat. For extremely short time scales, the distribution of energy within these five microns can be important, but we will approximate this as being a uniform distribution of energy. Assuming a track radius of 10 cm, a focal spot width of 0.5 mm, and a rotation rate of 100 revolutions per second, we estimate that a point on the anode is bombarded for only 10 microseconds. Even during this short time interval, the heat equation shows that heat actively migrate into the anode with a characteristic depth of less than 100 microns. Considering that the focal spot is 500 microns wide, and the skin of initial heat is 5 microns deep, and the characteristic migration distance <100 microns, we conclude that the heat migration is essentially in the depth direction only (that is, one dimensional) except at the outer periphery of the focal spot.

The heat equation is linear, and the fundamental solution (corresponding to an impulse) is:

$$T(x,t) = \frac{1}{\sqrt{4\pi\alpha t}} e^{-\left(\frac{x^2}{4\alpha t}\right)}$$

The thermal response of the anode can then be predicted by convolving the input distribution by the fundamental solution. To model the insulating boundary condition at the surface, we mirror the initial condition about the origin.

The thermal constraint states that all points within the volume should stay underneath the melting temperature of tungsten. We adjust the duration (focal spot width) of each of our simulations to reach the same peak temperature.

C. *Results*

We now calculate the time-varying temperature profile for three source profiles in Figure 3. The top model corresponds to a conventional focal spot which is bombarded for 10 microseconds. The other models split this profile into doublets and use a rest period equal to the bombardment period. In the middle model, this rest period is used to increase the spot width by 20%. In the bottom model, the intensity is increased by 8%. The temperature distribution at the end of the rest period is plotted for each (t = 6, or t = 5, respectively). In other words, by splitting the source into a doublet, the power density can be increased by 8%.

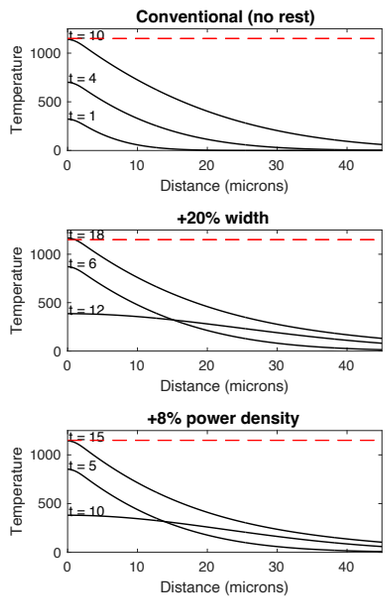


Figure 3. Temperature distributions throughout the anode at different time points (black lines). (Top) Conventional 10 μs bombardment. (Center) Two equal segments of 6 μs, separated by a rest of 6 μs. (Bottom) Two equal segments of 5 μs, separated by an equal rest duration, but 8% greater intensity.

REFERENCES

1. Meyer M, Haubenreisser H, Raupach R, Schmidt B, Lietzmann F, Leidecker C, et al. Initial results of a new generation dual source CT system using only an in-plane comb filter for ultra-high resolution temporal bone imaging. European radiology. 2015;25:178-85.
2. Oostveen LJ, Boedeker KL, Brink M, Prokop M, de Lange F, Sechopoulos I. Physical evaluation of an ultra-high-resolution CT scanner. European radiology. 2020;30:2552-60.
3. Hsieh J, Flohr T. Computed tomography recent history and future perspectives. Journal of Medical Imaging. 2021;8(5):052109-.
4. Rajendran K, Petersilka M, Henning A, Shanblatt ER, Schmidt B, Flohr TG, et al. First clinical photon-counting detector CT system: technical evaluation. Radiology. 2022;303(1):130-8.
5. Rajendran K, Voss BA, Zhou W, Tao S, DeLone DR, Lane JI, et al. Dose reduction for sinus and temporal bone imaging using photon-counting detector CT with an additional tin filter. Investigative radiology. 2020;55(2):91.
6. Shefer E, Altman A, Behling R, Goshen R, Gregorian L, Roterman Y, et al. State of the art of CT detectors and sources: a literature review. Current Radiology Reports. 2013;1(1):76-91.
7. Swank RK. Absorption and noise in x-ray phosphors. Journal of Applied Physics. 1973;44(9):4199-203.
8. Agostinelli S, Allison J, Amako Ka, Apostolakis J, Araujo H, Arce P, et al. GEANT4—a simulation toolkit. Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment. 2003;506(3):250-303.
9. Hsieh SS. Coincidence counters for charge sharing compensation in spectroscopic photon counting detectors. IEEE transactions on medical imaging. 2019;39(3):678-87.
10. Taguchi K. Multi-energy inter-pixel coincidence counters for charge sharing correction and compensation in photon counting detectors. Medical physics. 2020;47(5):2085-98.
11. Taguchi K, Iwanczyk JS. 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. 2021;48(9):4909-25.
12. Schardt P, Deuringer J, Freudenberger J, Hell E, Knüpfer W, Mattern D, et al. New x-ray tube performance in computed tomography by introducing the rotating envelope tube technology. Medical physics. 2004;31(9):2699-706.
13. Stayman JW, Noël PB, Gang GJ, editors. Multiple focal spots for high resolution CT. Medical Imaging 2023: Physics of Medical Imaging; 2023: SPIE.
14. Pelc NJ, Wang AS. Focal spot shape for computed tomography. Google Patents; 2022.
15. Hawkins T, Hyman S, Israel S. Thermal properties of solid and porous tungsten at temperatures to 5000 deg F. 1966.
16. Kandlakunta P, Thomas A, Tan Y, Khan R, Zhang T. Design and numerical simulations of W-diamond transmission target for distributed x-ray sources. Biomedical physics & engineering express. 2019;5(2):025030.

Single-Scan Quantitative Cone-Beam CT using a Primary Modulator and Dual-Layer Flat-Panel Detector: A Phantom Study

Linxi Shi, N. Robert Bennett, Adam Wang

***Abstract*—The use of flat-panel detector-based cone-beam CT (CBCT) in clinical settings has been increasing substantially in recent years due to its capability of providing low-dose, high resolution, compact, and low-cost solutions. The use of dual-energy imaging for CBCT further opens the window for quantitative imaging with material decomposition (MD). A dual-layer (DL) flat-panel detector can enable single-scan dual-energy (DE) CBCT by simultaneously acquiring motion-free DE images at each frame. However, the quantitative accuracy of MD is challenged by scatter contamination. Previously, we proposed a single-shot quantitative imaging (SSQI) method by combining the use of a primary modulator (PM) and dual-layer (DL) detector to enable DE imaging with scatter correction in a single frame. In this work, we use SSQI to achieve single-scan quantitative CBCT. Among numerous scatter correction methods, the primary modulator can simultaneously capture both primary and scatter signals without the need of a second scan, where alternating semitransparent and transparent regions of the PM encode the primary image into a known checkerboard pattern, while the low-frequency scatter image is largely unaffected. However, filtration from the PM is difficult to correct due to beam hardening effects. Fortunately, the DL detector provides DE images that can quantify both the PM's and object's material decomposition to remove the PM's checkerboard pattern from the projection images. We performed an experiment using a Gammex multi-energy phantom to demonstrate quantitative CBCT using the proposed method. The proposed method was able to perform quantitative estimation of the iodine inserts with less than 10% error.**

***Index Terms*— Cone-beam CT, dual-energy imaging, dual-layer detector, material decomposition, scatter correction, primary modulator**

I. Introduction

FLAT-PANEL based cone-beam computerized tomography (CBCT) has been widely used in image-guided procedures and diagnosis in the past two decades, such as in image-guided adaptive radiotherapy [1], C-arm assisted image-guided surgery [2], dedicated weight-bearing CT [3], and breast CT [4]. Substantial efforts have been made to improve the image quality and quantitative accuracy of CBCT. The adoption of dual-energy imaging is clinically promising due to its capability of generating material-specific images [5]. Typical implementations for DE imaging include kV switching, dual-source with different kV, and using a dual-layer detector [6]. While the first two methods can be tuned to have better spectral separation, their disadvantages are spatial and temporal misalignment between the DE images due to system rotation or patient motion. The dual-layer detector implicitly solves this problem by simultaneously capturing the DE images, enabling the acquisition of DE CBCT in a single scan with perfect spatial and temporal registration of DE projections. However, the scatter contamination for large-area imaging remains a fundamental challenge for DE imaging, which reduces the reliability of the MD results.

Numerous scatter correction methods have been developed [7][8], including software-based methods, such as analytical modeling, Monte Carlo simulation, and artificial intelligence, and hardware-based methods, such as using an anti-scatter grid, beam blocker, or primary modulator. Among all these methods, the primary modulator (PM) has the capability to capture both primary and scatter images in a single shot [9], thus avoiding an additional scan that has been usually implemented for other hardware-based scatter correction methods.

Previously, we have demonstrated the effectiveness of single-shot quantitative imaging (SSQI) for radiographic and fluoroscopic applications by combining the use of a PM and DL detector [10]. In this work, we will use the SSQI concept for single-scan quantitative DE CBCT. The quantitative accuracy of the proposed method will be demonstrated from a realistic measurement of the Gammex multi-energy phantom.

II. Methods

Single-scan quantitative DE CBCT requires the accurate estimation of primary images at each projection. To achieve this, SSQI leverages the complementary functions of a PM and DL detector, where the checkerboard pattern of the PM enables separation of the primary and scatter images, while the DL detector performs MD, and inherently corrects for beam hardening in the PM whose MD is known *a priori*. Thus, it

Research reported in this publication was supported by NIH R21EB030080 and Varex Imaging. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. (Corresponding author: Linxi Shi)

Linxi Shi, N. Robert Bennett, and Adam Wang are with the Department of Radiology, Stanford University, Stanford, CA 94305, USA (email: linxit@stanford.edu).

enables scatter-corrected DE imaging in each projection of the CBCT acquisition, which can then be reconstructed to form DE volumetric images.

A. Experimental Setup

Figure 1(a) shows the setup on our bench top system.

1) Primary Modulator

The PM (Figure 1(b)) includes alternating transparent and semi-transparent regions to form the checkerboard pattern for encoding primary images. In this work, a PM pitch size of 457 µm and a source-to-modulator distance (SMD) of 362 mm were selected for use. A detailed investigation on the impact of focal spot blur and PM pitch size can be found in Ref [11]. The semitransparent region in the PM is made of 210 µm copper, which reduces the intensity of a 120 kVp poly-energetic beam by about 50%. The PM implementation leads to the encoding of primary images into a checkerboard pattern whereas the scatter adds a low-frequency image without being modulated.

2) Dual-Layer Detector

The beam hardening in PM [12] leads to residual checkerboard patterns on the scatter corrected images, even after normalization. The dual-layer detector is used to solve this problem, where the modulated primary images and unmodulated scattered images are captured at the same time. By successfully removing the scatter, the residual PM pattern can be removed as the MD representation of the PM can be known beforehand. The dual-layer detector (Figure 1(c)) comprises two amorphous silicon flat-panel detectors, with a 200 µm CsI layer on the top detector to produce low-energy (LE) images and a 550 µm CsI layer on the bottom detector to produce high-energy (HE) images. Both layers have a native pixel size of 150×150 µm^2 that covers an active area of 43×43 cm^2. The combined use of a PM and DL detector allows simultaneous scatter correction and material decomposition in a single exposure.

3) DE CBCT Acquisition

A multi-energy CT phantom (Gammex, Middleton, WI) as shown in Fig. 2 was used for acquiring the DE CBCT scan with and without the PM. The source-to-detector distance (SDD) is 1412 mm, and the source-to-axis distance (SAD) is 1002 mm, resulting in an airgap of ~300 mm. Pulsed exposures with 120 kVp, 55 mA, and 20 ms were synchronized to the detector. The DE CBCT images were acquired at 7.5 frames per second with native resolution for 240 projections per rotation. In addition, a slit DE CT scan collimated to 2 cm on the detector surface was obtained as a scatter-free reference for later comparison.

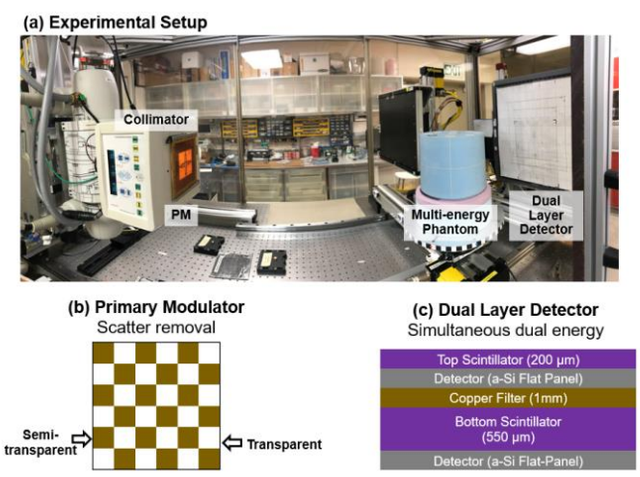


Fig. 1. Experimental setup of the proposed method.

B. Material Decomposition

We used an empirical material decomposition method in this work: A MD look-up table was created to map the DL-measured attenuation to the basis material thicknesses, i.e., f_a and f_c for acrylic and copper thickness, respectively [10]. In this work, we used acrylic/copper as the basis materials, which were selected for convenience but can be transformed into any other low/high Z material pairs as needed.

C. Scatter Correction

The SSQI algorithm is the foundation of this work, which fully utilizes the PM-encoded measurements from the DL detectors. The DL images obtained with the PM can be divided into 4 sub-measurements, which can be mathematically expressed as a combination of primary and scatter images as shown in Eqs. (1)-(4) [10].

$$M_{\text{top,st}} = \int I_{\text{eff,top}}(E) e^{-\mu_a(E) L_{\text{a,obj}} - \mu_c(E)(L_{\text{c,obj}} + L_{\text{c,PM}})} dE + S_{\text{top}}, \quad (1)$$

$$M_{\text{top,t}} = \int I_{\text{eff,top}}(E) e^{-\mu_a(E) L_{\text{a,obj}} - \mu_c(E) L_{\text{c,obj}}} dE + S_{\text{top}}, \quad (2)$$

$$M_{\text{bot,st}} = \int I_{\text{eff,bot}}(E) e^{-\mu_a(E) L_{\text{a,obj}} - \mu_c(E)(L_{\text{c,obj}} + L_{\text{c,PM}})} dE + S_{\text{bot}}, \quad (3)$$

$$M_{\text{bot,t}} = \int I_{\text{eff,bot}}(E) e^{-\mu_a(E) L_{\text{a,obj}} - \mu_c(E) L_{\text{c,obj}}} dE + S_{\text{bot}}, \quad (4)$$

where $I_{\text{eff,top/bot}}$ are the effective spectra of the DL layers; $\mu_a(E)$ and $\mu_c(E)$ are the attenuation coefficients of two basis materials: acrylic and copper, and their corresponding line integrals for the object are $L_{\text{a,obj}}$ and $L_{\text{c,obj}}$, respectively. The scatter images $S_{\text{top/bot}}$ are low frequency, and therefore modeled in each layer as an additive signal to the modulated images that remains constant in adjacent semitransparent (st) and transparent (t) regions of the PM checkerboard. The copper component of the PM, $L_{\text{c,PM}}$ is known *a priori* and can be measured empirically. Therefore, the MD images, $L_{\text{a,obj}}$ and $L_{\text{c,obj}}$, and scatter images, $S_{\text{top/bot}}$, can be directly solved. With proper calibration of the MD look-up tables, i.e, $f_{a/c}$, the problem can be further simplified to estimating $S_{\text{top/bot}}$ that makes $L_{\text{a,obj}}$ and $L_{\text{c,obj}}$ most consistent between semitransparent and transparent regions, thus achieving optimal MD in a single shot.

D. Single-Scan Quantitative DE CBCT

The SSQI was performed for each projection in the CBCT acquisition to achieve single-scan MD, and the resulting material-specific projections were reconstructed to material-specific images. Two iodine inserts with theoretical concentrations of 5 mg/ml and 10 mg/ml were used for quantitative validation. Since our basis materials were acrylic/copper, the basis material images were converted to



Fig. 2. Gammex multi-energy phantom for quantitative investigation of the proposed method. Two iodine inserts circled in red with 5 mg/ml and 10 mg/ml are used for validation.

iodine/water densities using their respective mass densities and attenuation curves.

The algorithm was implemented in MATLAB on a standard PC (11th Gen Intel(R) Core (TM) i7 @ 2.9GHz, 4 Core(s), 8 Logical Processors, and processing a single exposure set of DE images took about 2 sec without parallel computing or another runtime optimization.

III. Results

Figure 3 shows the material-specific images generated from three settings: slit scan, standard CBCT (without SSQI), and CBCT with SSQI. The slit scan reduced most scatter signals, so we considered the direct MD rendered accurate estimation for the two iodine inserts' concentration. A ~10% error was observed compared to theoretical values, which were 5 mg/ml and 10 mg/ml, respectively. This might be due to residual scatter, or the limited numbers of material pairs used in the calibration process. For CBCT without scatter correction, the estimation from direct MD is no longer reliable due to the scatter contamination, with errors of ~80% for the iodine densities. The MD accuracy was restored after implementing the proposed single-scan CBCT method using SSQI. While the mean concentration was preserved, we noted an increased noise level compared to that of the slit scan, which was caused by the scatter noise, even after successful removal of the deterministic scatter signals. The same observation was found for the water images.

Figure 4 compares the quantitative accuracy of the two iodine inserts from the three scans. It is clearly seen that the CBCT without SSQI or other scatter correction cannot provide reliable quantitative MD, while the slit scan and single-scan CBCT with SSQI are comparable to theoretical values.

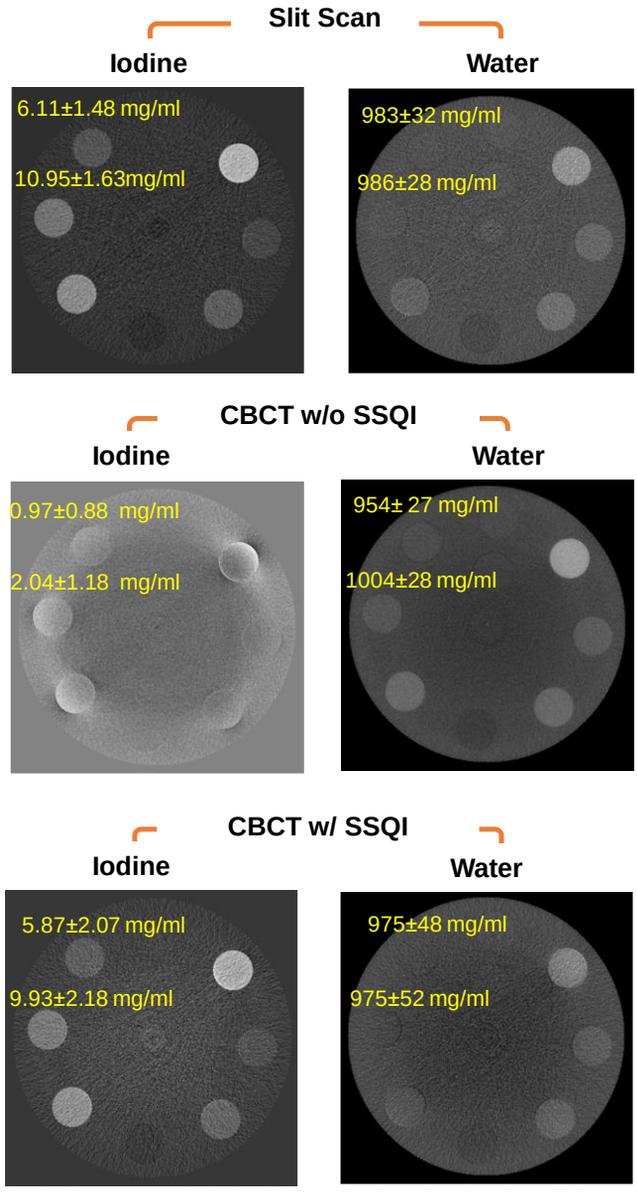


Fig. 3. Comparison of MD for slit scan, CBCT with and without SSQI scan.

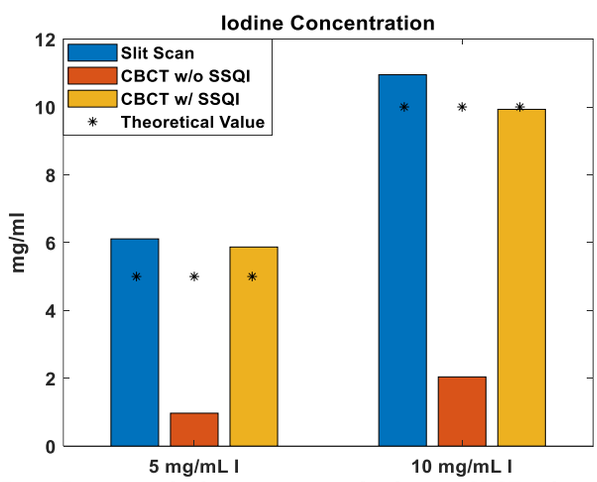


Fig. 4. Estimation of iodine concentration for slit scan, CBCT without SSQI, and CBCT with SSQI.

IV. Discussion and Conclusions

In this work, we have implemented a single-scan quantitative CBCT method based on our previously developed SSQI algorithm using a PM and a DL flat-panel detector, where the PM simultaneously captures both primary and scatter images without the need of a second scan, and the DL detector simultaneously captures DE projections that enable material decomposition of the object and accurate removal of the PM checkerboard pattern. The quantitative performance of the proposed method was evaluated using a multi-energy phantom. We found that the performance of the proposed method performed was comparable to a slit scan, which had an estimation error of less than 10% compared to the theoretical value.

With the promising preliminary work, more can be done to further investigate and improve the performance. As mentioned above, a 10% MD error was seen between the slit scan and theoretical values. This could be because there were only 42 pairs of thicknesses ranging from {0, 5.0, 10.0, 14.8, 19.9, 25.0} cm acrylic slabs and {0, 0.3, 0.6, 1.2, 1.8, 2.4, 3.0} mm copper sheets used for generating the calibration look-up table. A polynomial fitting scheme was used to interpolate the missing thickness. In the future, we can use more thickness pairs to improve the reliability of the look-up table.

Note that the material decomposition process from DE imaging will result in increased noise in the resulting basis image, and the scatter correction will even further amplify the noise. For this work or previous SSQI work, we did not implement any noise reduction scheme, which will be the focus of our immediate next step. We are planning to implement the Noise2Noise deep learning denoising method that was previously developed in our group for this purpose[13] .

We also plan to investigate the sensitivity of the proposed setting to system motion, which might affect the performance of the PM. A dedicated investigation might be needed to determine the optimal scanning scheme when a PM is present on a CBCT system.

Reference

[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, doi: 10.1088/0031-9155/40/11/007.

[2] D. Gulias-Soidan, N. M. Crus-Sanchez, D. Fraga-Manteiga, J. I. Cao-González, V. Balboa-Barreiro, and C. González-Martín, “Cone-beam ct-guided lung biopsies: Results in 94 patients,” *Diagnostics*, vol. 10, no. 12, Dec. 2020, doi: 10.3390/diagnostics10121068.

[3] C. F. T. Lôbo and M. Bordalo-Rodrigues, “Weight-bearing Cone Beam Ct Scans and its Uses in Ankle, Foot, and Knee: an Update Article,” *Acta Ortop Bras*, vol. 29, no. 2, pp. 105–110, 2021, doi: 10.1590/1413-785220212902236939.

