Title

Metric learning guided sinogram denoising for cone beam CT enhancement.

Authors

Li, Haoran
Tsai, Yun-Han
Liu, Hengjie
et al.

Publication Date

2024-10-01

DOI

10.1002/mp.17435

Metric Learning Guided Sinogram Denoising for Cone Beam CT Enhancement

Authors: Haoran Li[a], Yun-Han Tsai[a], Hengjie Liu[b] Dan Ruan[a, b, c]

[a] Department of Bioengineering, University of California Los Angeles, Los Angeles, CA 90095, USA.

[b] Graduate Program of Physics and Biology in Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA.

[c] Department of Radiation Oncology, University of California Los Angeles, Los Angeles, CA 90095, USA.

Corresponding Authors:

Dan Ruan

druan@mednet.ucla.edu

Abstract

Background: Cone beam CT is a widely available modality, but its clinical utility has been limited by detail conspicuity and quantitative accuracy. Convenient post-reconstruction denoising is subject to back projected patterned residual, but joint denoise-reconstruction is typically computationally expensive and complex.

Purpose: In this study, we develop and evaluate a novel **Me**tric-learning **g**uided w**a**velet **tr**ansform rec**on**struction (MEGATRON) approach to achieve image domain quality enhancement from projection-domain processing.

Methods: Projection domain-based processing has the benefit of being simple, efficient, and compatible with various reconstruction toolkit and vendor platforms. However, they also typically show inferior performance in the final reconstructed image, because the denoising goals in projection and image domains do not necessarily align. Driven by this insight, this work strives to develop a simple approach to translate the demand for quality enhancement from the quantitative image domain to the more easily operable projection domain. Specifically, the proposed paradigm consists of a metric learning module and a denoising network module. Via metric learning, enhancement objectives on the wavelet encoded sinogram domain data are defined with respect to post-reconstruction image discrepancy. The denoising network maps measured cone-beam projection to its enhanced version, driven by the learnt objective. In doing so, the denoiser operates in the convenient sinogram to sinogram fashion but reflects improvement in reconstructed image as the final goal. Implementation-wise, metric learning was formalized as optimizing the weighted fitting of wavelet subbands, and a res-Unet, which is a Unet structure with residual blocks, was used for denoising. To access quantitative reference, cone-beam projections were simulated using the X-ray-based Cancer Imaging Simulation Toolkit (XCIST). In both learning modules, a data set of 123 human thoraxes which was from Open-Source Imaging Consortium (OSIC) Pulmonary Fibrosis Progression challenge, was used. Reconstructed CBCT thoracic images were compared against ground truth FB and performance was assessed in Root Mean Square Error (RMSE), Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM).

Results: MEGATRON achieved RMSE in HU value, PSNR, and SSIM were 30.97±4.25, 37.45±1.78, and 93.23±1.62, respectively. These values are on par with reported results from sophisticated physics driven CBCT enhancement, demonstrating promise and utility of the proposed MEGATRON method.

Conclusions: We have demonstrated that incorporating the proposed metric learning into sinogram denoising introduces awareness of reconstruction goal and improves final quantitative performance. The proposed approach is compatible with a wide range of denoiser network structures and reconstruction modules, to suit customized need or further improve performance.

Keywords

Cone beam computed tomography (CBCT), Deep Learning, Metric Learning

1. Introduction

Cone-beam computed tomography (CBCT) is a widely available modality that can be used to generate fast volumetric anatomic imaging. While its quality is typically sufficient for qualitative assessment and depiction, the combined impact from scatter, motion, and low signal-to-noise ratio makes it an inferior source for quantitative information. CBCT enhancement or denoising is an active area of research, and the main methods can be categories as either postprocessing or with direct engagement in reconstruction. In the post-processing based enhancement category, fast filtered-back projection FDK reconstruction is first performed, and then adaptive noise reduction filters[1] or nonlocal means (NLM) method such as BM4D [2] can be applied to reduce noise in CBCT while preserving edges and details. The filtering approach is very intuitive and can be implemented with high efficiency, but usually fails to achieve very high quantitative accuracy as measurement uncertainty or noise was propagated first by back projections onto the complete spatial domain and cannot be fully decoupled after the fact. In the second category, regularization priors on the reconstruction can be incorporated directly into the optimization-based reconstruction, via basis encoding such as wavelet[3], tight frame[4], or more recently learning based features[5,6], and can be combined with variational operators[7]. This approach offers flexible ways to incorporate priors but usually requires iterative procedures for reconstruction that can be time consuming. More recently, deep learning-based reconstruction with unrolled network has demonstrated promise but its clinical stability could be sensitive to specific data characteristics and is still under exploration.

In this study, we work in the sinogram domain with an overall sequential procedure, performing projection data correction and standard reconstruction. Our major innovation and contribution lie in that we define the sinogram correction objective by conducting a metric learning study so that it reflects the targeted quality improvement in the image domain. Subsequently, a denoising network, trained with the learned metric, is used to denoise the sinogram data. The utility of the proposed **Me**tric-learning **g**uided w**a**velet **tr**ansform rec**on**struction (**MEGATRON**)

approach is validated and assessed using realistically simulated cone beam data against reference.