[4] S. Vedantham, L. Shi, A. Karellas, A. M. O’Connell, and D. L. Conover, “Personalized estimates of radiation dose from dedicated breast CT in a diagnostic population and comparison with diagnostic mammography.,” *Phys Med Biol*, vol. 58, no. 22, pp. 7921–7936, 2013, doi: 10.1088/0031-9155/58/22/7921.

[5] K. Müller *et al.*, “Interventional dual-energy imaging - Feasibility of rapid kV-switching on a C-arm CT system,” *Med Phys*, vol. 43, no. 10, pp. 5537–5546, 2016, doi: 10.1118/1.4962929.

[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, 2015, doi: 10.1148/radiol.2015142631.

[7] 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, 2011, doi: 10.1118/1.3599033.

[8] E. P. Rührnschopf and K. Klingenbeck, “A general framework and review of scatter correction methods in cone beam CT. Part 2: Scatter estimation approaches,” *Medical Physics*, vol. 38, no. 9. pp. 5186–5199, 2011. doi: 10.1118/1.3589140.

[9] L. Zhu, N. R. Bennett, and R. Fahrig, “Scatter correction method for X-ray CT using primary modulation: theory and preliminary results.,” *IEEE Trans Med Imaging*, vol. 25, no. 12, pp. 1573–87, 2006, doi: 10.1109/TMI.2006.884636.

[10] L. Shi, N. R. Bennett, A. Vezeridis, N. Kothary, and A. S. Wang, “Single-shot quantitative x-ray imaging using a primary modulator and dual-layer detector,” *Med Phys*, 2023, doi: 10.1002/mp.16789.

[11] L. Shi *et al.*, “Characterization and potential applications of a dual-layer flat-panel detector,” *Med Phys*, vol. 47, no. 8, pp. 3332–3343, 2020, doi: 10.1002/mp.14211.

[12] H. Gao, L. Zhu, and R. Fahrig, “Modulator design for x-ray scatter correction using primary modulation: Material selection,” *Med Phys*, vol. 37, no. 8, pp. 4029–4037, 2010, doi: 10.1118/1.3457472.

[13] S. Wang, Y. Yang, Z. Yin, and A. Wang, “Noise2Noise for denoising photon counting CT images: generating training data from existing scans,” SPIE-Intl Soc Optical Eng, Apr. 2023, p. 2. doi: 10.1117/12.2653975.

Nonparametric statistics for the quantification of noise and contrast in CT images

Mathias Anton, Marcel Reginatto, Ulf Mäder

***Abstract*—Nonparametric statistics are explored to define image quality parameters in simulated and CT images. If efficient parameters can be found, this would pave the way for the quality assessment of clinical images. Contrast and noise are investigated.**

***Index Terms*—CT, image quality, noise, contrast, nonparametric statistics**

I. Introduction

THE quantification of image quality is a challenge when modern image reconstruction and -processing methods are involved, since the assumptions made when using established methods for image quality assessment are often violated: modern CT systems can neither be considered linear nor shift-invariant, especially when image reconstruction via artificial intelligence is employed.

With the perspective of future applications in personalised medicine to optimise image quality vs dose, we have explored the use of nonparametric statistics for the quality assessment of clinical images, since these methods do not rely on any assumption about distributions, shift-invariance and the like. As examples, we have investigated noise and contrast.

II. Nonparametric statistics for noise and contrast

A. The noise parameter τ

A noise parameter τ was derived [1] based on work by Bielecka, Obuchowicz and others [2], [3] on quantifiers for noise obtained from non-maximum-suppression images. A threshold is varied from zero to a maximum, and for each threshold value t the number $n(t)$ of pixels is determined, where the grey value of all eight neigbouring pixels is either smaller or larger by t than the centre pixel's grey value. This yields a monotonously decreasing curve $n(t) = f(t)$. τ is the weighted mean of t, using $n(t)$ as a weight and is proportional to the square root of the variance of the noise, σ. The ratio of σ/τ depends on the covariance, or, more precisely, on how fast the covariance decays with increasing px-px distance. For white noise, $\sigma/\tau \approx 2.08$.

M. Anton and M. Reginatto are with the Department of Ionizing Radiation, Physikalisch-Technische Bundesanstalt, Braunschweig, Germany. e-mail: mathias.anton(at)ptb.de.

Ulf Mäder is with the Institute of Medical Physics and Radiation Protection, University of Applied Sciences, Giessen, Germany

Manuscript received Jan 23, 2024.

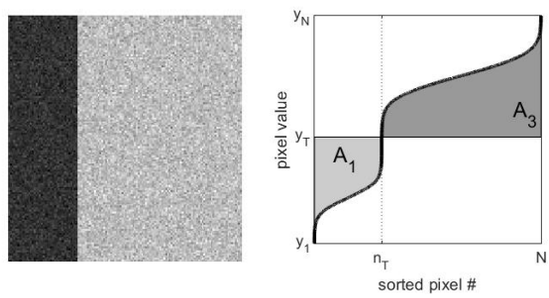


Fig. 1. Left: asymmetric step with white noise. Right: the corresponding order statistic, i.e. all N pixel values y plotted in ascending order. y_T indicates a threshold and n_T the number of the corresponding pixel.

B. The relative grey value distance δ and derived quantities

Contrast is often quantified by a grey value distance which is the difference of the mean grey values of two distinct regions. A mean relative grey value distance can also be quantified using the order statistic of the pixel values[1]. This is illustrated in Figure 1 with a simple example, an asymmetric single step function with some additional white noise. In the right part of the figure, the pixel values y are plotted in ascending order (the order statistic). When a threshold y_T is chosen, the ratio of the area between the order statistics and the horizontal line defined by y_T (the shaded areas A_1 and A_3) to the half of the total area of the graph yields a number δ which can vary between zero and two

$$\delta = \frac{A_1 + A_3}{\frac{N}{2} \cdot (y_N - y_1)}, \tag{1}$$

where N is the total number of pixels and y_1 and y_N are the lowest and the highest pixel value, respectively[2]. Multiplying δ by the range $R = y_N - y_1$ then yields the mean grey value distance $\Delta = \delta \cdot R$. In this example, the simulated grey value distance was 1, the distance Δ determined using $\delta = 0.52$ was Δ =1.05. With the help of the noise parameter τ, the distance-to-noise ratio DNR can be defined as

$$\text{DNR} = \frac{\delta \cdot R}{\tau} = \frac{\Delta}{\tau}. \tag{2}$$

C. Extension to more than two grey levels

It is possible to extend the method to multiple grey levels e.g. by applying Otsu's method for multiple thresholds [5].

[1]The parameters and their calculation are described in detail in a publication which will have been submitted to the journal *Physics in Medicine and Biology* by the time of the conference [4]

[2]For well-behaved gray value distributions, δ varies between 0 and 1.

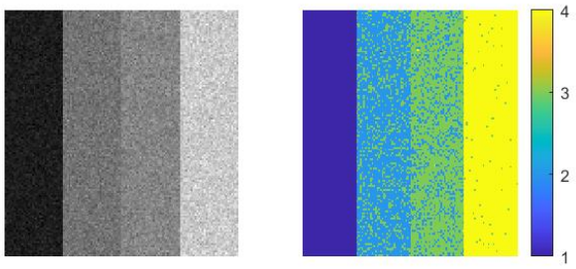


Fig. 2. Left: four-level staircase with additional white noise, Right: segmented image using the thresholds from Otsu's method, see Fig.3 The colorbar indicates the level number (1, 2, 3, 4)

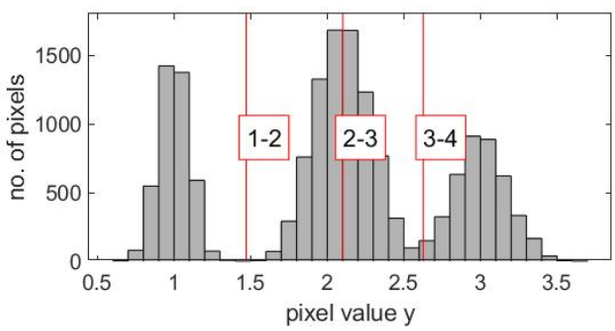


Fig. 3. Histogram of the pixel values of Fig.2, left. The red lines indicate the threshold values determined by Otsu's method, the label 1-2 denotes the transition from level 1 to level 2 and so on.

This is demonstrated with a simulated image with four grey levels and additional noise, where the middle two levels are rather close, see Figures 2 and 3. The order statistic (Fig.4) is treated section-wise, where the first section is from y_1 to the second threshold and is subdivided by the first threshold. The next section (shaded in Fig.4) is from the first to the third threshold, subdivided by the second threshold and so on. This way, three values of δ and Δ, corresponding to the three steps, can be obtained. The results for this simulated example are listed in Table I.

The leftmost column of Table I indicates the threshold (see also Figs. 3 and 4). The following columns list the mean relative grey value distance δ, the mean grey value distance Δ and the next column yields the corresponding input values to the simulation, Δ_{sim}. For the threshold (2-3), which appears in the middle of a single peak in the histogram Fig.3, the estimated Δ exceeds the true value by almost 50 %. The other two Δ-values are much closer to the simulated ones.

TABLE I
PARAMETERS DERIVED FROM THE EXAMPLE IN FIG.2

threshold	δ	Δ	Δ_{sim}	DNR
1-2	0.66	0.97	1.00	14.15
2-3	0.24	0.29	0.20	4.27
3-4	0.51	0.75	0.80	10.94

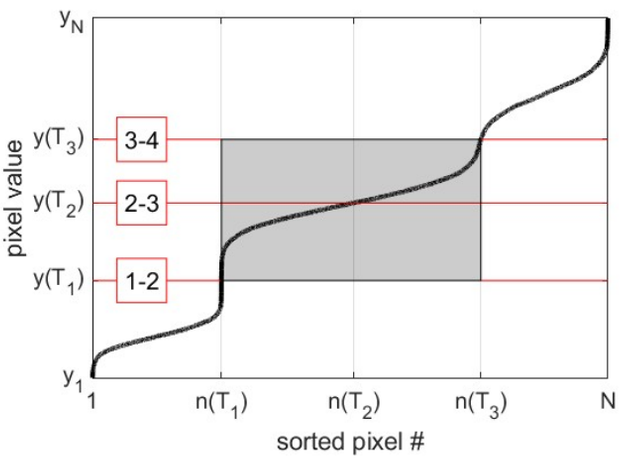


Fig. 4. The order statistic of the image in Fig.2 (left), i.e. all N pixel values y are plotted in ascending order as a function of the (ordered) pixel number. Red lines indicate the threshold values. The shaded region is the part which is used for the calculation of δ for the step 2-3.

D. *Validity limits for δ and Δ:* DNR_{lim}

Under low-noise conditions, $\Delta = \delta \cdot R$ coincides with the conventional grey value distance. The equivalence is lost, however, when noise dominates, since for the calculation of δ, no prior information on the location of pixels with "high" and "low" grey values is used. For low noise, the ratio of DNR to the signal-difference-to-noise ratio SDNR is simply σ/τ, i.e. the ratio of the noise standard deviation and the new noise measure τ. A sensible limit between "high" and "low" noise is SDNR=2. Using simulated data, it can be shown that for conditions which correspond to SDNR≥ 2, the relative (systematic) difference between Δ and the true grey value distance is less than 10 % and zero for SDNR≥ 4 [4]. From simulated data with different noise textures, comprising ranges which are realistic for CT [6] and mammography [1], a limiting value DNR_{lim} that corresponds to SDNR=2 is obtained using the proportionality $\mathrm{DNR}_{lim} = 2.37 \cdot \sigma/\tau$. The proportionality factor is independent of the shape of the decay of the covariance with increasing px-px distance. For the example of Fig. 2 with only white noise, $\sigma/\tau = 2.07 \pm 0.08$ on average, which suggest a DNR limit of $\mathrm{DNR}_{lim} \approx 4.9$. This is in accordance with the observation that Δ overestimates the true grey value distance Δ_{sim} for the threshold (2-3) since DNR $= 4.3 < 4.9$. The relative difference of Δ and Δ_{sim} for thresholds (1-2) and (3-4) is well below 10 %: their corresponding DNR values are significantly larger than 4.9 (see Table I). This means the validity of the grey value distance Δ as an estimate of the true grey value distance can be verified using DNR, provided the ratio σ/τ is available[3].

III. REALISTIC EXAMPLE: HEAD PHANTOM CT IMAGES

As a more realistic example, we have applied the methods outlined in the previous section to CT images of a head phantom. Data had been acquired on a Siemens Somatom

[3]All analyses were carried out using functions and scripts written in Matlab® (The MathWorks, Massachusetts, USA).

TABLE II
PARAMETERS OF THE HEAD PHANTOM CT IMAGES.

reconstruction kernel	exposure time-current product in mAs (A, B, C, D: labels of selected images)
H20f (soft)	7 (A), 15, 25, 40, 70, 100 (B)
H60f (hard)	7 (C), 15, 25, 40, 70, 100 (D)

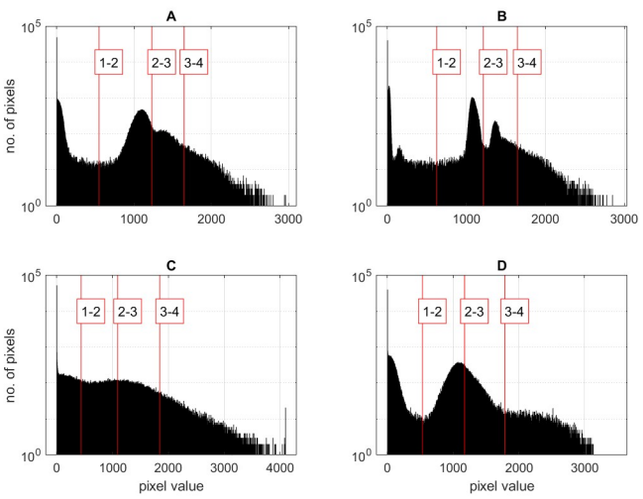


Fig. 5. Histograms of CT images of a head phantom (no. of pixels vs pixel value). Segentation using Otsu's method. Top: soft kernel (H20f); Bottom: hard kernel (H60f); left: 7 mAs; right: 100 mAs. Note the logarithmic y-axis.

CT (already in 2013) and reconstructed using filtered back projection with different convolution kernels. A protocol for the investigation of a sinusitis had been used, with a voltage of 100 kV and different exposures. The exposure-time current products are listed in Table II. Two series were selected for a comparison of the resulting image quality parameters. The four images shown as examples are indicated by the capital letters in parentheses[4].

A. Segmentation

The segmentation was performed using Otsu's method for multiple thresholds (Matlab's® built-in function `multithresh`) using the histograms shown in Fig.5 (counts of pixels given in logarithmic scale).

The resulting segmented example images are shown in Fig.6. Apparently, the segmentation is least successful for the image with the lowest dose and the hard kernel (image C, bottom left), especially the transitions between level 2 and 3 are hard to separate. The image with the highest dose and the soft kernel (B, top right) exhibits the clearest segmentation of the different materials. The harder kernel, however, appears to produce a sharper picture, which can still be inferred from the segmented image (D, bottom right).

[4]To obtain HU numbers instead of the given pixel values, a value of 1024 has to be subtracted. Dealing with negative numbers would cause undue difficulties when calculating the new parameters.

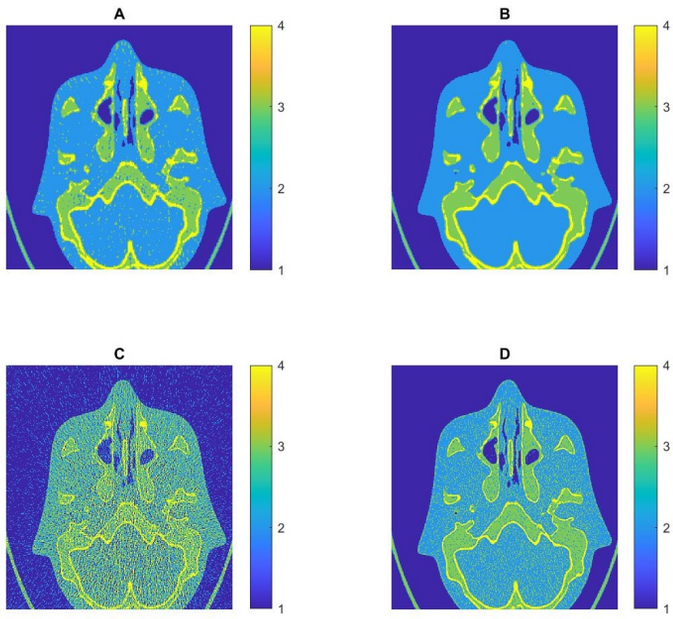


Fig. 6. CT images of a head phantom segmented into four levels / tissue types. Top: soft kernel (H20f); Bottom: hard kernel (H60f); left: 7 mAs; right: 100 mAs. The color axis denotes the level.

B. Image quality parameters

Image quality parameters were calculated according to the methods described in section II. For the estimation of σ/τ, a region of interest (ROI) was marked in the rather homogeneous, w-shaped region that represents a part of the brain, see the lower middle part of the images in Fig.6. Results for Δ and DNR are shown in Fig.7.

For the H20f convolution kernel, the grey value distances for all three thresholds are almost constant (Fig.7, top left), which is reassuring, since usually a higher exposure is not expected to change the contrast. The noise measure τ, however, is decreasing significantly with increasing exposure, which can be seen from the DNR plot on the right. Only for the smallest exposure of 7 mAs, the limit DNR_{lim} corresponding to SDNR=2 is approached for thresholds (2-3) and (3-4), the contrast between air and head (1-2) is nowhere affected by the noise, as is obvious. Only for the lowest exposure, Δ-values for the (2-3) and (3-4) threshold might be weakly hampered.

The situation is quite different for the H60f convolution kernel (lower part of Fig.7): The grey value distance (2-3) which decreases with increasing exposure is suspicious, but explained by the observation that the corresponding DNR value is lower than DNR_{lim}. The remaining two grey value distances increase with increasing exposure, as does their corresponding DNR value. For the transitions (1-2) and (3-4) only the grey value distances for the lowest two exposures are probably affected by noise.

Comparing the two series, it is observed that the grey value distance (1-2) is slightly lower for the harder convolution kernel, but that the contrast between regions 3 and 4 as measured by Δ (3-4) is significantly higher: this corresponds to the somewhat "sharper" impression in image D (Fig.6). However, the distance-to-noise ratio is always smaller for the

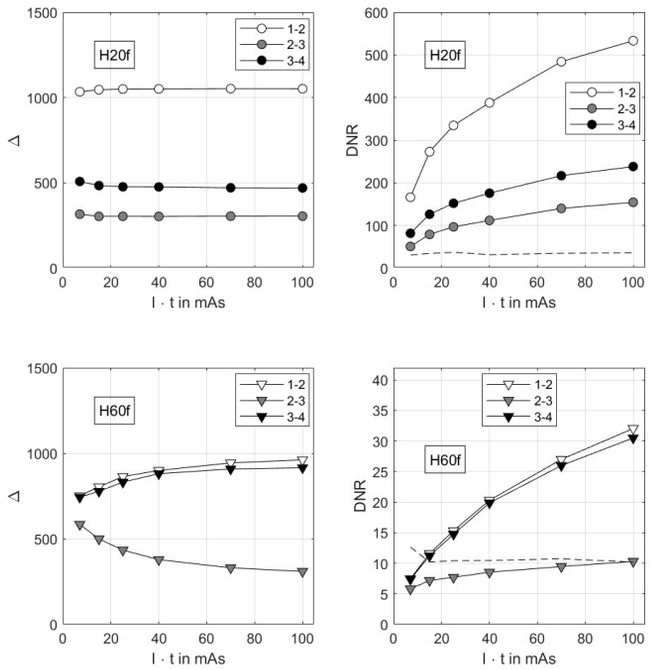


Fig. 7. Mean grey value distance Δ (left) and distance-to-noise ratio DNR (right) for the exposure series of the head phantom. Top: reconstruction with the soft kernel (H20f), Bottom: reconstruction with the hard kernel (H60f). The legends refer to the thresholds given in Fig.5. The dashed line in the DNR plots indicates the limit DNR_{lim}, corresponding to SDNR=2.

harder convolution kernel H60f.

IV. Conclusion

Using simulated data and realistic CT images of a head phantom, we have demonstrated how nonparametric statistics might be used to extract parameters for noise and contrast directly from clinical images, which would not be possible in general with conventional methods, unless the images have large homogeneous areas.

However, work remains to be done: firstly, a measure of the noise covariance based on nonparametric statistics is required (ongoing work), because a σ/τ-ratio from a homogeneous region in the image may not be available in all situations. Secondly, Δ and DNR quantify a mean grey value distance (-to-noise ratio) without referring to the spatial order in the image - the same Δ and DNR data would have resulted, if the same pixel values had been randomly distributed. An approach to solve this problem is sketched in Appendix A.

Appendix A
Mutual information of co-occurence matrices

The (relative) mutual information $U(x, y)$ obtained from grey-level co-occurence matrices [7] was explored to classify images in landscape ecology [8]. $U(x, y)$ can take on values between 0 and 1 and depends as well on the number of categories (grey levels) as on the spatial ordering of pixels.

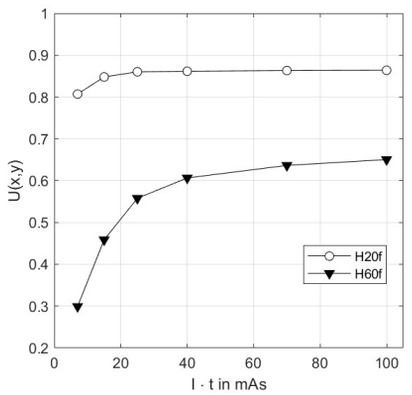


Fig. 8. Relative mutual information $U(x, y)$ as defined in [8] as a function of the exposure time-current product. Open circles: H20f; filled triangles: H60f.

These methods [8] can be directly employed for our purposes: Figure 8 shows $U(x, y)$ as a function of the exposure time-current product for the two exposure series of the head phantom. The grey levels used (i.e. the segmentations) are the same that were used for the results from Fig.7. Obviously, $U(x, y)$ indicates a higher spatial ordering of grey values for the exposure series with the softer kernel H20f. Interestingly, U remains constant for $I \cdot t > 20\ mAs$, which concides with constant grey value distances in this range (see Fig.7, top left). U is lower for the H60f series, which is in accordance with the visual impression, and also with the lower DNR compared to the H20f-series. Although Δ and DNR would remain the same when the order of the pixels were randomised, $U(x, y)$ would in such a case be close to zero. Therefore, $U(x, y)$ may add valuable supplementary information to Δ and DNR.

References

[1] M. Anton, U. Mäder, S. Schopphoven, and M. Reginatto, "A nonparametric measure of noise in x-ray diagnostic images – mammography," *Physics in Medicine & Biology*, vol. 68, no. 4, p. 045003, feb 2023. [Online]. Available: https://dx.doi.org/10.1088/1361-6560/acb485

[2] M. Bielecka, A. Bielecki, R. Obuchowicz, and A. Piórkowski, "Universal measure for medical image quality evaluation based on gradient approach," in *Computational Science – ICCS 2020*, V. V. Krzhizhanovskaya, G. Závodszky, M. H. Lees, J. J. Dongarra, P. M. A. Sloot, S. Brissos, and J. Teixeira, Eds. Cham: Springer International Publishing, 2020, pp. 406–417.

[3] R. Obuchowicz, M. Oszust, M. Bielecka, A. Bielecki, and A. Piórkowski, "Magnetic resonance image quality assessment by using non-maximum suppression and entropy analysis," *Entropy*, vol. 22, no. 2, p. 220, 2020.

[4] M. Anton, M. Reginatto, S. Schopphoven, C. Abou Jaoude, U. Mäder, M. Fiebich, F. Mauter, I. Sechopoulos, and R. van Engen, "A nonparametric measure of contrast in x-ray diagnostic images," *submitted to Physics in Medicine and Biology*, 2024.

[5] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[6] M. Anton, W. Veldkamp, I. Hernandez-Giron, and C. Elster, "RDI – a regression detectability index for quality assurance in x-ray imaging," *Physics in Medicine & Biology*, vol. 65, no. 8, p. 085017, 2020.

[7] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.

[8] J. Nowosad and T. F. Stepinski, "Information theory as a consistent framework for quantification and classification of landscape patterns," *Landscape Ecology*, vol. 34, no. 9, pp. 2091–2101, 2019.

Accelerating Diffusion-Based CT Image Restoration through Sampling Step Sharing

Xi Chen, Tzu-Cheng Lee, Liang Cai

***Abstract*— Denoising diffusion probabilistic models (DDPM) are a novel class generative model that outperformed supervised learning-based methods, as well as other generative models in many applications. Various studies demonstrate the potential for diffusion-based super-resolution and denoising of medical imaging. We applied a variation of DDPM to the task of CT image restoration. To address the practical concern of inference speed of diffusion-based methods, we proposed an acceleration strategy for volumetric CT slices using sampling step sharing. The proposed method enables 35× speed up compared to the conventional workflow while maintaining comparable image quality.**

***Index Terms*— DDPM, generative models, super-resolution, low-contrast detectability**

I. Introduction

Restoration of images degraded by noise, sub-sampling, or other artifacts is an important topic in medical image processing. The past decade saw an increase of deep learning (DL) based techniques applied to the restoration of medical image. A common formulation of the problem uses a supervised-learning framework, where a neural network model learns to produce a restored output from a lower quality input image through an optimization process based on gradient descent on the prediction error [1], [2]. Supervised learning methods achieved great denoising power, but exhibits a general tendency of over-smoothing the image, especially in low contrast regions, resulting in loss of fine texture and low contrast detail. Such behavior is especially undesirable in the clinical imaging context. Generative models such as variational autoencoder (VAE) [3] and generative adversarial model (GAN) [4] were shown to be less prone to over-smoothing compared to supervised learning. However, they pose new issues of sample quality and trainability. Denoising diffusion probabilistic models (DDPM) are a recently developed class of generative model [5]. Shown to outperform VAEs and GANs in many scenarios [6], DDPM started to be applied to the problem of image restoration with great success. Various pioneer studies of applying DDPM to the denoising and super-resolution of medical imaging presented promising result [7]–[10].

One major limitation of DDPM, compared to supervised learning and other generative models is the heavy computational cost. Sampling with DDPM involves reversing each diffusion step with a forward pass of the trained model. A typical selection for number of timesteps ranges from several hundreds to thousands. Thus, inference time consumption can become a hinderance for practical scenarios. Improving the efficiency of DDPM became a topic of great interest [11]. In existing studies on efficient diffusion models, a common approach to improve sampling speed is to reduce the number of sampling steps required to recover image in the data domain, which can be achieved using early stopping of the diffusion process [12] or condensing multiple sampling steps into one [13]. Alternatively, the diffusion process can be formulated into a non-Markovian process with the same training objective and reduced sampling time [14]. Applying diffusion models in the latent space can both accelerate the process and reduce amount of required computational resource [15]. These existing methods focus on accelerating inference on a single

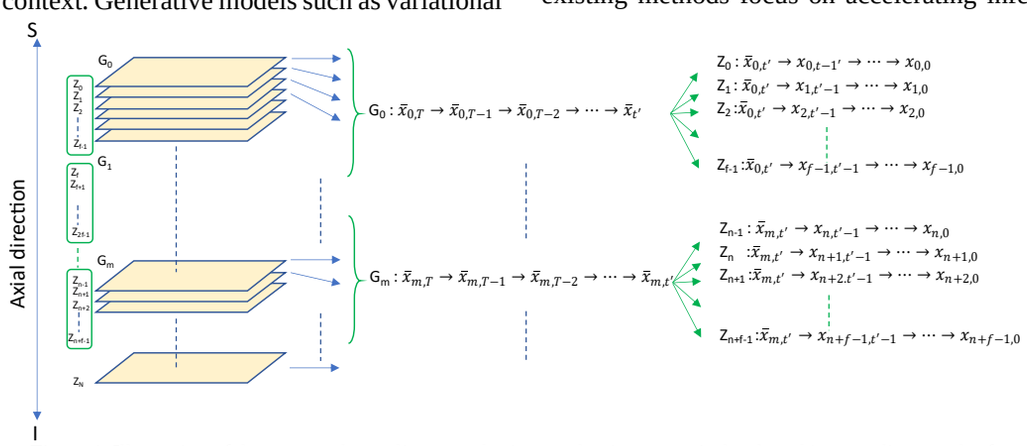


Figure 1. Illustration of the proposed sampling step sharing mechanism. $\bar{x}_{m,t}$ is the shared intermediate result of group m at step t, conditioned by a shared conditional image m. $x_{n,t}$ is the individual intermediate result of slice n at step t, conditioned by an individual conditional image n.

Xi Chen (xichen@mru.medical.canon), Tzu-Cheng Lee, and Liang Cai are with Canon Medical Research USA, Inc., Vernon Hills, IL 60061, USA.

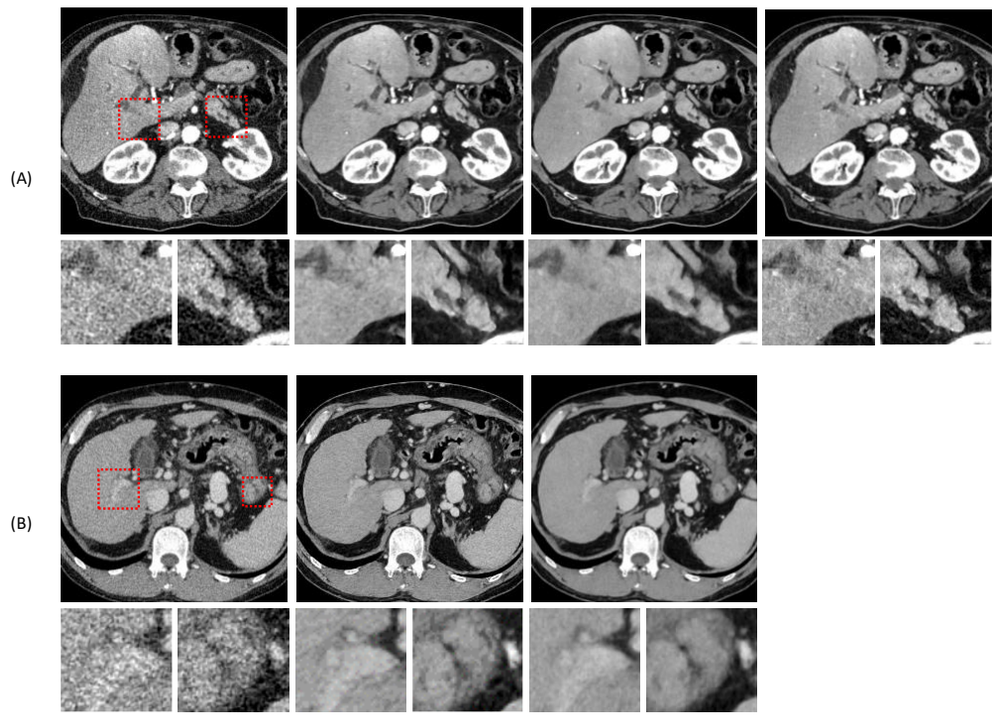


Figure 2. Representative images for DDPM-restored LR cases. (A) Test on training results with LR high noise conditional image, DDPM reconstruction, UNet-based supervised learning result, and HR target image. (B) DDPM applied on low-contrast clinical data with LR high noise conditional image, DDPM reconstruction and UNet-based supervised learning result. Zoom in views of two selected ROIs from the locations marked by dotted boxes. Window level / window width: 40/300 HU.

inference. However, clinical data such as CT, MRI and ultrasound often come in bulk of volumetric slices or temporal sequence, where adjacent instances share some high-level similarity. It would be beneficial to design an acceleration strategy targeting bulk processing of data with some degree of correlation between individual samples.

In this study, we propose a novel strategy of workflow acceleration for DDPM-based CT image restoration. The method targets processing stacks of 2D slices representing a 3D volume, where neighboring slices share some high-level features. The method provides two tunable parameters: the number of shared steps in sampling process and the sharing range in axial direction. By selecting an appropriate combination of sharing parameters for our testing data, we were able to achieve 35 × acceleration over the conventional workflow. The proposed method can be combined with any existing acceleration technique for single instance sampling to achieve greater performance gain.

II. Methods

A. Conditional DDPM

In this study we followed the adaptations proposed in the SR3 method [16]. For each iteration of the training process, a high resolution (HR) training sample x_0 and its corresponding low resolution (LR) image y were drawn from the training data distribution. The noise level α_t was picked from a predefined noise schedule. The noise image was sampled from a random gaussian noise $\epsilon \sim N(0, I)$. Each training step was performed by taking gradient descent step on $\Delta_\theta \left\| \epsilon - \epsilon_\theta \left(\sqrt{\overline{\alpha_t}} x_0 + \sqrt{1-\overline{\alpha_t}} \epsilon, \alpha_t, y \right) \right\|^2$, where $\overline{\alpha_t} = \prod_{s=1}^{t} \alpha_s$, θ are the trainable parameters, and ϵ_θ is the noise predicted by the model. The sampling process conditioned by a LR image y starts from a pure noise image $x_T \sim N(0, I)$. For each t, the image at t-1 is recovered by $x_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{1-\alpha_t}{\sqrt{1-\overline{\alpha_t}}} \epsilon_\theta (x_t, \alpha_t, y) \right) + \sqrt{1-\alpha_t} z$, where $z \sim N(0, I)$ for $t > 0$ and $z = 0$ for $t = 0$. The two important modifications are the inclusion of a LR conditional image y and using the noise level parameter α_t instead of the timestep t, which allows for different training and sampling noise schedule. We used 2000 timesteps with a linear noise scheduler for training, and 200 timesteps for sampling. We adopted the 2D residual UNet architecture described in the SR3 paper, but without the self-attention layers. The model was trained on the CT dataset for 5000 epochs. The baseline supervised learning method is a 2D residual UNet with the same architecture trained using the same set of training pairs as DDPM. The model was trained by minimizing the *L1* loss using an Adam optimizer. All data acquisition detail, reconstruction protocols, training datasets and simulation processes were described as in previous publication [17].

B. Sampling Step Sharing

In the reverse diffusion process, high level (low frequency) features can usually be recovered early in the inference steps (large t), while finer detailed features are revealed in the very last few steps (small t). Based on this assumption, the main idea of our sampling step sharing workflow is to share intermediate results among neighboring slices, where the low frequency features are similar. Figure 1 illustrates the inference process using sampling step sharing strategy. A volumetric CT scan Z with N slices in the axial direction is divided into groups (G_m) of f slices, where f is the sharing range size among which the sampling steps associated with large Ts are shared. The shared sampling steps were performed using the mean image $\overline{y_m}$ of the LR input slices y_i's in group m as the guide

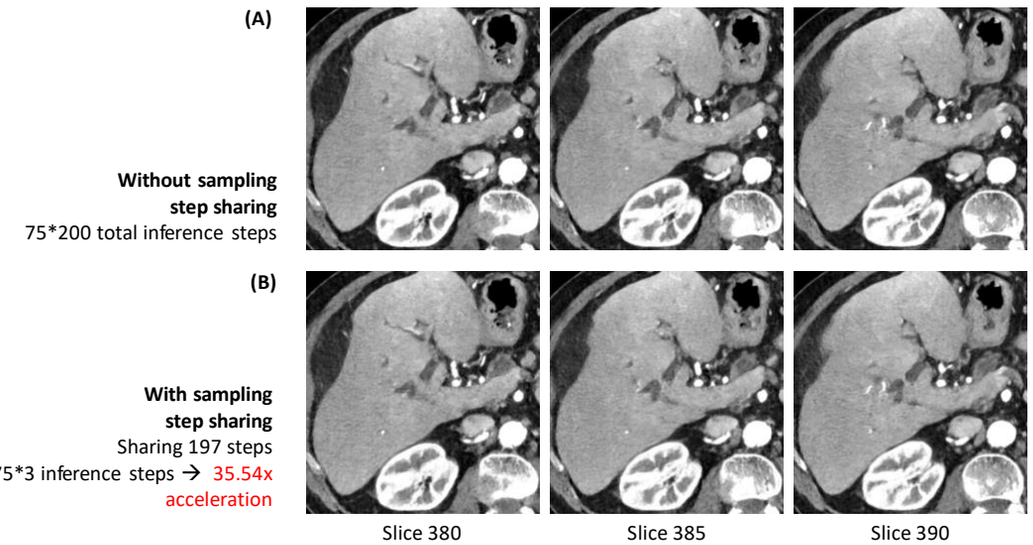


Figure 3 Neighboring image quality comparisons between (A) Single slice restoration result using the conventional sampling procedure without sampling step sharing. (B) Same slice image restoration obtained with sampling step sharing of 197 out of 200 steps (t'=3) and a sharing range of 75 slices (f=75). Window level / window width: 40/300 HU. Reconstruction pitch = 0.25mm.

image. Like the standard workflow, the shared sampling start from a pure noise image $\bar{x}_{m,T}$ to gradually recover noisy images $\bar{x}_{m,T-1}, \bar{x}_{m,T-2} \dots, \bar{x}_{m,t'}$, where t' is the timestep that sharing stops. The recovered HR image corresponding to each slice is obtained performing the remaining sampling steps $t'-1, t'-2, \dots, 0$ using the shared intermediate sample $\bar{x}_{m,t'}$ as the input at timestep t' and each individual LR input slices y_i as guide for each timesteps. The whole process requires t' forward passes of the network to recover the shared intermediate sample, and $f \times (T-t')$ forward passes for each individual slices, which is $(f-1) \times (T-t')$ steps less than the $f \times T$ steps required by the conventional workflow. The selection of sharing parameters t' and f affects both the acceleration factor and the quality of the HR images. This will be discussed further in section 3. Note that as a workflow acceleration solution, sampling step sharing strategy can be combined with any acceleration technique targeting single instance to further improve the overall processing speed.

III. Results

A. CT Image Restoration Using DDPM

The image restoration performance was first evaluated on an image sampled from the training distribution (see Figure 2.A), Supervised learning and DDPM were able to achieve similar denoising capability compared to the conditional image. However, in the selected two regions of interest (ROI), supervised learning showed obvious over-smoothing in zoomed ROI, while DDPM was able to retain a shaper contrast and cleaner boundaries around the hyperdense area and able to recover more of the low contrast hypodense features. We extended our evaluation to real clinical data to test the generalizability of the DDPM method (see Figure 2.B). Similarly, we selected two representative ROIs for a qualitative comparison between the performance of supervised learning and DDPM. In clinical case, DDPM restoration provided better contrast and clearer border for the hyperdense region within the liver (left ROI), as well as higher contrast for the ROI on the right.

B. Sampling Step Sharing Workflow Acceleration

Figure 3 showed DDPM-restored image (three neighboring slices, 1.25mm away) with proposed sampling step sharing approach had very similar image quality when compares to same slices without proposed acceleration method. To further quantify the effect of sharing parameter selection on the final restored image quality, we tested different combinations of sharing parameters on data from the same patient. The restoration was evaluated by computing the structural similarity (SSIM) between the restored image sequence using sampling step sharing, and the restored image sequence using the conventional DDPM sampling workflow without sampling step sharing (Figure 3 A). Note that due to the random sampling of noise in the reverse diffusion process, different iterations on the same image using the exact same model and processing method will have slight differences, resulting in ~0.995 SSIM between the restored images. Therefore, we consider a SSIM value greater than 0.99 to be an indication of minimal loss of quality. Trends showed that we are able to share as many as 195 out of 200 timesteps while maintaining relatively high fidelity of the restored image with SSIM > 0.994 compared to conventional workflow. As expected, increasing the number of shared sampling steps and increasing the number of slices within the sharing unit both negatively impact final image quality. We identify the sharing parameters represented by "green point" in Figure 4 A (sharing 197 steps among 75 slices) as the edge case setup that obtains the best acceleration factor without sacrificing too much image quality (SSIM ≈ 0.991). Using the green-point setup (see Figure 4 C), the full sampling process for all 75 slices requires only 422 total processing steps, which is 35.5 times acceleration than the total processing steps required using the conventional workflow. On the contrary, the "red-point" sharing parameters Figure 4. A is a failure case setup where "oversharing" happens: we aggressively shared 199 out of 200 steps, leaving only a single independent sampling step on each slice. The SSIM drops below 0.95 for a sharing range of 20 slices (see Figure 4. D).

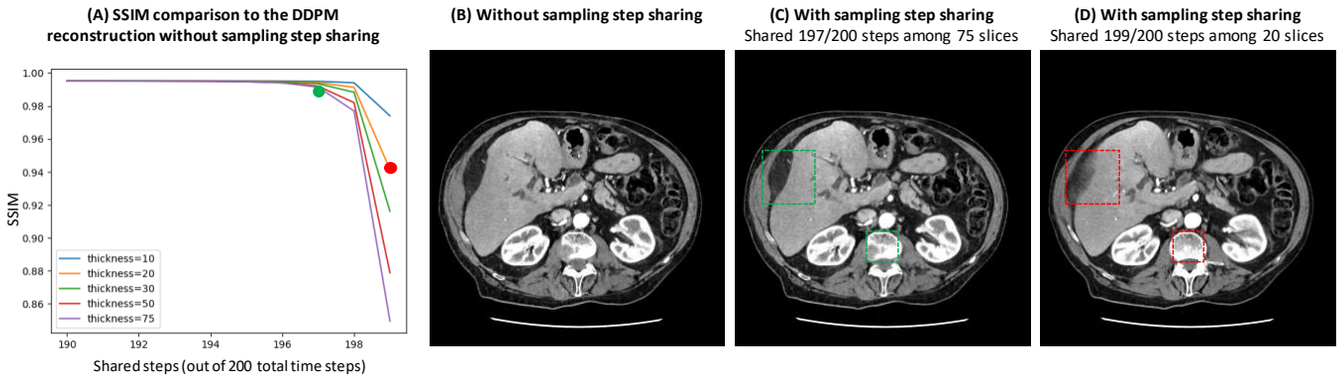


Figure 4 Trade-off plot showing the effect of different sharing parameter selection on the final restored image, measured by the SSIM between the restored image sequence and the DDPM restoration results without sampling step sharing. (B) Single slice restoration result using the conventional sampling procedure without sampling step sharing. (C) Edge case restoration obtained with sampling step sharing of 197 out of 200 steps and a sharing range of 75 slices as labeled in green spot in A. (D) Failure case result of the same slice obtained with sampling step sharing of 199 out of 200 steps and a sharing range of 20 slices as labeled in red spot in A. Window level / window width: 40/300 HU. Reconstruction pitch = 0.25mm

The late sampling steps (steps 197-199) had already started to reveal high frequency details supposedly unique to each individual slice conditioned by the mean image. As a result, each individual restored slice showed features present in other slices in the sharing range. Note that increasing sharing range while keeping a relatively small number of shared steps appears to be a more profitable strategy to obtain better acceleration factor: in the failure case, the acceleration factor is 18.26, which is less than that of the edge case obtained by sharing less steps among more slices.

IV. Discussion and Conclusion

The study trained a conditional DDPM model for CT image restoration. We showed that DDPM-based CT image restoration has less over-smoothing artifact compared to the conventional supervised learning, yielding better recovery of low contrast details and better preservation of noise texture. To address the limitation of sampling speed, we proposed a workflow acceleration strategy for DDPM-based CT image restoration. Different from existing acceleration techniques focusing on inference of single instance, the sampling step sharing workflow targets bulk processing of volumetric CT slices. We showed that the proposed method achieves up to 35x acceleration compared to conventional workflow without sharing, while maintaining good image quality (~0.99 in SSIM). Combined with timestep skipping acceleration technique, we can achieve 350x acceleration, which is less than an order of magnitude slower compared to supervised learning. The proposed method relies on the shared early-recovered structures between neighboring slices. How factors such as physiology of the patient and axial sampling frequency would affect acceleration capability needs to be studied further.

References

[1] Y. Li, B. Sixou, and F. Peyrin, “A Review of the Deep Learning Methods for Medical Images Super Resolution Problems,” *IRBM*, vol. 42, no. 2, pp. 120–133, Jan. 2021, doi: 10.1016/j.irbm.2020.08.004.

[2] S. Kaji and S. Kida, “Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging,” *Radiol Phys Technol*, vol. 12, no. 3, pp. 235–248, Jan. 2019, doi: 10.1007/s12194-019-00520-y.

[3] A. Abid and J. Zou, “Contrastive Variational Autoencoder Enhances Salient Features,” *arXiv preprint arXiv:1902.04601*, Jan. 2019.

[4] C. Ledig *et al.*, “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Jan. 2017, pp. 105–114. doi: 10.1109/CVPR.2017.19.

[5] J. Ho, A. Jain, and P. Abbeel, “Denoising Diffusion Probabilistic Models,” *Adv Neural Inf Process Syst*, vol. 33, pp. 6840–6851, Jan. 2020.

[6] P. Dhariwal and A. Nichol, “Diffusion Models Beat GANs on Image Synthesis,” *Adv Neural Inf Process Syst*, vol. 34, pp. 8780–8794, Jan. 2021.

[7] Q. Li *et al.*, “Ultra-low Dose CT Image Denoising based on Conditional Denoising Diffusion Probabilistic model,” in *2022 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, IEEE, Jan. 2022, pp. 198–205. doi: 10.1109/CyberC55534.2022.00041.

[8] Q. Gao and H. Shan, “CoCoDiff: a contextual conditional diffusion model for low-dose CT image denoising,” in *Developments in X-Ray Tomography XIV*, B. Müller and G. Wang, Eds., SPIE, Jan. 2022, p. 16. doi: 10.1117/12.2634939.

[9] W. Xia, Q. Lyu, and G. Wang, “Low-Dose CT Using Denoising Diffusion Probabilistic Model for 20 Speedup,” *arXiv preprint arXiv:2209.15136*, Jan. 2022.

[10] A. Kazerouni *et al.*, “Diffusion models in medical imaging: A comprehensive survey,” *Med Image Anal*, vol. 88, p. 102846, Jan. 2023, doi: 10.1016/j.media.2023.102846.

[11] A. Ulhaq, N. Akhtar, and G. Pogrebna, “Efficient Diffusion Models for Vision: A Survey,” *arXiv preprint arXiv:2210.09292*, Jan. 2022.

[12] Z. Lyu, X. XU, C. Yang, D. Lin, and B. Dai, “Accelerating Diffusion Models via Early Stop of the Diffusion Process,” *arXiv preprint arXiv:2205.12524*, Jan. 2022.

[13] T. Salimans and J. Ho, “Progressive Distillation for Fast Sampling of Diffusion Models,” in *International Conference on Learning Representations*, Jan. 2022.

[14] J. Song, C. Meng, and S. Ermon, “Denoising Diffusion Implicit Models,” *arXiv preprint arXiv:2010.02502*, Jan. 2020.

[15] 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*, Jan. 2022, pp. 10684–10695.

[16] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, “Image Super-Resolution Via Iterative Refinement,” *IEEE Trans Pattern Anal Mach Intell*, pp. 1–14, 2022, doi: 10.1109/TPAMI.2022.3204461.

[17] T.-C. Lee *et al.*, “Deep learning enabled wide-coverage high-resolution cardiac CT,” in *7th International Conference on Image Formation in X-Ray Computed Tomography*, J. W. Stayman, Ed., SPIE, Oct. 2022, p. 102. doi: 10.1117/12.2647130.

Lifelike and Deformable Lung Phantoms for 4DCT Imaging: A Three-Dimensional Printing Approach

Jessica Y. Im, Neghemi Micah, Amy E. Perkins, Michael Geagan, Sven Kabus, Kai Mei, Peter B. Noël

Abstract— **Respiratory motion phantoms can be used for evaluation of CT imaging technologies such as motion artifact reduction algorithms and deformable image registration. However, current respiratory motion phantoms do not exhibit detailed lung tissue structures and thus do not provide a realistic testing environment. This paper presents PixelPrint^{4D}, a method for 3D-printing deformable lung phantoms featuring highly realistic internal structures, suitable for a broad range of CT evaluations, optimizations, and research. The phantom in this study was designed with a patient 4DCT as a reference and 3D-printed using an extended version of the PixelPrint method for developing patient-specific CT phantoms. A flexible thermoplastic polyurethane (TPU) 3D-printing material was used, which produced regions with attenuation between -840 and -48 Hounsfield units (HU). A linear compression device was then designed and used to compress the phantom in the superior-inferior (SI) direction, and the phantom was scanned at different compression levels matched to the diaphragm displacements measured on the reference patient 4DCT. Deformable image registration (DIR) was performed, and motion vector fields were obtained for both patient and phantom images. SI displacements of selected features in the lung had mean errors of 0.5 mm difference from the patient, or less than the reconstructed slice thickness. In conclusion, the deformable lung phantom developed in this study exhibits realistic lung structures and deformation characteristics under compression, indicating potential for advancing more lifelike respiratory motion phantoms.**

Index Terms— **Deformable lung phantom, 3D-printed phantom, Respiratory motion phantom, CT imaging phantom**

I. Introduction

Respiratory motion lung phantoms for computed tomography (CT) imaging have been developed and utilized for various applications in radiation oncology and diagnostic radiology . Some examples of applications include evaluation of dose estimation [1-4], motion artifact reduction [5], four-dimensional CT (4DCT) [4, 6], tumor tracking [7], deformable image registration (DIR) [8], and CT ventilation imaging [8, 9]. Phantoms used for these applications require realistic and reproducible motion characteristics.

Current respiratory motion phantoms include both commercially available phantoms and those that have been developed in academic research settings. These can be broadly grouped into several categories: fully rigid phantoms which undergo bulk translation [5, 7], phantoms with separate rigid body and tumor portions which can slide relative to each other [10-12], compressible foam or 3D-printed phantoms with higher density inserts embedded inside [1-3, 6, 8, 9, 13, 14], and *ex vivo* lung phantoms [15].

While many of these respiratory motion systems are capable of precise motion control with various breathing patterns, most of the lung phantoms used in these examples have simplified geometries, consisting of a homogenous lung phantom with one or more higher density geometric inserts. A few phantoms include simplified airway structures and/or vessels [8, 9, 14], but do not appear fully realistic due to simplified background lung parenchyma. Those that do appear more realistic are either rigid [7], or use *ex vivo* lungs which can only be used for a limited time [15]. To our knowledge, there are no prior existing anatomically accurate, deformable CT lung phantom suitable for repeated use.

In this work we introduce technology for manufacturing of lifelike and deformable lung phantom by expanding upon the previously developed PixelPrint method [16-19]. PixelPrint has been shown to produce detailed patient-based phantoms with high levels of structural and Hounsfield unit (HU) accuracy by utilizing the partial volume effect to emulate regions of varying HU [19, 20]. Whereas previously PixelPrint has only been used to produce rigid phantoms, this work introduces PixelPrint^{4D}, an adaptation to the PixelPrint method, for fabricating deformable CT imaging phantoms using flexible 3D-printing material. The lung phantom produced in this study demonstrates detailed internal lung structures with realistic deformation patterns when compressed.

II. Methods

A. Phantom Design

The lung phantom was printed with a flexible 3D-printing material, thermoplastic polyurethane (TPU) (Ninjaflex (85A), Ninjatek, PA, USA) on a fused filament fabrication (FFF) printer (Lulzbot TAZ Sidekick with M175 v2 tool head, Fargo Additive Manufacturing Equipment 3D, LLC Fargo, ND, USA) with a 0.25 mm brass nozzle. The TPU has a specific gravity of 1.19 g/cc, a tensile yield strength of 4 Mpa, and 65% elongation at yield. Several other flexible filaments including other brands of TPU and thermoplastic elastomers (TPE) were tested, and

We acknowledge support through Philips Healthcare.

J. Y. Im is with the Department of Bioengineering and Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. N. Micah is with Swarthmore College, Swarthmore, PA, USA. A. E. Perkins and S. Kabus are with Philips Healthcare, in Cleveland, OH, USA and Hamburg, Germany respectively. M. Geagan, K. Mei, and P. B. Noël are with the Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. (e-mail: imjess@seas.upenn.edu).

Ninjaflex was found to be the most reliable at printing thin lines, as low as 0.15 mm, required for PixelPrint. The 3D printer geometric code (g-code) instructions for the lung phantom were generated using PixelPrint software [19]. Filament was extruded at 220°C in a line pattern which leaves sub-voxel sized air spaces to achieve lung-like attenuation levels. The infill ratio of the printed phantom was adjusted on a voxel-by-voxel basis by varying the speed of the nozzle while maintaining a constant extrusion rate.

Calibration phantom

To determine the relationship between the printed infill ratio and HU of TPU, a 10 cm calibration phantom with 10 sectors of infills ranging from 15% to 100% was 3D-printed and then scanned at 120 kVp with an exposure of 150 mAs. The mean HU values in circular regions of interest (ROIs) placed in each sector were measured and used to determine the calibration function relating target HU to printed infill percentage.

Patient-based deformable lung phantom

The deformable lung phantom was designed using a retrospectively selected patient 4DCT chest scan (Table 1) with 8 gated time points. The image at the state of maximum inhalation (gate 0.0%), was selected as the reference image for printing. The other phases of the respiratory cycle were replicated by compressing the phantom. The right lung was segmented using an automated deep-learning based segmentation algorithm [21]. The segmented lung, along with about 4 mm of diaphragm directly below the lung, was converted to g-code instructions using PixelPrint software and the function relating HU to infill obtained from the calibration phantom (1). The region under the segmented lung and diaphragm was filled in with a uniform infill percentage of 24% or about -750 HU to support the phantom as it was printed (Fig. 1A). The complete phantom (Fig. 1B) was 23 cm tall and total printing time was about 2 weeks.

Compression device

A preliminary compression device was designed to compress the lung phantom to emulate different stages of the respiratory cycle. Since physiological lung motion is primarily in the superior-inferior (SI) direction due to the motion of the diaphragm [22], the lung motion was modeled approximately as linear compression and expansion along the SI axis. The compression device (Fig. 1D) was designed as a clamp-like structure with the phantom placed between a stationary wall on the superior side and a moving wall controlled by a lead screw on the inferior side. Additionally, an inverse of the top half of the lung phantom was 3D-printed as a rigid cap (Fig. 1C) using PLA and placed between the stationary wall and the phantom to ensure even distribution of forces to the phantom during compression.

B. Image Acquisition

The uncompressed lung phantom was scanned on a spectral-detector CT scanner (Spectral CT 7500, Philips Healthcare, OH, USA) and conventional images were acquired. The phantom was then placed in the compression device with the diaphragm side adjacent to the moving wall and the top of the lung phantom toward the stationary wall (Fig. 1C). The phantom was compressed by advancing the moving wall by 0.0, 4.2, 9.0, 13.1, 14.8, 14.1, 9.1, and 3.5 mm from resting position to match the motion of the patient diaphragm at each gate of the patient 4DCT and was scanned at each compression level. Image acquisition and reconstruction parameters are listed in Table 1.

TABLE I
IMAGE ACQUISITION AND RECONSTRUCTION PARAMETERS

	Patient 4DCT scan	**Phantom scans**
CT Scanner	Gemini TF Big Bore (Philips)	Spectral CT 7500 (Philips)
Tube voltage [kVp]	120	120
Exposure [mAs]	602	600
Scan mode	Helical	Helical
Slice thickness [mm]	3	1 (static scan) 3 (compression scans)
Field of View [mm]	600	300
Matrix size	512 x 512	512 x 512
Pixel spacing [mm]	1.1719	0.5859
Reconstruction Filter	B	B

C. Image Analysis

The lung phantom was first compared in its non-deformed state to the reference patient image. The static scan of the phantom was rigidly registered to the patient image, and attenuation profiles along a line drawn through the same anatomical location in the patient and phantom images were compared.

Next the deformations of the compressed lung phantom and its internal structures were compared to the patient 4DCT time series. DIR was performed on both the patient and phantom image sequences using the fast elastic image registration tool (FEIR) [23] to obtain motion vectors and to map all images to the same mid-position state. The resulting phantom images were then rigidly registered to the patient images to align the internal structures. Then motion vectors in each of the three orthogonal directions (right/left (RL), anterior/posterior (AP),

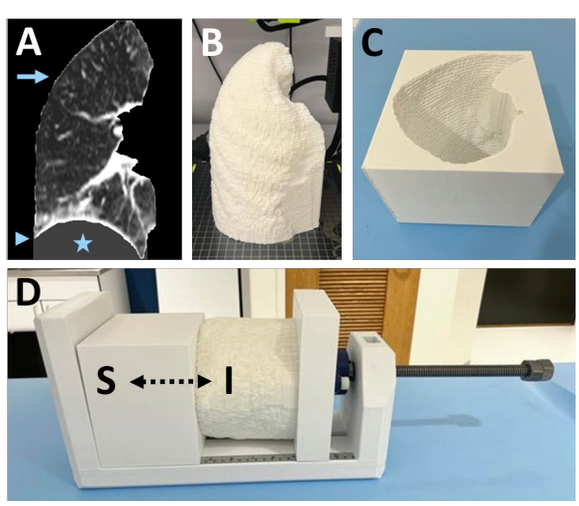


Fig. 1. (A) Coronal view of input image volume to PixelPrint showing the segmented patient lung (arrow), diaphragm (arrowhead), and region under the diaphragm filled with -750 HU (star). (B) 3D-printed deformable lung phantom (C) 3D-printed cap of inverse of the top portion of lung phantom (D) Assembled compression device with cap and lung phantom inserted.

and superior/inferior (SI)), at locations selected in the tumor, vessels, and diaphragm were compared between the patient and phantom lungs.

III. Results

A. Calibration Phantom

The calibration phantom showed a linear relationship between printed infill ratio of TPU and resultant HU (Fig. 2). The measured HU values in the calibration phantom ranged from -840 HU to -48 HU, and the equation correlating HU and infill percentage was:

$$HU = -981 + 9.54 * Infill \quad (1)$$

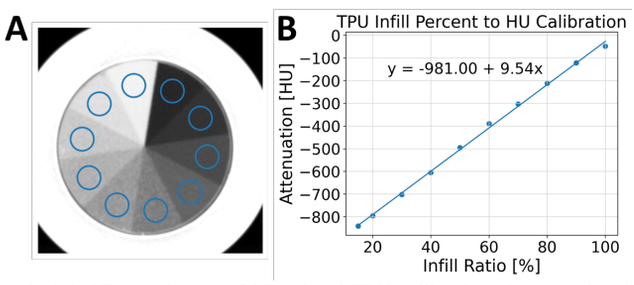


Fig. 2. (A) CT scan image of 3D-printed TPU calibration phantom showing circular ROIs where mean HU values were measured (ww: 1000, wl: -500). (B) Calibration curve showing the relationship between printed infill ratio and CT attenuation value of TPU.

B. Lung Phantom

The printed phantom demonstrates realistic structures such as vessels, bronchi, tumor, and lung fissures which visually match the structures seen in the patient images (Fig. 3A-B). The attenuation profiles of the patient and phantom scans show a strong correlation between the phantom and patient (Fig. 3C) with a mean attenuation difference of 18 HU along the measured line. A video showing a pseudo-4DCT obtained from playing the compression scans of the phantom in sequence compared to the patient 4DCT was also produced.

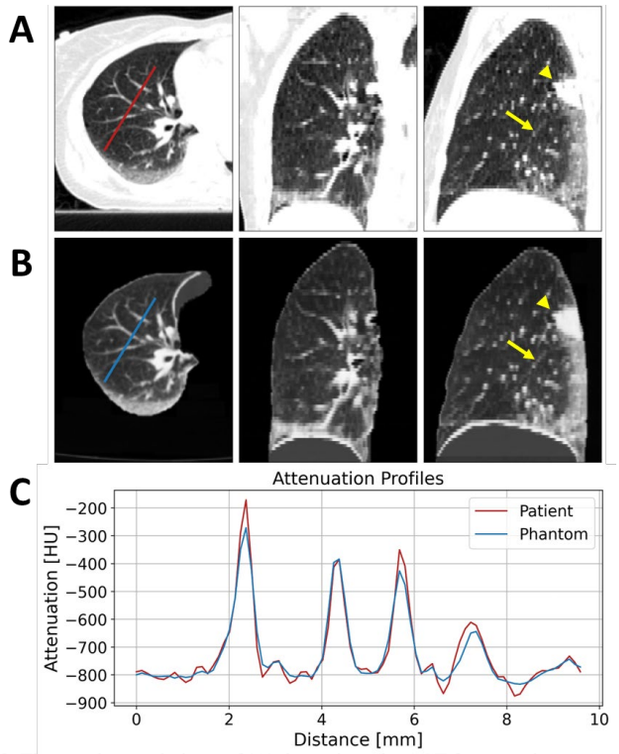


Fig. 3. Three orthogonal views of (A) the patient chest CT from respiratory gate 0.0% and (B) phantom scanned in uncompressed state demonstrating internal lung structures including the tumor (arrowhead) and lung fissure (arrow). (C) Attenuation profiles corresponding to the lines shown in the axial views of the patient (A) and phantom (B) images. (ww: 1000, wl: -500)

Motion vectors of both the phantom and patient lung in the three selected locations (tumor, vessel, and diaphragm) (Fig. 4) were greatest in the SI direction with amplitudes up to 15 mm. In contrast there was less motion in the RL and AP directions, with amplitudes < 2 mm and < 5 mm respectively. The phantom motion vectors measured across the respiratory cycle in the three selected locations show mean errors of 0.5 and 0.7 mm in the SI and RL directions respectively compared to the patient. These mean errors are smaller than the spatial resolution of the

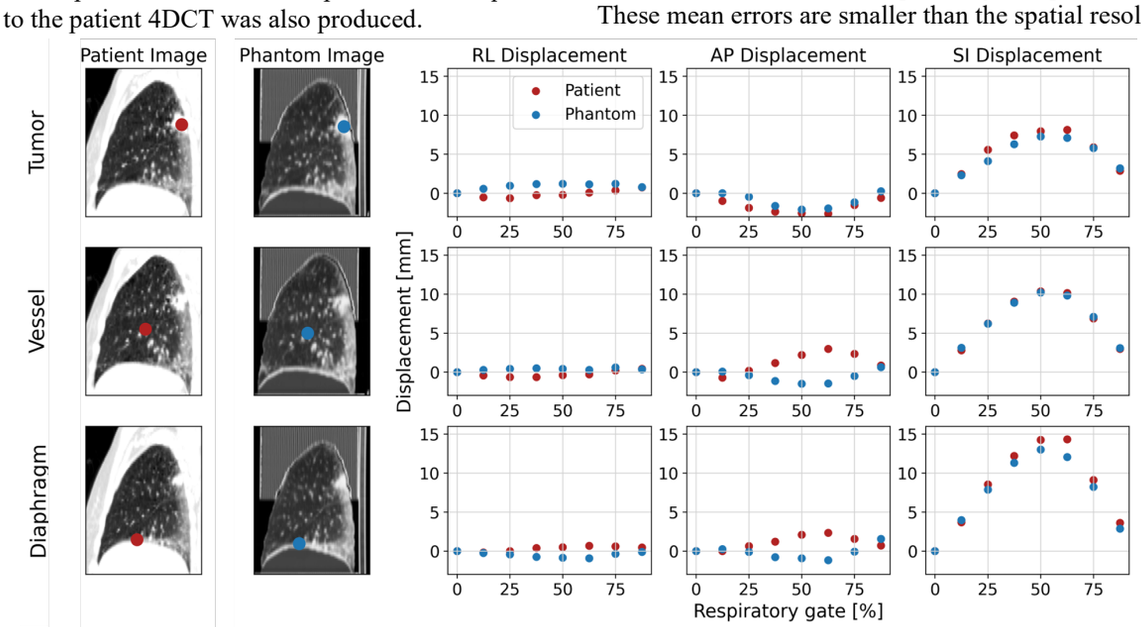


Fig. 4. (Left) Patient and phantom images showing locations where motion vectors were measured (ww: 1000, wl: -500). (Right) Plots comparing displacements in the RL, AP, and SI directions of each location over the course of the respiratory cycle, between the patient and phantom.

patient scan. The motion of the phantom differs slightly more from the patient in the AP direction, with a mean error of 1.3 mm, which is roughly the size of one pixel in the patient images.

IV. Discussion

This work introduces PixelPrint^{4D}, a technology for the design and fabrication of deformable CT imaging lung phantoms which demonstrate realistic details and compression characteristics. CT scans of the phantom show detailed internal lung structures including the lung fissures, making this phantom a valuable tool for assessing motion effects on fine structures in the lungs. Under compression, the deformations of various lung structures in the SI and RL directions matched the motion of the same structures found in the patient scan. At the compression levels implemented in this study, no permanent deformation was observed, although further testing is needed to assess the mechanical robustness of the phantom. Overall, PixelPrint^{4D} has the potential to provide a realistic testing environment for various technologies such as motion correction, tumor tracking algorithms, and CT ventilation. The PixelPrint^{4D} phantom fabrication process described in this study is also highly customizable and can be adapted in future iterations to match different clinical scenarios, to allow for fitting of additional components such as dosimetry devices, or to be customized for use in various respiratory motion actuation devices.

There are several limitations to this study. First, the deformation of the lung phantom was only limited to the SI direction whereas motion in the AP and RL directions were not controlled. This likely is the cause of the discrepancy between the AP motion characteristics of the phantom and patient. Second, the deformations in this study were performed statically rather than dynamically. We are currently in the process of building a dynamic motion system for more lifelike compression and expansion of the phantom, and more comprehensive analysis of its motion characteristics. Finally, the HU range of the printed phantoms is currently limited in range to -840 to -48 HU, so features with higher or lower attenuations are clipped. Future iterations may explore different printing materials or techniques to increase the HU range.

Acknowledgment

We acknowledge support through the National Institutes of Health (R01EB035092). The authors thank Dr. Boon-Keng Kevin Teo for selecting the reference patient 4DCT for this study and Magdalena Bazalova-Carter for fruitful and inspiring discussions.

References

[1] S. Gholampourkashi, J. E. Cygler, B. Lavigne, and E. Heath, "Development of a deformable phantom for experimental verification of 4D Monte Carlo simulations in a deforming anatomy," *Phys Med,* vol. 51, pp. 81-90, Jul 2018, doi: 10.1016/j.ejmp.2018.05.012.

[2] H. Zhong, J. Adams, C. Glide-Hurst, H. Zhang, H. Li, and I. J. Chetty, "Development of a deformable dosimetric phantom to verify dose accumulation algorithms for adaptive radiotherapy," *J Med Phys,* vol. 41, no. 2, pp. 106-14, Apr-Jun 2016, doi: 10.4103/0971-6203.181641.

[3] S. Ehrbar *et al.*, "Respiratory motion-management in stereotactic body radiation therapy for lung cancer - A dosimetric comparison in an anthropomorphic lung phantom (LuCa)," (in English), *Radiother Oncol,* vol. 121, no. 2, pp. 328-334, Nov 2016, doi: 10.1016/j.radonc.2016.10.011.

[4] M. Ranjbar, P. Sabouri, C. Repetto, and A. Sawant, "A novel deformable lung phantom with programably variable external and internal correlation," *Med Phys,* vol. 46, no. 5, pp. 1995-2005, May 2019, doi: 10.1002/mp.13507.

[5] 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.

[6] Y. Cheung and A. Sawant, "An externally and internally deformable, programmable lung motion phantom," *Med Phys,* vol. 42, no. 5, pp. 2585-93, May 2015, doi: 10.1118/1.4918581.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] J. Medwig, S. Gaede, J. J. Battista, and S. Yartsev, "Effect of lateral target motion on image registration accuracy in CT-guided helical tomotherapy: a phantom study," *J Med Imaging Radiat Oncol,* vol. 54, no. 3, pp. 280-6, Jun 2010, doi: 10.1111/j.1754-9485.2010.02171.x.

[12] 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.

[13] 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: ARTN 025008 10.1088/2057-1976/acb7f7.

[14] E. Colvill *et al.*, "Anthropomorphic phantom for deformable lung and liver CT and MR imaging for radiotherapy," *Phys Med Biol,* vol. 65, no. 7, p. 07NT02, Apr 2 2020, doi: 10.1088/1361-6560/ab7508.

[15] 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.

[16] N. Shapira *et al.*, "PixelPrint: Three-dimensional printing of realistic patient-specific lung phantoms for CT imaging," *Proc SPIE Int Soc Opt Eng,* vol. 12031, Feb-Mar 2022, doi: 10.1117/12.2611805.

[17] K. Mei *et al.*, "PixelPrint: Three-dimensional printing of patient-specific soft tissue and bone phantoms for CT," *Proc SPIE Int Soc Opt Eng,* vol. 12304, Jun 2022, doi: 10.1117/12.2647008.

[18] K. Mei *et al.*, "PixelPrint: A collection of three-dimensional printed CT phantoms of different respiratory diseases," *Proc SPIE Int Soc Opt Eng,* vol. 12463, Feb 2023, doi: 10.1117/12.2654343.

[19] 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.

[20] 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.

[21] J. Hofmanninger, F. Prayer, J. Pan, S. Röhrich, H. Prosch, and G. Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental,* vol. 4, no. 1, 2020-12-01 2020, doi: 10.1186/s41747-020-00173-2.

[22] H. H. Liu *et al.*, "Assessing respiration-induced tumor motion and internal target volume using four-dimensional computed tomography for radiotherapy of lung cancer," *Int J Radiat Oncol Biol Phys,* vol. 68, no. 2, pp. 531-40, Jun 1 2007, doi: 10.1016/j.ijrobp.2006.12.066.

[23] 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.

Long-term quantitative stability of a first-generation dual-source photon-counting CT

Leening P. Liu[1,2], Pouyan Pasyar[1,2], Olivia F. Sandvold[1,2], Pooyan Sahbaee[3], Harold I. Litt[1], Peter B. Noël[1]

Abstract— **The introduction of the first clinical photon-counting CT (PCCT) presents an opportunity to improve and expand quantitative imaging to new applications with its high spatial resolution and stellar quantitative capabilities. Despite this potential, PCCT employs a photon-counting detector that introduces unknowns including temporal stability that is critical to separating biological changes from scanner changes and variation in longitudinal studies. For the purpose of determining the temporal stability of a first-generation dual-source PCCT, a phantom was subjected to near-weekly scans across a two-year period, in both single-source and dual-source modes. Virtual monoenergetic images (VMI) at 40, 70, 100, and 190 keV and iodine density maps were analyzed to determine changes in relative error and noise both related and unrelated to software/hardware changes. VMIs demonstrated improvements in quantification for dual-source mode associated with software and hardware updates but otherwise illustrated invariance with variation ranging from 0.03 to 0.08%. VMI noise similarly exhibited stability between and with major scanner updates with a maximum change of 4 HU. Iodine density maps also displayed stability between scanner updates with variation up to 0.1 mg/mL but significant improvements in quantification, especially in dual-source mode, that allowed relative error in single-source and dual-source modes to match at -0.04 and -0.02 mg/mL, respectively. Spectral results in PCCT showed temporal stability over time that improved quantification accuracy particularly for dual-source mode. This stability will boost confidence in quantitative metrics such as in longitudinal studies and thus facilitate more clinical applications that may change the workflow of diagnostic radiology.**

Index Terms— **Photon-counting CT, Spectral CT**

I. Introduction

COMPUTED tomography (CT) is an important imaging tool for the diagnosis and screening of diseases, which may require longitudinal evaluation to characterize disease progression. Imaging is also critical to assessment of the response to treatment. Both disease progression and response to therapy can be characterized with quantitative imaging, which implements quantitative metrics to establish changes in disease states. With the addition of dual-energy CT (DECT), the first realization of the spectral CT, quantitative imaging has flourished as a result of a variety of quantitative maps, or spectral results [1]. These spectral results, compared to conventional CT images, diminish the effect of acquisition parameters, such as tube voltage and patient size, on quantification [2], enabling the universal use of quantitative metrics across the patient population for different diseases. They include virtual monoenergetic images (VMI) that represent the attenuation at a specific energy and iodine density maps that isolate iodinated contrast to provide information about flow of blood in the vasculature and into/out of lesions [3-5]. Despite the quantitative capabilities of DECT, there remain challenges that are currently being addressed with the introduction of the latest photon-counting detectors.

Photon-counting CT, the latest realization of spectral CT, improves upon DECT and similarly facilitates quantitative imaging while also expanding clinical application of quantitative imaging with its superior quantification capabilities. These capabilities can be attributed to the replacement of the typical energy-integrating detector in DECT with an energy-discriminating detector [6]. The new detector technology enables increased spatial resolution and reduced electronic noise that enhances PCCT's low-dose imaging capabilities for a range of patient sizes. Since its introduction into clinical practice, many clinical applications have taken advantage of both PCCT's high spatial resolution and quantitative capabilities across organ systems from cardiovascular [7-10] to abdominal imaging [11-13]. With both the adaptation of existing applications to PCCT and development of new applications, quantitative imaging continues to play a critical role in clinical decision making.

The increasing emphasis and utilization of quantitative imaging in the clinical workflow highlight the importance of quantitative accuracy and temporal stability, particularly for longitudinal applications. Quantitative accuracy establishes the degree of confidence in disease characterization while temporal stability ensures differentiation of biological changes over time from scanner-related changes and variation. For PCCT, quantitative accuracy has already been characterized with a variety of acquisition parameters for different materials [14]. Temporal stability, on the other hand, has yet to be described. Although temporal stability has been evaluated in CTs with energy-integrating detectors [15], photon-counting detectors impart distinct challenges, such as temperature management, that may affect its temporal stability. Thus, in this study, we

This study was funded by the National Institutes of Health (R01EB030494) and Siemens Healthineers.

L.P. Liu. P. Pasyar, and O.F. Sandvold is with the Department of Bioengineering and Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. H.I. Litt and P.B. Noël are with the Department of Radiology, Perelman School of Medicine, Philadelphia, PA, USA. P. Sahbaee are with Siemens Healthineers, Malvern, PA, USA.

assess the temporal stability of a first-generation dual-source PCCT (NAEOTOM Alpha, Siemens Healthineers) across two years in both VMIs and iodine density maps.

II. Methods

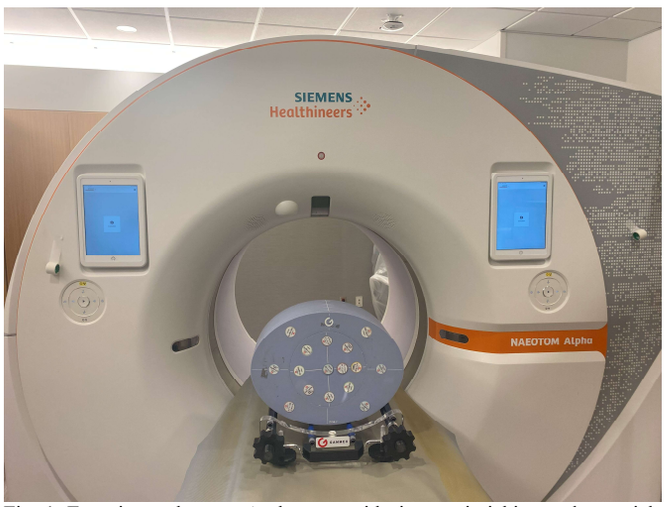


Fig. 1. Experimental setup. A phantom with tissue-mimicking and material-specific inserts was scanned approximately weekly over a period of two years.

A phantom with tissue and material specific inserts was utilized to evaluate the temporal stability of spectral results (**Fig. 1**). The inserts included adipose, brain, blood 70, blood 100, blood + 2 mg/mL iodine, blood + 4 mg/mL iodine, calcium 50 mg/mL, calcium 100 mg/mL, calcium 300 mg/mL, iodine 2 mg/mL, iodine 5 mg/mL, and iodine 10 mg/mL Using the manufacturer provided elemental composition and physical density, the expected attenuation at 40, 70, 100, and 190 keV were calculated by using the elemental and energy-specific attenuation from NIST. Additionally, a simple material decomposition was implemented to determine the expected iodine density for iodine-containing inserts. These expected values were treated as the ground truth for calculating relative error.

For 104 weeks from November 23, 2021 to November 23, 2023, the phantom was scanned on a first-generation dual-source PCCT approximately every week except during holidays and service events. Specifically, software updates occurred at weeks 8, 35, and 69, while hardware changes were at weeks 8, 18, and 80. Every week three scans were performed in single-source and dual-source mode for a total of 6 scans. All scans utilized a tube voltage of 120 kVp and a volumetric CT index ($CTDI_{vol}$) of 10 mGy. For each scan, VMI 70 keV were reconstructed at both iterative reconstruction levels (IR) of QIR 0 and 2. All other acquisition and reconstruction parameters are in **Table I**.

TABLE I
Acquisition And Reconstruction Parameters

Tube voltage [kVp]	120
Exposure time [s]	0.25
Spiral pitch factor	1
Exposure [mAs]	127
$CTDI_{vol}$ [mGy]	10
Collimation width	144 x 0.4 mm
Slice thickness [mm]	3
Convolution kernel	Qr48
QIR	0, 2
Field of view [mm]	450
Matrix size	512x512
Pixel spacing [mm]	0.88

Regions of interest were then placed on ten consecutive central slices and mean and standard deviation (noise) measured. The values were then averaged across the 10 slices from each of the three repetitions for every week (30 slices each). The percent relative error was then computed for VMI 70 keV. These relative errors in addition to noise were then plotted over time to visualize variation between software and hardware changes and the quantitative changes associated with software and hardware updates. To quantify variation, the standard deviation of the relative error between scanner changes was calculated for each individual insert and then averaged across inserts. Additionally, both relative error and noise was averaged across rods between each scanner modification. A change point search algorithm, pruned exact linear time algorithm, was also applied to the relative error and noise to identify timepoints where quantification changed for each combination of spectral result, source type, and IR level.

III. Results

VMI 70 keV quantification remained invariant between software and hardware updates, with certain updates improving the accuracy. VMI 70 keV demonstrated variation ranging from 0.03 to 0.07 and from 0.04 to 0.08% for single-source and dual-

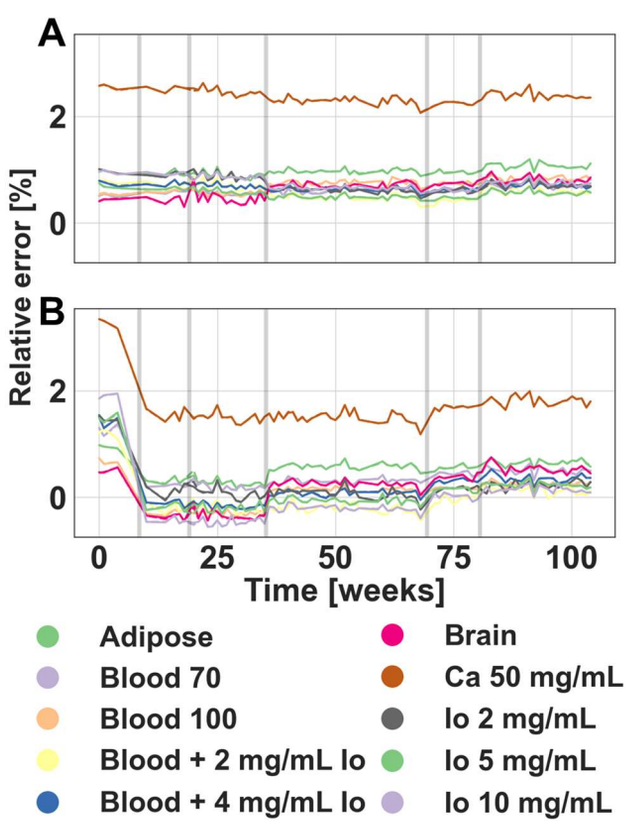


Fig. 2. VMI 70 keV difference relative to expected values for a range of inserts. VMI 70 keV in single-source mode (A) demonstrated stability between and across scanner changes represented by gray vertical bars. In dual-source mode (B), quantification was similarly stable between scanner changes with distinct improvements in quantification at weeks 8 and 68.

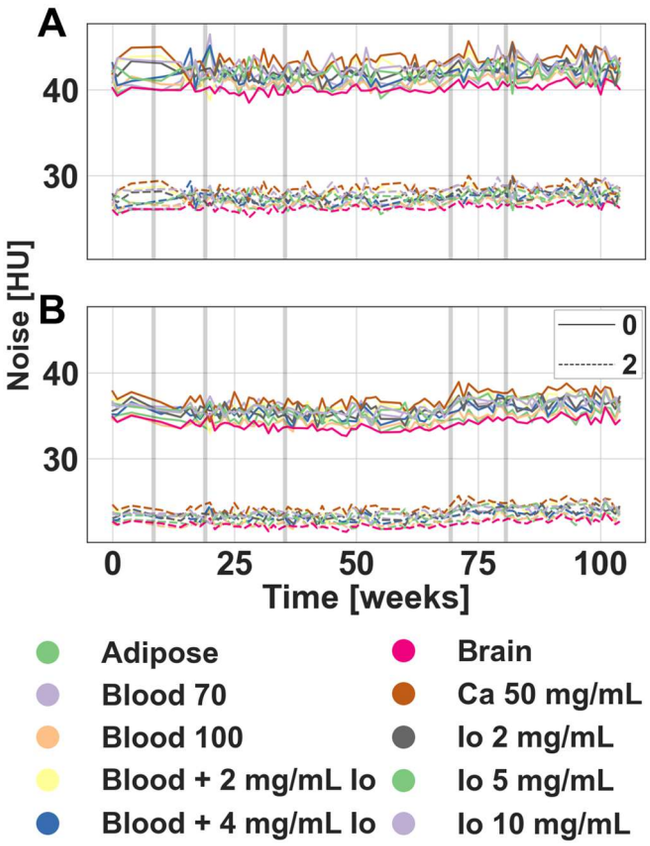


Fig. 3. Stability of VMI 70 keV noise over time for different inserts. In both single-source (A) and dual-source modes (B), noise was stable with minimal quantitative changes associated with scanner changes for two different levels of IR.

source modes, respectively, for the different periods between scanner updates (**Fig. 2**). This variation corresponded to approximately 0.8 HU. However, VMI 70 keV in single-source mode was more stable than in dual-source mode. Additionally, the change point with the most quantitative change occurred at week 8 with average relative error decreasing from 1.4 to 0.1% in VMI 70 keV. While at week 8 a majority of inserts (8/10) demonstrated a significant change in VMI 70 keV in dual-source mode, other change points include weeks 35 and 69, each for 2/10 inserts, suggesting minor change points.

Noise comparatively demonstrated stability with minimal changes resulting from scanner updates. Noise in VMI 70 keV averaged 41.9 ± 1.0, 27.4 ± 0.7, 35.5 ± 0.9, and 23.4 ± 0.6 HU across inserts for single-source + IR level 0, single-source + IR level 2, dual-source + IR level 0, and dual-source + IR level 2, respectively (**Fig. 3**). Additionally, only one major change point was identified at week 69 for dual-source + IR level 0, which amounted to an average change in noise of 1.3 HU.

IV. Discussion

The first-generation dual-source PCCT represents a significant advancement in medical imaging technology. This study aimed to rigorously evaluate the temporal stability of this novel system. By systematically analyzing spectral results over time, we established the consistency of PCCT performance amidst various software and hardware updates. This approach provided a comprehensive understanding of the system's reliability over extended periods.

Unlike its predecessors, conventional CTs and DECTs, which have undergone extensive evaluation in clinical settings, PCCT is a relatively new concept in the field of diagnostic imaging. Its distinctive feature lies in the utilization of photon-counting detectors, a departure from the traditional energy-integrating detectors used in other CT technologies. This transition marks a significant leap, potentially offering enhanced image quality and diagnostic capabilities.

V. Conclusion

The first clinical dual-source PCCT system has illustrated quantitative stability over time, even amid scanner updates. These updates have led to significant improvements in quantification accuracy. This demonstrated temporal stability provides essential insights, ensuring the reliability of quantitative metrics in longitudinal studies.

Acknowledgements

We acknowledge support from National Institutes of Health (R01EB030494) and Siemens Healthineers.

References

[1] M. C. Jacobsen *et al.*, "Multi-energy computed tomography and material quantification: Current barriers and opportunities for advancement," *Med Phys*, vol. 47, no. 8, pp. 3752–3771, Aug. 2020, doi: 10.1002/mp.14241.

[2] A. P. Sauter *et al.*, "Accuracy of iodine quantification in dual-layer spectral CT: Influence of iterative reconstruction, patient habitus and tube parameters," *Eur J Radiol*, vol. 102, pp. 83–88, 2018, doi: 10.1016/j.ejrad.2018.03.009.

[3] S. Oda *et al.*, "Myocardial late iodine enhancement and extracellular volume quantification with dual-layer spectral detector dual-energy cardiac ct," *Radiol Cardiothorac Imaging*, vol. 1, no. 1, p. 180003, Apr. 2019, doi: 10.1148/RYCT.2019180003

[4] D. Deniffel, A. Sauter, J. Dangelmaier, A. Fingerle, E. J. Rummeny, and D. Pfeiffer, "Differentiating intrapulmonary metastases from different primary tumors via quantitative dual-energy CT based iodine concentration and conventional CT attenuation," *Eur J Radiol*, vol. 111, pp. 6–13, Feb. 2019, doi: 10.1016/J.EJRAD.2018.12.015.

[5] L. Wu *et al.*, "Spectral CT Analysis of Solitary Pulmonary Nodules for Differentiating Malignancy from Benignancy: The Value of Iodine Concentration Spatial Distribution Difference," *Biomed Res Int*, vol. 2018, 2018, doi: 10.1155/2018/4830659.

[6] 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.

[7] V. Sandfort, M. Persson, A. Pourmorteza, P. B. Noël, D. Fleischmann, and M. J. Willemink, "Spectral photon-counting CT in cardiovascular imaging," *Journal of Cardiovascular Computed Tomography*, vol. 15, no. 3. Elsevier Inc., pp. 218–225, May 2021. doi: 10.1016/j.jcct.2020.12.005.

[8] T. Emrich *et al.*, "Coronary Computed Tomography Angiography-Based Calcium Scoring: In Vitro and In Vivo Validation of a Novel Virtual Noniodine Reconstruction Algorithm on a Clinical, First-Generation Dual-Source Photon Counting-Detector System," *Invest Radiol*, vol. 57, no. 8, pp. 536–543, Aug. 2022, doi: 10.1097/RLI.0000000000000868.

[9] T. Allmendinger *et al.*, "Photon-Counting Detector CT-Based Vascular Calcium Removal Algorithm: Assessment Using a Cardiac Motion Phantom," *Invest Radiol*, vol. 57, no. 6, pp. 399–405, Jun. 2022, doi: 10.1097/RLI.0000000000000853.

[10] V. Mergen *et al.*, "Extracellular Volume Quantification With Cardiac Late Enhancement Scanning Using Dual-Source Photon-Counting Detector CT," *Invest Radiol*, vol. 57, no. 6, pp. 406–411, Jun. 2022, doi: 10.1097/RLI.0000000000000851.

[11] J. A. Decker *et al.*, "Low-dose CT of the abdomen: Initial experience on a novel photon-counting detector CT and comparison with energy-integrating detector CT," *Eur J Radiol*, vol. 148, p. 110181, Mar. 2022, doi: 10.1016/J.EJRAD.2022.110181.
[12] V. Mergen *et al.*, "Virtual Noncontrast Abdominal Imaging with Photon-counting Detector CT," *Radiology*, vol. 305, no. 1, pp. 107–115, Oct. 2022, doi: 10.1148/RADIOL.213260.
[13] T. Sartoretti, V. Mergen, L. Jungblut, H. Alkadhi, and A. Euler, "Liver Iodine Quantification With Photon-Counting Detector CT: Accuracy in an Abdominal Phantom and Feasibility in Patients," *Acad Radiol*, May 2022, doi: 10.1016/J.ACRA.2022.04.021.
[14] L. P. Liu *et al.*, "First-generation clinical dual-source photon-counting CT: ultra-low-dose quantitative spectral imaging," *Eur Radiol*, vol. 32, no. 12, pp. 8579–8587, Dec. 2022, doi: 10.1007/S00330-022-08933-X.
[15] M. Jacobsen, C. Wood, and D. Cody, "SU-G-206-07: Dual-Energy CT Inter- and Intra-Scanner Variability Within One Make and Model," *Med Phys*, vol. 43, no. 6Part25, pp. 3641–3641, Jun. 2016, doi: 10.1118/1.4956948.

Feasibility of an open-standard, accessible, reproducible modular insert phantom for quantitative imaging

John M. Hoffman and Morgan A. Daly

***Abstract*—This work introduces a design and a method for producing low-cost, high-quality phantom bodies for quantitative imaging. Our phantom design has modular insert regions designed to accept task-specific 3D printed modules. Additionally, we have designed features to enhance the reproducibility of our phantom for quantitative imaging tasks, such as repeatable insert orientation, and phantom alignment and orientation markers. We perform an initial evaluation of three phantom bodies produced with our method and using three simple inserts: two developed from published digital reference objects, and a "texture" insert consisting of regularly spaced cylindrical rods. We found that our phantom bodies cast in epoxy resin were highly uniform and the associated radiomic measurements were well-reproduced between two different sets of modular inserts, suggesting that this approach is viable for the production of phantoms. Future work will explore reproducibility in greater depth and other modular insert designs.**

***Index Terms*—phantoms, 3D printing, quantitative imaging, radiomics**

I. Introduction

PHYSICAL imaging phantoms are widely used for the measurement of image quality and quantitative imaging measurements in computed tomography (CT). The most commonly used phantoms are designed for CT quality assessment and usually consist of uniform materials with minimal texture (i.e., local variation in pixel intensity). However, interest in quantitative imaging, including radiomics and machine learning, has driven demand for phantoms with varying texture patterns. These "texture phantoms" or "radiomics phantoms" typically fall into one of two categories: phantoms that consist of hollow vessels filled with available materials that vary in texture (such as wood, cork, foams, rice, or nuts), or custom-designed commercially manufactured phantoms (sometimes still involving filled inserts) [1]–[3].

While these types of phantoms are effective, no single design is accepted as a standard radiomics phantom, nor are they readily accessible (in the sense that even manufactured radiomics phantoms are often one-of-a-kind). Textures made from granular materials, such as rice or metal shavings, are also inherently one-of-a-kind. Alternatively, 3D printing offers a low-cost, repeatable, and relatively accessible method for similar structures with significantly greater reproducibility.

To date, the focus on 3D printed phantoms in medical physics has been on the direct printing of the phantom (e.g., [4], [5]). This unfortunately is often impractical for anything other than a professional quality 3D printer since phantom bodies present a relatively large and dense object. Such objects require significant time and material to print accurately. Large objects also often pose printing challenges (e.g., bed separation, curling, etc.), and the required longer prints have a higher chance of failure.

We propose that quantitative imaging research would benefit from the development of a phantom with the following properties:

1) **Open-standard:** a standard, freely available for use, modification, and redistribution
2) **Accessible:** readily obtainable from manufacturers at low relative cost, or producible using easily obtained materials and tools
3) **Reproducible:** able to be replicated with little variation, particularly with respect to quantitative imaging feature values of interest

This work investigates the feasibility of producing such a phantom. The phantom proposed could ostensibly be produced for a relatively low cost compared to commercial offerings, with minimal specialized equipment or expertise. Additionally, once the phantom body is obtained, 3D printed modules are easily, quickly, and affordably produced on consumer-grade 3D printers, or via university 3D printing labs or 3D printing services. Furthermore, the 3D design files for modular inserts can be readily shared between researchers if a shared phantom body design is utilized. Such an approach could facilitate collaboration between research groups and potentially improve the reproducibility of some phantom-based studies.

II. Methods

A. Materials for 3D printing

The 3D printing was done using a Bambu Labs Carbon X1 printer with a 0.4 mm printer nozzle. Three printer filament materials were tested: PLA (polylactic acid), PET-G (polyethylene terephthalate glycol), and ABS (acrylonitrile butadiene styrene). All printer filaments were acquired from Bambu Labs, Shenzhen, Guangdong, China. Slicing of the 3D models for printing was done using Bambu Studio version 1.8.2.56. All design work was done using Fusion 360 (version 2.0.17954 arm64) from AutoDesk on a Macbook Air M2.

B. Phantom body mold development

An approach for 3D printing molds for the casting of epoxy phantoms that would be accessible for the average CT

JM Hoffman and MA Daly are with the Department of Radiology, University of California, Los Angeles, Los Angeles, CA, 90024 USA
e-mail: jmhoffman@mednet.ucla.edu

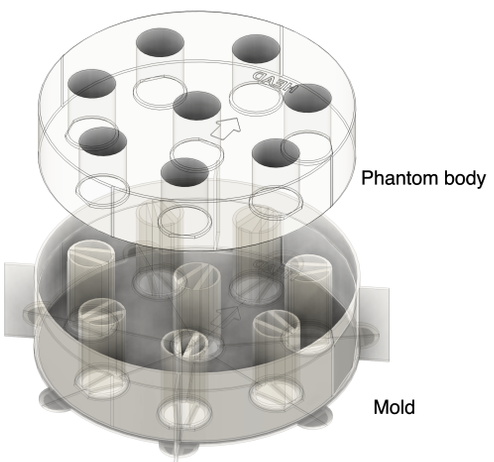


Fig. 1. Rendering of the base phantom design, and corresponding mold. Phantom body is designed to be 200.0mm in diameter, and 40mm in depth. Modular inserts regions are 30mm in diameter. The phantom design can be customized in Fusion 360, and the mold will be automatically recalculated for the new phantom.

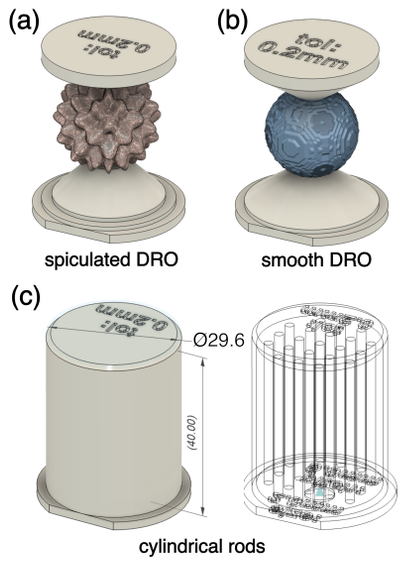


Fig. 2. Rendering of the 3D printed modular inserts. (a) Stanford digital reference object (DRO) with spiculated surface; (b) DRO with smooth surface; (c) cylindrical rod modular insert, shaded view (left) and wireframe (right) to show interior structure.

physics researcher was developed over the course of several iterations. The proposed design is single-use and separates from the phantom relatively easily using basic tools (utility or other razor knife, screwdriver, pliers). The unique challenges presented by the insert molds were overcome by leveraging the deformable nature of plastic to "crush" the mold insert away from the phantom. The final mold and phantom designs are shown in Figure 1.

The molds were printed at a layer height of 0.28 mm and an infill of 100%. No other special settings were used. While all three printer filament materials were tested, ABS was the only material capable of being safely and reliably demolded from the phantom body.

C. Phantom casting and demolding

The phantom bodies were cast using "Deep Pour" epoxy resin (which cures slowly and is suitable for large, solid objects) obtained from Amazon (Teexpert, China). A generic vacuum chamber and pump, also obtained from Amazon, were used to remove air bubbles introduced into the liquid resin during mixing. The liquid resin was placed in chamber after initial mixing and held under vacuum for 20-30 minutes until minimal additional bubbles were observed.

The phantom bodies were demolded after 24 hours and allowed to cure an additional 48 hours to final hardness. Any sharp edges were sanded down or trimmed off. No additional post-processing was performed.

D. Modular insert development

We developed an initial set of three modular inserts to investigate the repeatability of the 3D printing process for texture modules. Two inserts were based on digital reference objects (DROs) from the Stanford DRO Tookit [6] that vary in z, one with a smooth surface and one with a spiculated surface. The third insert contained uniformly sized and spaced cylindrical rods. Renderings of the three modules are shown in Figure 2.

The modules were sliced with the same process used for the mold (including 100% infill) except for the layer height which was reduced to 0.2 mm. All inserts were printed using PLA.

E. Analysis of reproducibility

For the evaluation of the process presented here, we manufactured three phantom bodies and printed two sets of the modular inserts. Table I provides a summary of any noteworthy characteristics of the phantoms.

The phantom bodies were scanned on a Siemens Definition AS 64 (Siemens Healthineers, Forcheim, Germany) at our institution. We utilized our institution's routine chest protocol with tube current modulation disabled (120 kVp, 180 eff. mAs).

Two sets of seven ROIs (equal to the number of modular insert regions) were placed at 40 mm ("inner") and 80 mm ("outer") radii from phantom center. 2D ROIs were placed such that they only contained phantom body material. Mean and standard deviation were calculated for each. Inter-phantom variability was considered by comparing the uniformity measurements between phantoms.

External physical dimensions were measured on CT using ROI annotations, and the module insert tolerances were measured using a 3D-printed set of tolerance calibration modules.

III. Results

A. Dimensional accuracy

Dimensional accuracy results are given in Table II. Overall agreement between the three phantom bodies is shown in CT images in Figure 3. We observed high dimensional accuracy and good modular insert tolerance values.

TABLE I
PHANTOMS USED IN OUR MEASUREMENTS. THE MOLD LID WAS FOUND TO BE CAUSING SEPARATION/AIR BUBBLE ISSUES AND WAS EVENTUALLY REMOVED FROM THE MOLD DESIGN. FINAL DESIGN IS PHANTOM BODY 1, BOLDED.

Phantom Body	Mold Material	Mold Lid	Notes
1	**ABS**	**no**	**Most up-to-date mold (no lid); no remarkable manufacturing defects.**
2	ABS	yes	Mold separation around edge of phantom.
3	PET-G	yes	Mold separation around edge of phantom; small epoxy break around middle insert occurred during demolding.

TABLE II
PHANTOM PHYSICAL DIMENSIONS. PERPENDICULAR OUTER DIAMETERS SHOULD BE 200.0 MM. MODULE TOLERANCE IS IDEALLY 0.0 MM. "Ø" IS USED TO INDICATE DIAMETER. ALL UNITS ARE IN MILLIMETERS.

Phantom	Ø (horizontal)	Ø (vertical)	module tolerance
1	199.8	199.3	0.2
2	200.7	199.4	0.2
3	199.2	200.0	0.15

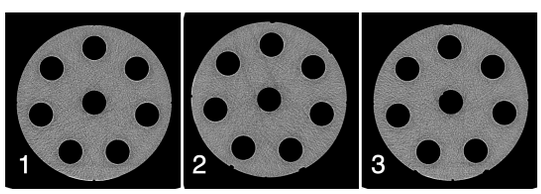


Fig. 3. The three manufactured phantom bodies display good agreement and uniformity in CT images. Shown with window width 200HU and window level 100HU. Numbers correspond to phantom body described in Table I.

B. Uniformity

All phantoms were observed to have high intra-phantom uniformity. Although only three phantoms were manufactured for this experiment, the inter-phantom uniformity was observed to be high and consistent with the intra-phantom results. See Table III for details.

TABLE III
UNIFORMITY MEASUREMENTS FOR THE TESTED PHANTOM BODIES. MEAN VALUES WERE CALCULATED FOR EACH ROI, FROM WHICH THE AGGREGATE MEAN, STANDARD DEVIATION, AND MIN/MAX VALUES WERE CALCULATED. ALL UNITS ARE HU.

Phantom	mean	std	min	max
1 (inner)	106.06	0.52	105.11	106.83
1 (outer)	105.66	1.05	104.50	107.39
2 (inner)	106.55	0.55	105.94	107.55
2 (outer)	106.89	1.15	105.71	108.74
3 (inner)	107.30	0.81	106.42	108.76
3 (outer)	107.88	0.87	106.43	108.80

C. Reproducibility of the modular inserts

Two sets of the modular inserts were printed for an initial investigation into their reproducibility. Qualitative agreement between the sets in CT images was very good, as shown in Figure 4. The design of the phantom ensures that module orientation is repeatable. A slight difference can be seen in the spiculated DRO between the two images, however, this is due to imperfect axial alignment of the selected slices and the insert's heterogeneity in z (shown in Figure 2a) rather than differences between the physical objects themselves.

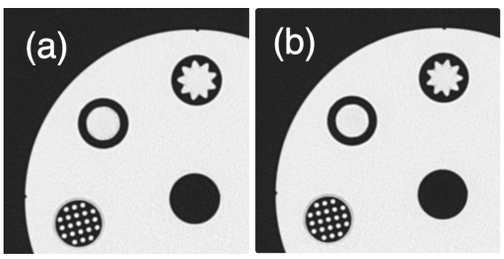


Fig. 4. Agreement in CT images of (a) insert set 1 and (b) insert set 2, both within phantom body 1. Module orientation is repeatable in our phantom design, which can be seen from the circular rod insert (lower left). Images shown with window width 1400HU and level -500HU.

The reproducibility of quantitative imaging feature values was evaluated by extracting a selection of six features from a 2D ROI segmentation generated for each insert. These results are shown in Figure 5. Overall, there appears to be a high degree of agreement in feature values between the two sets of inserts. The largest apparent difference is seen between the two smooth DRO inserts; upon further investigation, this may be due to a minor defect in the 3D printing process visible in one of the smooth DRO inserts.

IV. DISCUSSION

At this time, the only viable printing material tested for the mold was ABS. This presents some challenges in that ABS is slightly less accessible to 3D print safely for users (requiring an enclosure and good ventilation due to unhealthy fumes). However, newer printers such as the Bambu X1 Carbon used in this experiment make ABS much more accessible than previously possible.

An alternative to using your own 3D printer is to utilize local printing lab resources often readily and affordably available at many major universities, or online services that will print and ship the part. This can be a good option if a suitable 3D printer is unavailable, or ABS is not printable.

More important than the molding of the phantom body is the ability to print the custom modular inserts. For this task, we recommend PLA, which is fortunately one of the *easiest* materials to 3D print (and readily accessible on entry-level 3D printers). Other work [7] has evaluated the x-ray attenuation properties of PLA. Although non-conclusive, we observed good consistency between our two samples. In the future, we intend to explore consistency between modules printed from different sources (printers, filament providers, etc.)

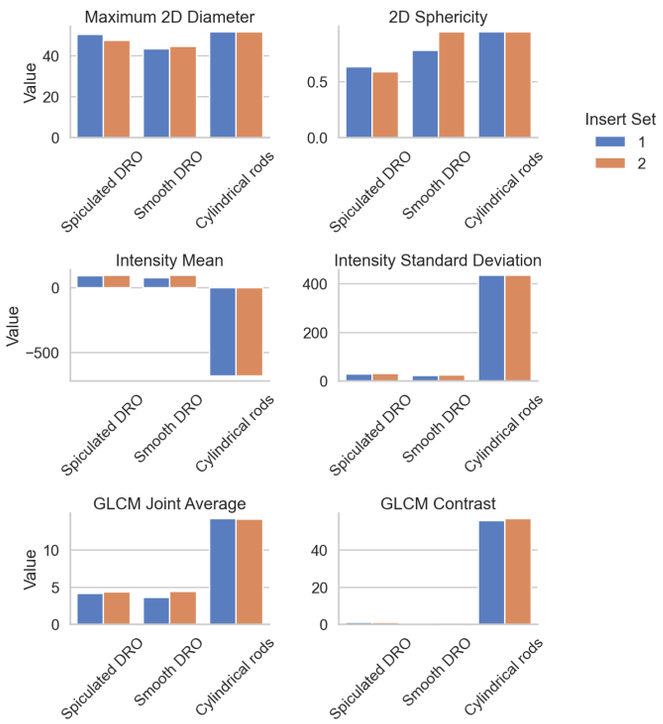


Fig. 5. Comparison of selected quantitative imaging feature values between the two sets of 3D printed modular inserts. We observe minimal difference in feature values between the two insert sets (shown in blue and orange).

Although we observed several small manufacturing defects (due to the use of the lid in earlier mold designs), we did not observe any systematic impact on our other measurements near the site of the defects. This is encouraging since the "do-it-yourself" approach recommended by this paper may raise some concerns around the achievable consistency of phantoms made with the proposed process. We have found the initial achievable consistency, with respect to physical dimensions and radiologic properties, to be extremely encouraging.

We did employ a vacuum chamber and pump, which are unlikely to be equipment most CT researchers will have readily available. Use of the vacuum system eliminates some of the variables while mixing resin, however is possible that the degassing process we employed is not fully required to produce a consistent, uniform phantom. This will likely require additional care to limit air bubbles when mixing two-part resins.

V. Conclusions

We see the critical long-term contribution of this work as the potential library of possible printable modular inserts for cutting-edge research topics. The easily shareable, affordable, user-makeable phantom body discussed here is merely a stepping stone to access these possible libraries or collections or to build custom measurement tools that can be shared with colleagues. Specifically, we hope to help develop modular inserts for areas that are not currently well-served by existing phantom manufacturers (e.g., texture) can easily be standardized, shared, and evaluated. Eventually, we hope to be able to provide disease-specific printed inserts. Cutting-edge printing technology such as Pixel Print could make this possible [7].

This process was undertaken to develop a method that would be repeatable by others. Although access to 3D printing, and comfort level with basic hand tools should play a role in the decision to undertake this project, we do feel that utilizing our designs and following instructions that we will provide should make this achievable for most researchers. Furthermore, we highly encourage the exploration of other manufacturing methods using our designs and design methods (such as CNC milling, casting with other materials, etc.)

All 3D designs presented in this work will be made publicly available including the mold, modular inserts, and any other details required for reproduction of this work, including instructions and eventually a tutorial video illustrating the complete process.

References

[1] D. Mackin, X. Fave, L. Zhang, *et al.*, "Measuring CT scanner variability of radiomics features," *Investigative radiology*, vol. 50, no. 11, pp. 757–765, Nov. 2015, ISSN: 0020-9996. DOI: 10.1097/RLI.0000000000000180.

[2] B. A. Varghese, D. Hwang, S. Y. Cen, *et al.*, "Identification of robust and reproducible CT-texture metrics using a customized 3D-printed texture phantom," *Journal of Applied Clinical Medical Physics*, vol. 22, no. 2, pp. 98–107, Feb. 2021, ISSN: 1526-9914, 1526-9914. DOI: 10.1002/acm2.13162.

[3] M. Daly, J. M. Hoffman, A. M. Hernandez, *et al.*, "Harmonizing quantitative imaging feature values in CT using image quality metrics as a basis," in *SPIE Medical Imaging 2024: Physics of Medical Imaging*, 2024.

[4] J. Solomon, A. Ba, F. Bochud, and E. Samei, "Comparison of low-contrast detectability between two CT reconstruction algorithms using voxel-based 3D printed textured phantoms," *Medical Physics*, vol. 43, no. 12, pp. 6497–6506, Dec. 2016, ISSN: 0094-2405, 2473-4209. DOI: 10.1118/1.4967478.

[5] E. D. Ehler, B. M. Barney, P. D. Higgins, and K. E. Dusenbery, "Patient specific 3D printed phantom for IMRT quality assurance," *Physics in Medicine and Biology*, vol. 59, no. 19, pp. 5763–5773, Oct. 2014, ISSN: 0031-9155, 1361-6560. DOI: 10.1088/0031-9155/59/19/5763.

[6] A. Jaggi, S. A. Mattonen, M. McNitt-Gray, and S. Napel, "Stanford DRO Toolkit: Digital Reference Objects for Standardization of Radiomic Features," *Tomography*, vol. 6, no. 2, pp. 111–117, Jun. 2020, ISSN: 2379-1381. DOI: 10.18383/j.tom.2019.00030.

[7] K. Mei, M. Geagan, L. Roshkovan, *et al.*, "Three-dimensional printing of patient-specific lung phantoms for CT imaging: Emulating lung tissue with accurate attenuation profiles and textures," *Medical Physics*, vol. 49, no. 2, pp. 825–835, Feb. 2022, ISSN: 0094-2405, 2473-4209. DOI: 10.1002/mp.15407.

Organ-Level Noise Estimation in CT Images Using Deep Learning

Wonkyeong Lee, Sen Wang, Maria Jose Medrano, Jennie Cao, Grant Stevens, Justin Ruey Tse, and Adam S. Wang

***Abstract*—We propose an automatic, organ-level noise estimation framework for Image Quality Assessment (IQA) in Computed Tomography (CT) images, designed to accurately characterize the distinctive and non-stationary noise levels within CT images, without the need for reference images. This innovative IQA framework combines an accurate noise model with an open-source segmentation program to calculate the average noise level for each organ. This methodology allows for the automatic estimation of organ-specific noise, independent of anatomical structures, across various dose levels, with a RMSE of 1.09 HU. Rigorous evaluation of the framework across diverse anatomical organs demonstrates its robustness, marking a substantial advancement in IQA for CT imaging. This development holds the potential to facilitate radiation dose reduction while maintaining acceptable diagnostic quality.**

***Index Terms*—Image quality assessment (IQA), organ-level IQA, noise estimation, Computed Tomography.**

I. Introduction

ONE of the notable advantages of computed tomography (CT) is its proficiency in delineating structures with minimal contrast within a subject. This capability is predominantly constrained by the presence of noise, closely linked to the radiation dose. That is, an increase in the radiation dose during image acquisition corresponds to a reduction in image noise, thereby enhancing the perception of low-contrast structures [1]. Therefore, optimizing the balance between radiation dose and image quality is crucial in CT imaging. While higher radiation doses can produce images of sufficient diagnostic quality with less noise, there is a concurrent risk as ionizing radiation can potentially induce radiation-induced malignancy [2]. To address this challenge, numerous methods have been developed for image quality assessment (IQA) in CT, aimed at capturing the diagnostic efficacy of images while optimizing the tradeoff between image quality and radiation dose.

However, a limitation of these methods is that they focus on slice-level IQA, whereas in actual clinical practice, there is also a need for organ-level IQA. Organ-level IQA can potentially reduce radiation dose to the patient by focusing radiation dose to the region of interest and avoiding excessive radiation dose to radiosensitive organs [3]. For instance, Boss et. al. [4] applied an organ-specific dose-reduction algorithm into pediatric chest CT imaging and reported that this algorithm does not adversely affect image quality, and therefore, it can be employed in clinical settings to diminish radiation exposure without compromising the diagnostic integrity of the images. However, organ-level IQA presents substantial challenges due to the difficulty of estimating noise free from anatomical structures without clean reference images of the same CT scan. Additionally, the noise in CT images is non-stationary and influenced by various factors, including patient anatomy, quantum noise, and electronic noise [5].

To address these limitations, we propose a novel IQA framework for organ-level noise estimation. This framework processes a CT image of any arbitrary dose level as input and employs a deep learning model to generate an accurate standard deviation (stdev) map representing the noise intensity for each pixel without requiring any reference image. Simultaneously, a segmentation map for each organ is generated using an open-source segmentation program. These two maps are then combined, facilitating the noise estimation for each organ. To the best of our knowledge, this is the first IQA framework that automatically calculates organ-specific noise by directly combining a stdev map, generated through a deep learning model, with an organ segmentation map.

II. Method

We present a novel IQA framework for organ-level noise calculation. As depicted in Figure 1, the framework is composed of two parts: stdev map generation and organ noise estimation. The framework processes a 2D CT image as input and employs a deep learning model to generate a stdev map, which is devoid of the anatomic features present in the CT images. Simultaneously, a segmentation map for seven different organ including lungs, kidneys, liver, bladder, spleen, pancreas, and stomach is produced using an open-source segmentation program TotalSegmentator v2 [6]. These two maps are subsequently combined, enabling noise estimation for each organ through the averaging of pixel values in the stdev map corresponding to each organ. In the following section, the detailed methodologies for each component of the framework will be described.

A. Standard Deviation Map Generation

Noise in CT imaging can be decomposed into quantum and electronic components [7]. Additionally, image quality is substantially influenced by the reconstruction process [8].

W. Lee, S. Wang, M. Medrano, J. Cao, J. Tse, and A. Wang are with the Department of Radiology, Stanford University, Stanford, CA 94305 USA (e-mail: uulee@stanford.edu; senwang@stanford.edu; medranom@stanford.edu; jcao19@stanford.edu; jrtse@stanford.edu; adamwang@stanford.edu).

G. Stevens is with GE HealthCare, Waukesha, WI 53188 USA (email: grant.stevens@gehealthcare.com).

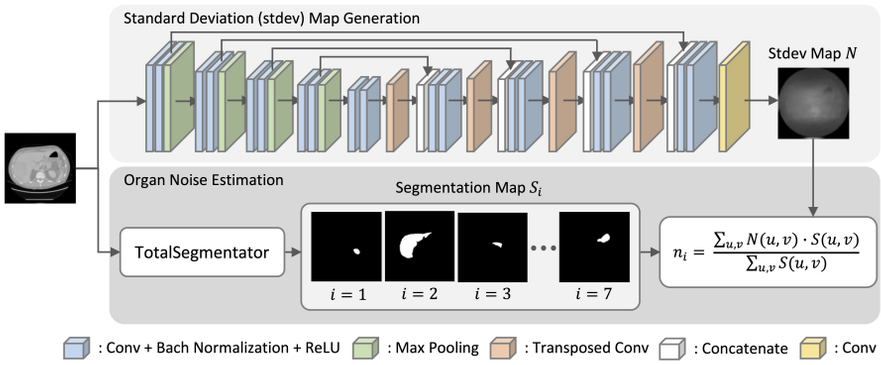


Fig. 1: Overall architecture of the proposed IQA framework, comprising two main components: stdev map generation and organ noise estimation. In the stdev map generation phase, a U-Net model with batch normalization is employed. The framework calculates the noise for the ith organ (n_i).

Moreover, due to the complex underlying anatomical structures, simply calculating the standard deviation cannot accurately extract complete noise information. Therefore, given the nonlinear pattern of noise and the presence of anatomical structures in the background, isolating noise solely from the CT image is challenging without clean reference images. Consequently, a deep learning model is required to learn the nonlinear relationship between the CT image and the noise, devoid of anatomical structures, without the need for reference images. To aid in this process, we generate a reference stdev map that accounts for quantum, electronic, and reconstruction-induced noise, devoid of anatomical structures, using an empirically developed noise model (as detailed in Section III-A). By using a CT image as input and its corresponding stdev map as the reference, a deep learning architecture is trained in a supervised manner. This training equips the model to learn the complex relationship between a CT image and its inherent noise structure, enabling it to generate an accurate stdev map when given an arbitrary CT image.

We employed a U-Net architecture [9] with batch normalization [10] for generating stdev maps, as illustrated in Figure 1. In this architecture, all convolutional layers, except for the final layer, feature a kernel size of 3 and padding of 1. The transposed convolutional layers are configured with a kernel size of 2, a stride of 2, and padding set to 0. The feature channels sequentially expand to {64, 128, 256, 512, 1024} across four encoder blocks and a bottleneck layer, and then reduce to {512, 256, 128, 64} through four decoder blocks. Ultimately, the final convolutional layer, with a kernel size of 1 and padding of 0, reduces the channel count to 1, yielding the final stdev map.

B. Organ Noise Estimation

After acquiring the stdev map, it is necessary to determine the specific location of each organ to calculate organ-specific noise using this map. Consequently, we utilize TotalSegmentator to obtain organ segmentation maps, selected for its ease of use and robust performance in clinical images [6]. Through TotalSegmentator, organ segmentation map S_i are generated, where $i = \{1, 2, 3, 4, 5, 6, 7\}$ correspond to the kidneys, liver, pancreas, bladder, spleen, lungs and stomach. While these organs are the focus of our study, many more are available and can be similarly utilized. The noise for the ith organ n_i is then calculated by multiplying the stdev map N with the segmentation map S_i, and then averaging the pixel values of N within the area segmented by S_i.

$$n_i = \frac{\Sigma_{u,v} N(u,v) \cdot S_i(u,v)}{\Sigma_{u,v} S_i(u,v)} \tag{1}$$

Here, $N(u,v)$ represents the pixel value at position (u,v) in the stdev map N, while $S_i(u,v)$ denotes the pixel value at the same position in the segmentation map S_i.

III. Experimental Setting

A. Dataset

We utilized abdominal CT images from 37 patients, acquired using the Revolution Apex scanner from GE HealthCare, with full-dose settings at varying tube potentials (100, 120, and 140 kVp). To calculate reference noise levels for training our model, scans representing 80% and 20% of the full-dose were generated using a low-dose emulation pipeline [11]. When full-dose images I_f and emulated low-dose images I_e are subtracted, the common underlying structures are removed, leaving the noise inserted to emulate 80% and 20% dose. Stdev maps of these difference images can be scaled to account for the non-linear effect of desired dose level d on image noise [12].

$$\text{Stdev}(d) = \sum_e p_e(d)\sigma(I_e - I_f) \tag{2}$$

$$p_e(d) = \alpha_e d^{-1} + \beta_e d^{-0.5} + \gamma_e \tag{3}$$

Here, σ denotes a standard deviation filter with a window size of 11×11. In equation (3), d^{-1}, $d^{-0.5}$, and γ represent electronic noise, quantum noise, and other reconstruction effects, respectively. The values of e were set to {20%, 80%}. The parameters α, β, and γ are calibrated for these dose e by fitting the model to manually selected uniform ROIs where noise could be directly measured. This method produces reference noise levels with RMSE around 1-2 HU compared to the manually selected ROIs.

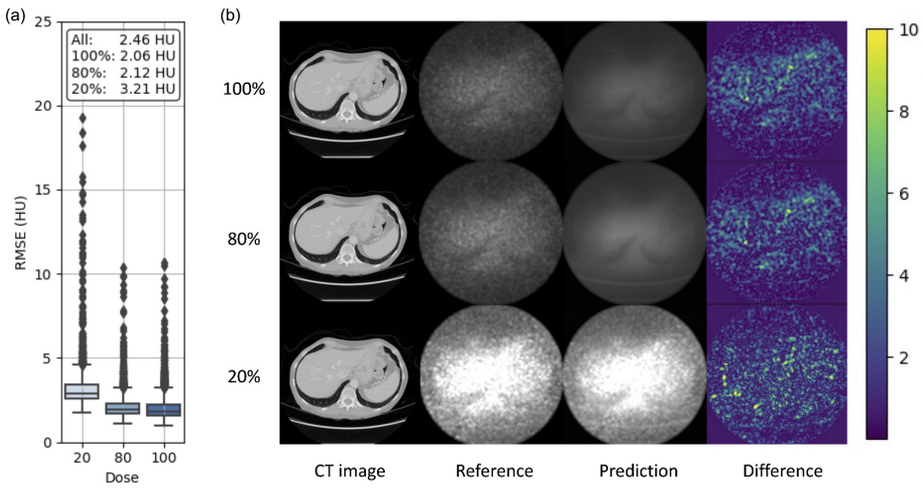


Fig. 2: Visualization of stdev map errors: (a) RMSE values (HU) across full, 80%, and 20% dose levels, with the *All* label summarizing all levels and average RMSE displayed in the plot's text box. (b) CT images, reference and predicted stdev maps, and difference heatmaps for full, 80%, and 20% doses; stdev maps [0-50] HU, difference images [0-10] HU.

B. Model Training

Observing that the model demonstrated more robust performance with CT images at varied dose levels, we included images at full, 80%, and 20% of the full dose for training. The dataset was divided into training, validation, and test sets, consisting of 12,588 images from 20 patients, 2,106 images from 3 patients, and 9,732 images from 14 patients, respectively. Training utilized Mean Squared Error loss and was conducted with a learning rate of 1e-4, following a cosine annealing scheduler ($T_{max} = 50$, and $eta_{min} = 0$) that included warm restarts [13]. The batch size was 8, and training lasted 100 epochs. The final stdev map generation model was selected based on achieving the minimum loss. Data augmentation included horizontal and vertical flips.

IV. Results

A. Performance of Standard Deviation Map Generation

For accurate organ noise calculation, the precision of the initial stdev map generated is paramount. This section evaluates the accuracy of stdev maps produced by the U-Net model. Figure 2a presents the RMSE values in Hounsfield Unit (HU) scale across all dose levels, including full, 80%, and 20% doses, as well as the RMSE for each specific dose level. Overall, the model demonstrates robust performance, with an RMSE of 2.46 HU across all dose levels. However, a dose-dependent increase in RMSE is observed, with RMSE values of 2.06, 2.12, and 3.21 HU for full, 80%, and 20% doses, respectively. Furthermore, while the maximum RMSE values for full and 80% doses remained around 10 HU, the maximum RMSE for the 20% dose exceeded 15 HU. This indicates a relatively higher RMSE for the 20% dose compared to the stdev maps of full and 80% doses.

This trend is also evident in Figure 2b, which displays sample images consisting of the input CT image, the reference stdev map, the model-predicted stdev map, and the difference map obtained by subtracting the predicted stdev map from the reference, all represented in the HU scale. Note that the predicted stdev maps exhibit a smoother texture compared to the reference, which itself is a noisy estimate and contributing to some of the RMSE.

B. Performance of Organ Noise Estimation

Our ultimate objective is to quantify the noise level of each organ using a single numerical value. Therefore, this section evaluates the accuracy of organ noise estimation by comparing the scalar organ noise scores predicted by the framework with the scalar reference scores, which are calculated using the reference stdev map. Consequently, alongside RMSE, we employed Pearson's Linear Correlation Coefficient (PLCC) and Spearman's Rank Order Correlation Coefficient (SROCC) as evaluation metrics. These coefficients assess the extent to which the relationship between the predicted and reference organ noise scores is linear (PLCC) and monotonic (SROCC), respectively. Table I presents the final evaluation metrics calculated between the reference and the predicted noise values for each organ. Overall, the framework achieved an average RMSE of 1.09 HU and correlation scores above 99% for both PLCC and SROCC, demonstrating a strong correlation with the reference organ noise. As detailed in Table I, the RMSE was highest for full-dose images and lowest for 20% doses, in contrast to the trend observed in stdev map prediction accuracy. The correlation coefficients, PLCC and SROCC, exhibited a similar trend: the scores were highest for 20% dose images and lowest for full-dose images. Among the seven organs analyzed, the lungs showed the highest accuracy in terms of RMSE, with scores of 0.90 HU, while the liver and stomach showed the second-highest accuracy, with an RMSE of 0.98 HU. In contrast, the bladder displayed the lowest accuracy with an RMSE of 2.02 HU. However, the framework consistently exhibited high correlation scores across all organs, with the lungs and liver showing the highest SROCC and PLCC values, respectively.

TABLE I: Performance metrics of the predicted organ noise scores. The *All* column represents scores calculated using all images across various dose levels and organs. The numbers in the brackets indicate the number of organs evaluated.

Metric	All (13,593)	100% (4,531)	80% (4,531)	20% (4,531)	Lungs (3,576)	Kidneys (1,968)	Liver (2,607)	Bladder (849)	Spleen (1,383)	Pancreas (1,311)	Stomach (1,899)
RMSE (HU)	1.09	1.20	1.11	**0.96**	**0.90**	1.20	0.98	2.02	1.05	1.29	0.98
PLCC (%)	99.7	97.0	98.1	**99.6**	99.6	99.7	**99.8**	99.3	99.7	99.6	99.7
SROCC (%)	99.2	97.0	98.0	**99.6**	**99.4**	98.3	99.0	98.7	98.5	98.4	98.9

V. Discussion

In this paper, we introduce a novel IQA framework for organ-specific noise estimation, which combines stdev maps with segmentation maps for seven different organs. Through statistical analysis, our framework has demonstrated the capability to accurately calculate the noise level for each organ, effectively disregarding anatomical structures and accounting for the non-stationary characteristics of CT noise.

However, there are limitations that future research could address. Firstly, although the predicted stdev map demonstrates a low RMSE when compared to the reference stdev map, further reduction in per-pixel RMSE could enhance the accuracy of organ noise estimation. For example, while the liver, lungs, and stomach exhibit substantial accuracy in noise scores, the bladder presents a comparatively higher RMSE. This discrepancy may be attributed to the smaller size of the bladder, which encompasses fewer pixels for averaging, as opposed to larger organs such as the liver, lungs, and stomach. Consequently, larger organs benefit from averaging across a greater number of pixels, which can help to offset errors in individual pixel measurements. Thus, generating a more precise stdev map with reduced pixel-level errors could narrow the RMSE gap between the bladder and larger organs. Secondly, the U-Net architecture was employed for stdev map generation. While U-Net is broadly effective across various research domains, its simplicity may not fully capture the complex nature of noise in CT imaging. Exploring advanced architectures like Vision Transformers [14] could potentially enhance performance. Thirdly, the use of RMSE for pixel-level accuracy assessment of stdev maps might not comprehensively reflect the performance of stdev map generation, as minor pixel shifts can result in large errors. Given our ultimate goal of generating a single numeric noise score for organs, where we average noise values within the organ area, pixel-level discrepancies might be less critical if the average noise value closely approximates the reference. Finally, this study focused exclusively on noise. While noise is a critical artifact in CT imaging related to radiation dose, other types of artifacts should also be considered for the IQA model to be broadly applicable in clinical practice.

Consequently, our future research endeavors will address these identified limitations and broaden the spectrum of artifacts under consideration. This expansion aims to enhance the IQA framework's capability to accurately assess organ quality levels in the context of a diverse array of artifacts beyond noise.

VI. Conclusion

In this study, we introduced an innovative IQA framework designed for calculating organ-specific noise levels. This was achieved by integrating a generated stdev map with segmentation maps of seven distinct organs. Our comprehensive experimental analysis confirmed that both the stdev maps produced by our framework and the subsequent noise calculations for each organ closely correspond with the reference values. Looking ahead, we plan to expand our framework's scope to include a variety of artifacts beyond noise, enhancing its applicability for organ-level IQA in broader clinical contexts.

References

[1] L. W. Goldman, "Principles of CT: radiation dose and image quality," *Journal of nuclear medicine technology*, vol. 35, no. 4, pp. 213–225, 2007.

[2] J. M. Albert, "Radiation risk from CT: implications for cancer screening," *American Journal of Roentgenology*, vol. 201, no. 1, pp. W81–W87, 2013.

[3] C. M. Sandino, E. K. Cole, C. Alkan, A. S. Chaudhari, A. M. Loening, D. Hyun, J. Dahl, A. S. Wang, S. S. Vasanawala *et al.*, "Upstream machine learning in radiology," *Radiologic Clinics*, vol. 59, no. 6, pp. 967–985, 2021.

[4] J. Boos, P. Kröpil, D. Klee, P. Heusch, L. Schimmöller, J. Schaper, G. Antoch, and R. S. Lanzman, "Evaluation of the impact of organ-specific dose reduction on image quality in pediatric chest computed tomography," *Pediatric radiology*, vol. 44, pp. 1065–1069, 2014.

[5] T. M. Benson and B. K. B. De Man, "Synthetic CT noise emulation in the raw data domain," in *IEEE Nuclear Science Symposuim & Medical Imaging Conference*, 2010, pp. 3169–3171.

[6] 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, 2023.

[7] A. S. Wang, J. W. Stayman, Y. Otake, S. Vogt, G. Kleinszig, A. J. Khanna, G. L. Gallia, and J. H. Siewerdsen, "Low-dose preview for patient-specific, task-specific technique selection in cone-beam CT," *Medical physics*, vol. 41, no. 7, p. 071915, 2014.

[8] S. E. Divel and N. J. Pelc, "Accurate image domain noise insertion in CT images," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1906–1916, 2020.

[9] 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: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[10] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International conference on machine learning*. pmlr, 2015, pp. 448–456.

[11] T. M. Benson and B. D. Man, "Synthetic ct noise emulation in the raw data domain," *IEEE Nuclear Science Symposuim & Medical Imaging Conference*, pp. 3169–3171, 2010. [Online]. Available: https://api.semanticscholar.org/CorpusID:19674609

[12] A.-A.-Z. Imran, D. Pal, S. Wang, S. Dutta, E. Zucker, and A. Wang, "Personalized CT organ noise estimation from scout images," in *Medical Imaging 2022: Physics of Medical Imaging*, W. Zhao and L. Yu, Eds., vol. 12031, International Society for Optics and Photonics. SPIE, 2022, p. 120310R. [Online]. Available: https://doi.org/10.1117/12.2610986

[13] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.

[14] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale. arxiv 2020," *arXiv preprint arXiv:2010.11929*, 2010.

Image Resolution Analysis Using Line-pair and Disk-based MTF Methods For Deep Learning CT Image Denoising

Md Selim, Prabhat KC, Rongping Zeng
Division of Imaging, Diagnostics and Software Reliability, Office of Science and Engineering Laboratories,
Center for Devices and Radiological Health, Food and Drug Administration.
10903 New Hampshire Ave., Silver Spring, MD 20993

***Abstract*—It is important to characterize the image resolution, often estimated using modulation transfer function (MTF), of CT image reconstruction and denoising algorithms. For nonlinear algorithms that blur low-contrast features while reducing image noise, disk edges of varying contrast have been proposed as a practical method to measure the contrast-dependent MTF. However, it is not clear whether the MTF measured using disk edges would overestimate the underlying image resolution of novel deep learning-based CT image reconstruction or denoising (DLIR/D) methods because the disk edges may be relatively easy to be detected and enhanced by DLIR/D. To investigate this problem, we assemble a library of deep learning (DL)–based CT image denoisers. Then, we compare their cd-MTF performance on disk edges against those measured using line-pair patterns. We also investigated the influence of the surrounding background on the MTF performance by inserting disk and line-pair objects in uniform and patient backgrounds. Our results show that the cd-MTF-based performance remains consistent in the patient and uniform backgrounds. However, the object feature used in the cd-MTF estimation impacts the outcome. In particular, disk-based cd-MTF may overestimate the resolving power of a denoising algorithm.**

***Index Terms*—Deep learning, CT image denoising, Image sharpness.**

I. Introduction

Spatial resolution is a measure of the smallest object that can be resolved by an imaging system, or the minimum distance between two objects such that they can be resolved as separate details. The spatial resolution of a CT system is mainly determined by the hardware system (e.g., focal spot size, detector element size, magnification, etc.). However, resolution is also greatly impacted by the image reconstruction and denoising methods. For example, using a smooth kernel in filtered back projection (FBP)-based reconstruction will lower image noise but at the same time, it may degrade image resolution compared to that from an FBP of a sharp kernel. In CT, the in-plane spatial resolution is often characterized by modulation transfer function (MTF) measured using image quality (IQ) phantoms that contain objects like small beads, thin wire or edges [1]. These objects are of relatively high contrast to minimize the estimation error caused by noise. A high-contrast MTF can fully characterize the resolution property of a linear FBP reconstruction method. However, for advanced nonlinear reconstruction and denoising methods such as model-based iterative reconstruction methods (MBIR) that tend to preserve high-contrast edges while blurring low-contrast features to reduce image noise, their spatial resolution changes with the object contrast. Thus high-contrast MTF alone is insufficient to cover the full spectrum of their spatial resolution property. To overcome this deficiency, a practical method was proposed to estimate MTFs using edges of disk objects of varying contrast in the standard CT IQ phantom to approximate contrast-dependent spatial resolution property [2].

Now deep learning (DL)-based methods have become a new trend in CT image reconstruction and denoising. It is data-driven, nonlinear and shown to have potential to create a visually clean image at low-dose scan. Moreover, it is unclear whether DL-based denoising techniques can yield a spatial resolution to a degree as that obtained from the normal-dose-based FBP reconstruction. As such, it is necessary to revisit how well the disk edge-based MTF methods developed to characterize contrast-dependent image resolution reflect the intrinsic image resolution of the novel type of reconstruction and denoising methods. The concern lies in that the edge of a disk in the uniform IQ phantom may be relatively easy to be detected and enhanced by nonlinear denoising methods such as DL denoisers. MTF measured using disk edges may overestimate the resolution performance on finer structures. To investigate this problem, we assemble a library of DL-based CT image denoising methods, and compare their contrast dependent MTF (cd-MTF) performance measured using disk and line-pair (LP) objects. We consider the MTF estimated using LP objects as the reference MTF in this work since LP objects more closely resemble sinusoidal signals and the LP based MTF should better represent the intrinsic spatial resolution. In addition, we also compare the MTF performance in uniform and patient backgrounds to investigate the impact of background type on the spatial resolution of DL-based denoising methods, similar to what we did previously in [3]. Our work is based on in silico modeling, that is, we simulate the CT images of disk and LP test objects in both uniform and patient backgrounds to facilitate the MTF based image resolution analysis.

Corresponding author: Md Selim, md.selim@fda.hhs.gov

II. Method

A. Deep denoiser models

Four deep learning-based convolutional neural network (CNN) models were selected to perform low-dose CT (LDCT) image denoising: a 3-layer CNN (CNN3), 10-layered U-shaped CNN (Unet), a 17-layer feed-forward CNN (DnCNN), and a 10-layer residual encoder-decoder CNN (REDCNN) [4]. The denoising networks were trained and optimized using paired low-dose and full-dose patient CT images provided in the low-dose CT grand challenge (LDGC) dataset [5]. The training dataset contained paired quarter-dose and full-dose two-dimensional (2D) patches extracted from six patient CT volumes reconstructed with FBP of sharp filter and 3 mm slice thickness. Air patches outside of the body boundary were automatically detected and removed from the training dataset. CNN3 and Unet were optimized with L1 loss function. REDCNN was optimized with the loss function of mean squared errors (MSE) and a total-variation (TV) prior. DnCNN was optimized with the loss function combining a MSE term and an L1 prior. Details of the implementation can be found in our previous publication [4].

B. Phantom design and CT simulation

The test objects for cdMTF measurements included both multiple line pairs (2-9 lp/cm) and small disks (12 mm diameter), representing two distinct feature types. This dual-feature setup allowed an assessment of whether the cdMTF performance of DL denoisers is object-dependent. LP and disk objects were separately inserted and analyzed on two different backgrounds: a uniform phantom and a patient background (Fig. 1). The LP and disk objects were designed with HU contrasts of 340, 150, 100, and 80 relative to the background.

In the uniform water-equivalent background of a 2D circular phantom (200 mm diameter), LP and disk objects were inserted. The disk and LP objects layout followed the ACR CT accreditation phantom module 1 and 4 respectively [6]. The acquisition settings for generating LP images were kept consistent across different contrast levels. Notably, our LP object module included an additional disk-shaped auxiliary object with the same HU as the corresponding LP objects to calculate LP MTFs [7].

To investigate the influence of patient background on cdMTF performance, especially in scenarios where achieving an absolute uniform background is clinically challenging, the test objects were inserted into a patient background (Fig.1e & 1f). Before object insertion, a TV denoiser was applied to the full-dose patient image to approximate a noise-free patient background. Subsequently, a relatively large organ in the patient slice was selected for object insertion. Due to the limited area of the chosen organ, each time only two LP objects were inserted, resulting in four different patient test designs (Fig. 1e contains a LP layout with frequencies of 2 and 3 lp/cm).

DL denoisers were assessed using CT images of uniform and patient background with test objects. The images were generated through a virtual 2D fan-beam CT scanner with geometry matching the Siemens SOMATOM Definition AS+ from which the LDGC data (our training data) were acquired. The sinograms were reconstructed using FBP with a Hann filter, chosen to match the sharp filter in the LDGC patient scans. Poisson noise was simulated at the detector, excluding electronic noise. For cdMTF measurements, noiseless and noisy scans were simulated, and the average cdMTF measure was derived from 20 noisy scans. All reconstructed CT images had a pixel size of 0.66 mm. Subsequently, these FBP-reconstructed CT images underwent further processing by the DL denoisers for performance evaluation.

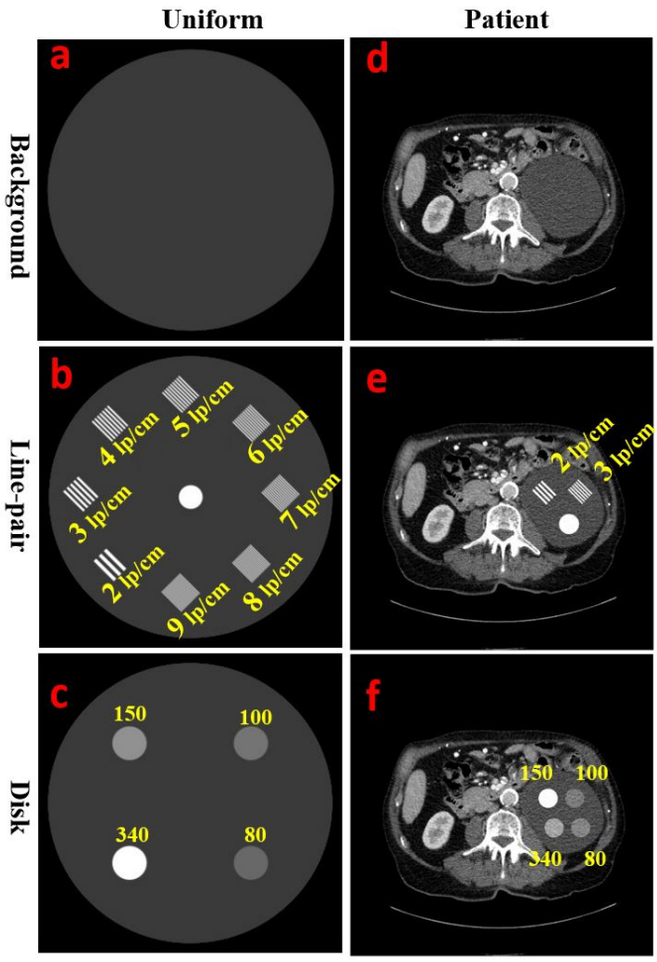


Fig. 1: Test objects and background designs evaluated in this work. a). Uniform water-equivalent phantom background. b). Uniform background with eight LP objects of 340 HU. Similar approach was used to create the LP objects for 80, 100, and 150 HUs. c). Uniform background phantom with four disk objects of 80, 100, 150, and 340 HUs. d). Patient background. e). Patient background with two LP objects and a disk with 340 HU. The disk object is an auxiliary component here to support the cdMTF calculation. Similar approach was used to generate other LP configuration and HU contrast. f). Patient background phantom with four disk objects of 80, 100, 150, and 340 HU.

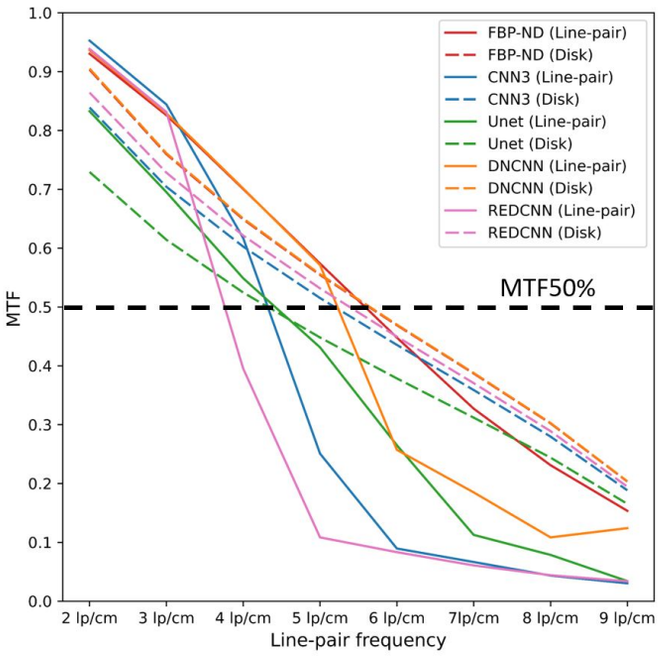


Fig. 2: MTF values from line-pair (solid line) and disk (dashed line) objects. The horizontal dashed line indicates the MTF50% marks for each models including FBP-ND. Here, MTF were estimated for line-pair and disk having HU 340 on uniform background.

C. MTF calculation

Four disks with HU 340, 150, 100, and 80 were used to derive the cdMTF using the edge spread method [2]. The MTF50% value, indicating the frequency at which the normalized MTF decreases by half, was estimated from each MTF curve as a metric for image sharpness. LP cdMTF was calculated using the method outlined in [7]. The method requires two additional ROIs in addition with the LP, one from the uniform background and another from the auxiliary disk area of uniform HU of the corresponding LP, to calculate the signal magnitude. Following the estimation of LP MTFs, Gaussian fitting was applied to the MTF values of the eight LP frequencies to determine the MTF50%. To ensure accurate MTF estimation from patient images, background image subtraction was applied, considering the potential lack of absolute uniformity in patient images that could influence the estimation process [8].

III. Results

Fig. 2 presents the MTF curves derived from both Line Pair (LP) and disk objects. Each LP object contributes to one point on the LP MTF curve, resulting in eight discrete points from the eight LP objects. On the other hand, the edge spread method enables continuous MTF estimation from a single disk object. The MTF50% values were subsequently calculated from the MTF curves. Additionally, Fig. 2 highlights that using disk objects tends to overestimate the MTF in the high frequency range when compared to using LP objects.

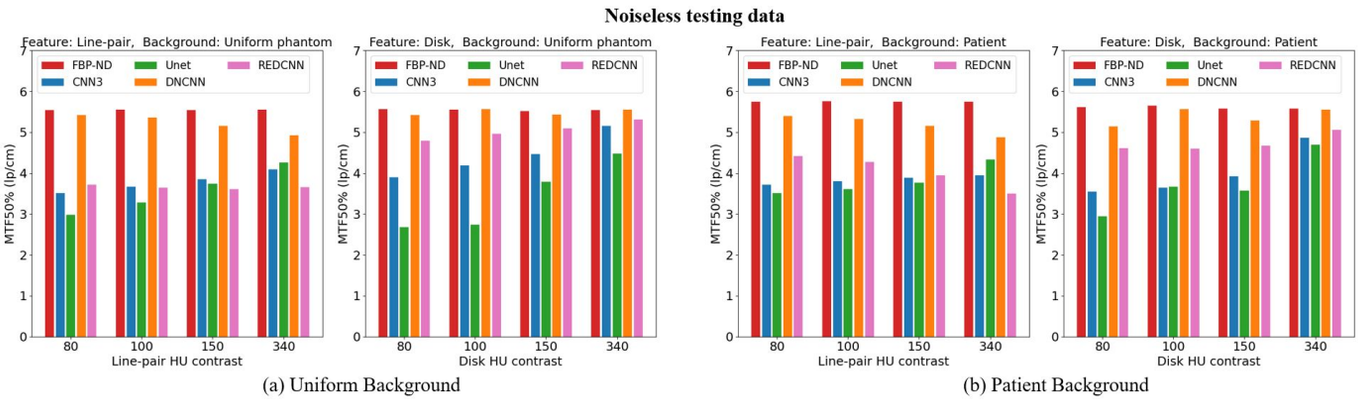
Fig. 3 and 4 show the cdMTF performance results of the four DL denoising models compared with the baseline FBP for different image background conditions and testing feature objects, in noiseless and noisy cases respectively. Overall, FBP's MTF50% value remains consistent across different contrast levels for both testing features in both backgrounds particularly in the noiseless cases. This further validates our implementation of the cdMTF calculation since the results agree with the fact that FBP reconstruction method is linear and has an image resolution independent of the testing images. The MTF50% value of FBP in noisy cases fluctuates possibly due to the impact of noise. The denoising models in general did not preserve the MTF50% values compared with the FBP baseline, especially at lower contrast levels. For example, Fig. 3 Unet had better MTF50% performance on the 340HU LP test object in the uniform background (only 7% drop) than the 80HU (44% reduction relative to FBP). The results also show that the image background did not greatly affect their cdMTF performances. For example, using the LP features in the uniform background and patient background resulted in similar MTF50%-vs-HU trend for each denoising model.

Regarding test features, it appears that the MTF50% measurements obtained using disks and LPs are largely consistent. However, we observe that the LP based MTF50% values for some models could drop substantially in lower contrast levels compared with the corresponding disk-based MTF results in the noisy setting. For instance, in Fig. 4, CNN3's MTF50% for LP dropped 25%, and 12% compared with the MTF50% of disk in HU 80 and 100 respectively . This performance drop could be due to the fact that the edges of low-contrast and dense line pairs are more difficult to be preserved for these nonlinear DL denoising models. In the future, we will include LP and disk features of contrast lower than 80HU to further investigate the magnitude of difference on LP and disk-based low-contrast MTF measurements.

In addition, we observed that in the LP MTF measurements, DNCNN had better MTF performance on the 80HU level in the patient background (only 5% drop) than the 340HU level (13% reduction relative to FBP). A similar behavior can be observed in REDCNN too. Better performance at a lower contrast for a nonlinear DL denoiser is unexpected. We will further investigate the reason in the future.

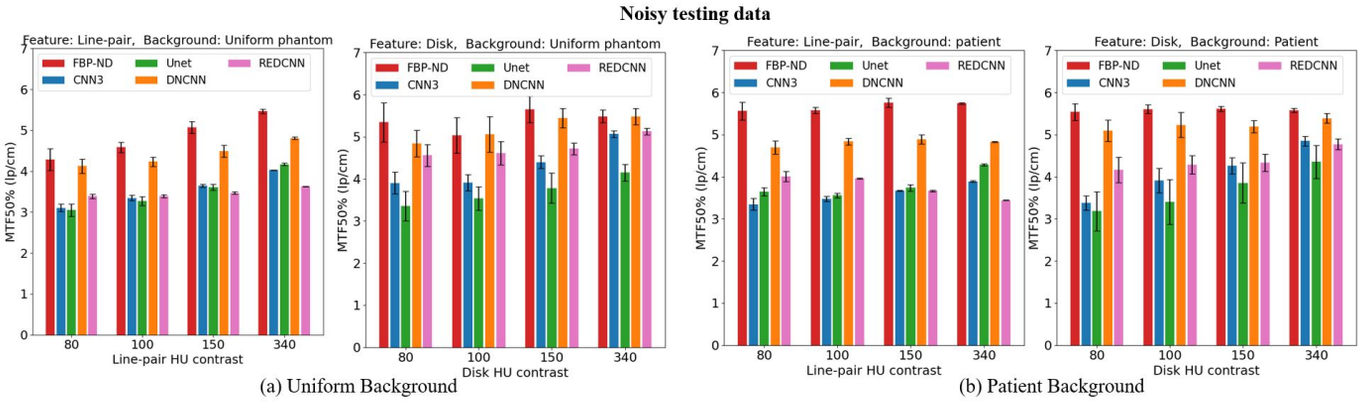
IV. Conclusion

In this study, we conducted a comparison of the contrast-dependent Modulation Transfer Function (cdMTF) performance among four low-dose CT deep learning denoising models. Measurements were taken using both a uniform phantom background and an anatomical patient background. Although the cdMTF performance measure is shown to be consistent between the two backgrounds, it's crucial to point out potential discrepancies introduced by the object type (line-pattern vs disk) at low-contrast levels. It appears that the cdMTF method using disk objects can overestimate the spatial resolution of the DL denoising methods at lower contrast levels compared to the LP MTF method.



(a) Uniform Background (b) Patient Background

Fig. 3: Bar plots comparing MTF50% values for FBP, CNN3, Unet, DNCNN, and REDCNN using LP and disk objects in both uniform phantom and patient background during noiseless reconstruction.



(a) Uniform Background (b) Patient Background

Fig. 4: Bar plots comparing MTF50% values for FBP, CNN3, Unet, DNCNN, and REDCNN using LP and disk objects in both uniform phantom and patient background during noisy reconstruction.

ACKNOWLEDGMENT

This study was supported in part by NIH grant 5R01EB032716-03 and Md Selim acknowledge funding by appointment to the research participant program at the Center for Devices and Radiological Health administrated by the Oak Ridge Institute for Science and Education through an interagency agreement between the US department of Energy and the US Food and Drug Administration.

REFERENCES

[1] Ehsan Samei, Donovan Bakalyar, Kirsten L Boedeker, Samuel Brady, Jiahua Fan, Shuai Leng, Kyle J Myers, Lucretiu M Popescu, Juan Carlos Ramirez Giraldo, Frank Ranallo, et al. Performance evaluation of computed tomography systems: summary of aapm task group 233. *Medical physics*, 46(11):e735–e756, 2019.

[2] Samuel Richard, Daniela B Husarik, Girijesh Yadava, Simon N Murphy, and Ehsan Samei. Towards task-based assessment of CT performance: system and object mtf across different reconstruction algorithms. *Medical physics*, 39(7Part1):4115–4122, 2012.

[3] Rongping Zeng, Prabhat KC, and Brandon Nelson. Bench testing performance of deep learning-based CT image denoising methods: influence of object background on image sharpness and noise texture. In *Medical Imaging 2023: Physics of Medical Imaging*, volume 12463, pages 349–355. SPIE, 2023.

[4] KC Prabhat, Rongping Zeng, M Mehdi Farhangi, and Kyle J Myers. Deep neural networks-based denoising models for ct imaging and their efficacy. In *Medical Imaging 2021: Physics of Medical Imaging*, volume 11595, pages 105–117. SPIE, 2021.

[5] C McCollough, B Chen, D Holmes, X Duan, Z Yu, L Xu, S Leng, and J Fletcher. Low dose CT image and projection data (LDCT-and-projection-data)(version 4). *Med. Phys*, 48:902–911, 2021.

[6] Cynthia H McCollough, Michael R Bruesewitz, Michael F McNitt-Gray, Krista Bush, Thomas Ruckdeschel, J Thomas Payne, James A Brink, and Robert K Zeman. The phantom portion of the american college of radiology (acr) computed tomography (ct) accreditation program: practical tips, artifact examples, and pitfalls to avoid. *Medical physics*, 31(9):2423–2442, 2004.

[7] Ronald T Droege and Richard L Morin. A practical method to measure the mtf of CT scanners. *Medical physics*, 9(5):758–760, 1982.

[8] Zhongxing Zhou, Hao Gong, Scott Hsieh, Cynthia H McCollough, and Lifeng Yu. Contrast-and noise-dependent spatial resolution measurement for deep convolutional neural network-based noise reduction in ct using patient data. In *Medical Imaging 2023: Physics of Medical Imaging*, volume 12463, pages 318–322. SPIE, 2023.

Author List

A

Abadi Ehsan 245
Abdurahman Shiras 344
Abella M. 248
Alamo Juan C. del 384, 388, 431
Ali Abdul Salam Rasmi Asraf 154
An Fengwei 70
Andrew Matthew 280
Andreyev Andriy 280
Anthoine S. 292
Anton Mathias 467

B

Baader Edith 198
Bal Guillaume 194, 221
Bayer Siming 130
Bennett N. Robert 463
Bernard Guillaume 376
Bert Christoph 78
Berthe Daniel 420
Bismark Richard N.K. 142
Blum Nele 332
Bollig Jonathan 209
Bortfeldt Jonathan 420
Boursier Y. 292
Brown Kevin M. 340, 451
Butala Mark D. 42
Buzug Thorsten M. 332
Byl Achim 86
Byrd Darrin W. 245

C

Cai Liang 471
Cannet F. 292
Cao Jennie 320, 487
Cerro F. C. Del 248
Chanda Rajib Kumar 380
Chen Buxin 22, 34, 201, 241
Chen Guang-Hong 174
Chen Maxin 420
Chen Ruifeng 126
Chen Xi 471
Chen Yang 126
Chen Zhennong 66
Chen Zhiqiang 98, 348
Contijoch Francisco 384, 431
Costa Paulo R. 439

D

Daerr Heiner 451
Daly Morgan 483
Darambara Dimitra 118
Debnath Anupama 368
Dedes George 420
Delnooz C. 316
Deng Zhi 182
Desari Paul 245
Desbat Laurent 376
Desco M. 248
Dettmer Jannik 284
Dong Di 364
Dorosti Tina 114
Du Muge 392
Duan Baojun 134
Duan Xiaoyu 22
Dupont M. 292
Dörfler Arnd 344

E

Ehtiati T. 166
Eulig Elias 54, 86, 110

F

Fan Fuxin 78, 82, 94, 130, 308
Fan Ying 368
Fang Mengjie 364
Faria Andreia V. 62
Fietkau Rainer 78
Fitzgerald Paul 368
Frank David 237
Frysch Robert 213, 344
Fu Zhiyang 138
Fung George S.K. 284, 336
Fusiello Andrea 154

G

Gagnon Daniel 400
Gang Grace J. 186, 190, 324, 447, 451
Gao Hewei 182, 276, 348
Gao Jianpeng 134
Ge Yinglin 268
Geagan Michael 475
Gizewski Elke 416
Glocker Ben 118
Gomaa Ahmed 78
Gonzales B. 316
Gonzalez F. 166
Gonzalo Alejandro 384, 431
Goorden Marlies C. 288
Gorges Sebastien 376
Gruber Nadja 416
Gu Mingxuan 82, 130, 158

Guo Jinyuxuan 384, 431
Guo Peiyuan 233
Guo Z. 336
Gupta Rajiv 98

H
Haltmeier Markus 416
Hammermann Maximilian 110
Haneda Eri 368
Haseljic Hana 213
He Bingxi 364
Hei Dongwei 134
Heinz Nico 304
Heinzl Christoph 264
Herl Gabriel 90, 122
Hernández-Girón Irene 435
Hietanen Ari 296
Hoffman John 483
Hsieh Scott S. 229, 459
Hu Guyue 412
Hu Rui 66, 98
Huang Hailing 22
Huang Heyuan 166, 316
Huang Yixing 78, 94

I
Ibrahim Michael S. 221
Im Jessica Y. 475

J
Jadick Giavanna 372
Jiang Xiao 30, 186, 324
Jobin James 368
Jolivet Frederic 376
Jung Alexander 122
Jäger Fabian 74, 209

K
Kabelac Anton 54
Kabus Sven 475
Kachelrieß Marc 54, 74, 86, 110, 198, 209
Kahn Andrew M. 388
Kastner Johann 264
Kattau Margarete 118
KC Prabhat 491
Khurana Gurnoor Singh 384, 431
Kim Joonbeom 150
Kinahan Paul E. 245
Koehler Thomas 451
Kolehmainen Ville 296
Krishna Arjun 328
Kristof Gernot 86
Kröncke Thomas 412
Kulvait Vojtech 213
Kumar Ankit 264
Kurz Christopher 412

L
Labno Thomas 427
Landi Claudio 154
Landry Guillaume 412
Lang Stefan 344
Lasser Tobias 114, 237
Laurendeau Matthieu 376
Le Min 146
Lee Duhgoon 98
Lee Okkyun 150, 178
Lee Ting-Yim 62
Lee Tzu-Cheng 471
Lee Wonkyeong 320, 487
Leenen Jacco van 288
Leibold David 288
Li Ke 174
Li Mengzhou 360
Li Quanzheng 66, 408
Li Shudong 30
Li Xiaoyue 42
Li Yang 134
Li Yinsheng 126
Li Yunxiang 46
Liang Dong 126
Liang Li 134, 146, 225
Litt Harold I. 356, 479
Liu Leening P. 340, 356, 447, 479
Liu Stephen Z. 186
Liu Xiangyuan 50, 70
Liu Yinong 392
Lopez-Montes Alejandro 316
Lorenzon Altea 186
Lorraine Peter 368
Lu A. 166
Lu Xin 106, 256
Lucking Hannes 344
Lutz Raffael 217

M
Ma Jianhua 364
Ma Yiqun Q. 447
Mader Ulf 467
Maebe Jens 272
Magnusson Maria 404
Mahony Michael 368
Maidu Bahetihazi 384, 431
Maier Andreas . . . 26, 78, 82, 90, 94, 122, 130, 158, 304, 308, 312
Maier Joscha 54, 110
Maltz Jonathan S 360
Malusek Alexandr 404
Man Bruno De 360, 368
Manhart Michael 344
Manjeshwar Ravindra M. 451
Maul Noah 158
McSkimming T. 316
McVeigh Elliot R. 388
Medrano Maria Jose 320, 487
Mei Kai 427, 475
Mei Siyuan 82, 130, 158
Men Longchao 233
Mengke Tielige 170
Micah Neghemi 475
Michen Markus 380
Michielsen Koen 252, 396
Mikerov Mikhail 252, 396
Morel C. 292
Mu Wei 364

Mueller Klaus 328
Mur Lorente 248

N

Nassi Martina 396
Niepel Katharina 412
Niinimaki Kati 296
Niu Chuang 360
Noo Frédéric 336, 455
Noël Peter B. . . . 268, 340, 356, 427, 447, 451, 475, 479

O

Olivas Katie 245
Ongie Gregory 241, 423
Oostveen Luuk J. 439

P

Pack Jed 368
Pan Tinsu 170
Pan Xiaochuan 22, 34, 194, 201, 221, 241
Parodi Katia 412, 420
Pasyar Pouyan 356, 479
Paysan Pascal 74
Perkins Amy E. 340, 451, 475
Pfaff Laura 158
Pfeiffer Daniela 114
Pfeiffer Franz 114, 420
Phillips John Paul 22, 194, 221
Pimenta Elsa B. 439
Pramanik Aniket 300
Prekas Georgios 400
Proksa Roland 268, 451
Pu Shaozhou 225
Punzet Daniel 344
Putz Florian 78

Q

Qi Jinyi 38
Qin Yanwei 106, 256
Qiu Jingna 94

R

Rahman Obaidullah 300
Rajbhandary Paurakh 229
Ran Zhizhong 50
Reddy Nandan 368
Reginatto Marcel 467
Reiser Ingrid S. 22, 194, 221
Risch Franka 412
Rit Simon 376
Riviere Patrick La 372
Rose Georg 142, 213, 344
Roshkovan Leonid 427
Ross Steven 427
Rybertt Martin V. 356

S

Sahbaee Pooyan 479
Samei Ehsan 245
Sandborg Mikael 404
Sandvold Olivia F. 268, 451, 479
Sarti Cristina 154
Schaart Dennis R. 288
Schaff Florian 114
Schielein Richard 304
Schneider Linda-Sophie 26, 90, 130, 308, 312
Schultheiß Manuel 114
Schwab Johannes 416
Schwarz Florian 412
Scienti Oliver Pickford 118
Sechopoulos Ioannis 252, 396
Segars W. Paul 62, 162, 245, 284
Selim Md 491
Severance Lauren M. 384, 388, 431
Shao Hua-Chieh 46, 170
Shapira Nadav 356
Sharma Shobhit 427
Shen Le 352
Shen Yuan 30
Sheng Liang 134
Shi Linxi 463
Sidky Emil Y. 22, 34, 194, 201, 221, 241
Siewerdsen J. H. 166
Sisniega A. 166, 316
Skeats A. 316
Smit Ewoud J. 252
Song Jiadan 225
Song Yan 134
Song Yukang 182
Soulen Michael C. 340
Speck Oliver 213
Stayman J. Webster 30, 186, 324, 447, 451
Stevens Grant M. 102, 320, 487
Stierstorfer Karl 284, 336
Stille Maik 332
Subramanian Shalini 62, 162
Sun Tianjiao 106
Sun Yipeng 82, 130, 308, 312
Syben Christopher 304

T

Taguchi Katsuyuki 62, 162
Tang Xiangyang 443
Tedgren Asa Carlsson 404
Terzioglu Fatma 194, 221
Thalhammer Johannes 114
Thies Mareike 82, 130, 158
Thompson Richard 427
Tian Jie 364
Tivnan Matthew 66, 408
Tomal Alessandra 396
Tse Justin Ruey 320, 487
Tseng Hsin Wu 138
Tunissen Sjoerd A.M. 252

V

Vandenberghe Stefaan 272
Vauhkonen Marko 296
Vedantham Srinivasan 138
Venkatakrishnan Singanallur V. 300
Vervenne Boris 272
Vysotskaya Nastassia 158

W

Wagner Fabian 82, 158
Walczak Michał 74

Wang Adam S. 102, 320, 463, 487
Wang Gaoang 42
Wang Ge 328, 360
Wang Sen 102, 320, 487
Wang Shuo 364
Wang Xuezhi 182
Wang Yongbo 364
Wang Yuang 98
Wang Zhentian 233
Wang Zhilei 276
Werncke Thomas 213
Wettenhovi Ville-Veikko 296
Wiedeman Christopher 360
Wu Dufan 66, 98, 408
Wu Mingye 368
Wu Xiangyi 22
Wurl Matthias 412

X

Xia Dan 22, 34, 201, 241
Xia Dimeng 50, 70
Xia Yingxian 348
Xie Jiacheng 46
Xing Yuxiang 260, 348, 352, 392
Xu Jingyan 455

Y

Yang Faguo 280
Yang Yi 443
Yang Yirong 102
Yang Zeyu 217
Ye Chengze 312
Yoon Siyeop 66, 98
Yu Baihui 98
Yu Haijun 158
Yu Xinran 106
Yu Zhicong 400
Yu Zhou 427
Yuan Nimu 38
Yuan Yuedong 26

Z

Zbijewski W. 316
Zeng Gengsheng L. 58, 205
Zeng Rongping 491
Zhan Linying 174
Zhang Jiayong 368
Zhang Leo Y. 22
Zhang Li 98, 233, 348
Zhang Weibin 50, 70
Zhang You 46, 170
Zhang Zheng 22, 34, 201, 241
Zhao Shengzi 352
Zhao Shusen 50, 70
Zhao Wei 22
Zhao Xing 256
Zhao Yunsong 106, 256
Zhou Jian 38
Zhou Yuzhong 130, 308
Zhu Hongyang 106, 256
Ziabari Amirkoushyar 300
Zollner Frank G. 217
Zong Chunguang 146
Zou Xiang 260

Schedule CT Meeting 2024

	Monday *August 5*	**Tuesday** *August 6*	**Wednesday** *August 7*	**Thursday** *August 8*	**Friday** *August 9*
8:00 - 9:40 CEST		**Recon-struction** General opening remarks + 4 talks *(Page 7)*	**Photon-Counting** Basic short course + 3 talks *(Page 12)*	**AI and Physics** Basic short course + 3 talks *(Page 16)*	**Image Quality I** Basic short course + 3 talks *(Page 20)*
		Coffee	**Coffee**	**Coffee**	**Coffee**
10:00 - 11:40 CEST		**Missing Data** 5 talks *(Page 7)*	**Spectral** 5 talks *(Page 12)*	**Artifacts** 5 talks *(Page 16)*	**Image Quality II** 4 talks + Preview of Fully 3D 2025 and CT Meeting 2026 *(Page 20)*
		Lunch	**Lunch**	**Lunch**	**Lunch**
13:20 - 15:00 CEST		**Poster** Poster FF + 24 Posters *(Page 8)*	**Poster** Poster FF + 24 Posters *(Page 13)*	**Poster** Poster FF + 24 Posters *(Page 17)*	**Visit to EZRT** *(Page 21)*
		Coffee	**Coffee**	**Coffee**	
15:20 - 17:00 CEST	**Registration**	**Motion** 5 talks *(Page 10)*	**Deep and Industrial** 5 talks *(Page 15)*	**Novel CT** 5 talks *(Page 19)*	
17:00 - 17:40 CEST		**Plenary** Stephan Achenbach			
19:00 - 22:00 CEST	**Registration and Reception at Restaurant Plückers**	**Dinner at Residenz-schloss**		**Dinner at Klosterbräu**	