2. Methods and Materials

2.1 MEGATRON Metric learning

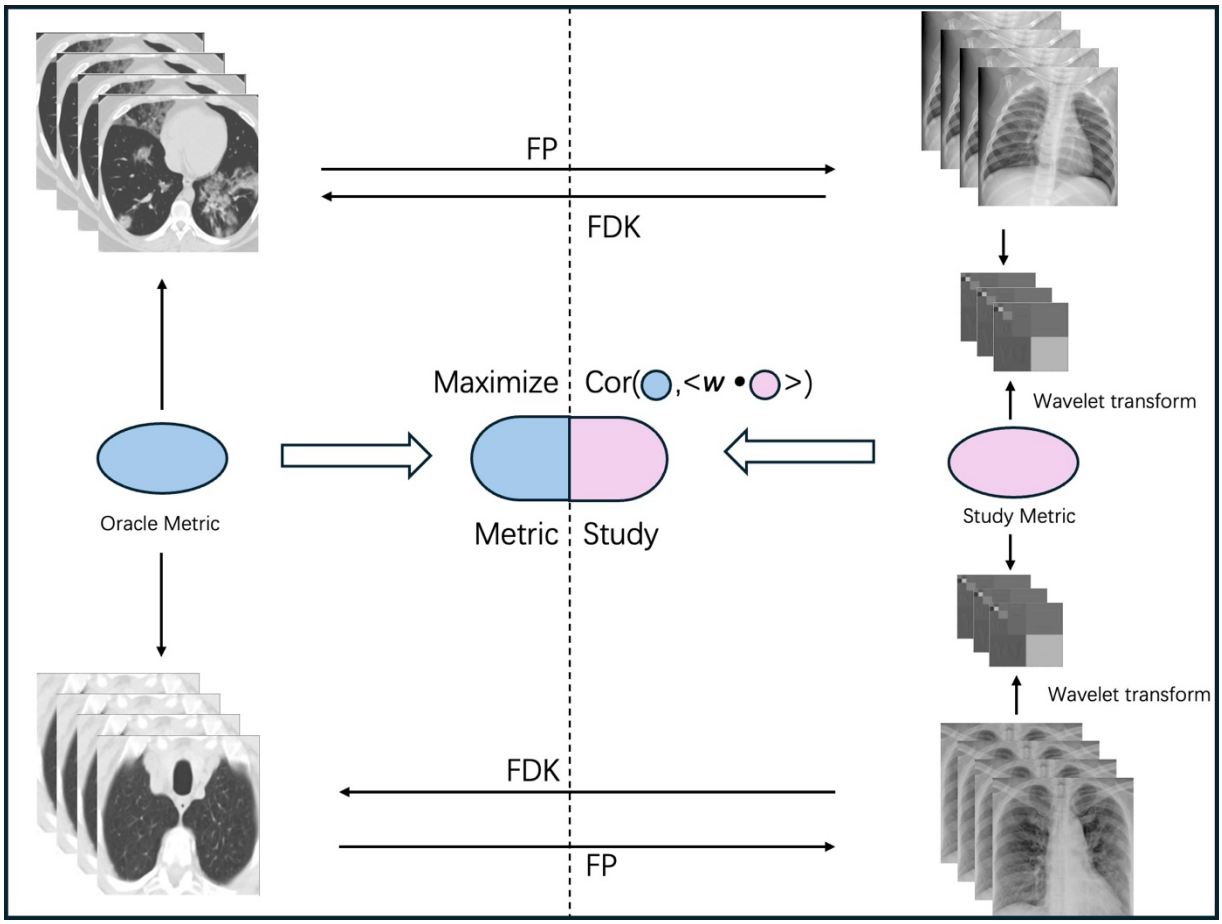

Figure 1. Schema of the metric transfer module in MEGATRON. For any two datasets, known as a pair, the left-side panel describes the images and the relationships in the image domain. The oracle metric is defined based on the quantitative electron density from reference fan-beam CT which we aim to learn but is inaccessible in practice. On the right-side panel are the corresponding projection sets that are presented with the corresponding sinograms wavelet representations. A set of wavelet band weights w is sought to maximize the correlation between these two objectives.

Figure 1 illustrates the main idea of metric transformation in MEGATRON. To enhance reconstructed CBCT quality, it is desirable to have its reconstruction match that provided by its Fan-beam counterpart[8]. Ideally, one wants to perform denoising with respect to the difference between image reconstruction and the ground truth fan-beam quantitative map, a metric that is

ideal but typically inaccessible, termed “Oracle” metric on the reconstructed image domain. On the other hand, processing directly on sinogram makes it easier to cope with scatter, beam hardening, and starvation. It also provides a more modular way to integrate with openly or commercially available reconstruction software. Since forward projection is not unitary, applying the same quantitative metric (e.g., mean squared difference) to the image versus the sinogram domain would have different denoising impact. We propose to work with the sinogram domain and modify the commonly used objective with wavelet-band-variant-weights. By maximizing the correlation between the weighted wavelet representation discrepancy to the Oracle metric, we introduce reconstruction awareness into the objective function w.r.t. sinogram domain, shown in **Equation 1**.

$$Maximize_w \quad Corr(\left\|\mathrm{Image_i} - \mathrm{Image_j}\right\|_2, \langle \left\|\mathrm{Wavelet(Sino)_i} - \mathrm{Wavelet(Sino)_j}\right\|_2, \mathrm{w}\rangle) \quad (1)$$

$$\text{s.t.} \quad \sum_{k=1} w_k = 1, \; w_k \geq 0$$

where w is the weight vector corresponding to the contribution of the K wavelet bands. It is constrained to reside on the simplex to avoid trivial nulls-space caused by trivial scaling and the corresponding oscillations in optimization convergence. The correlation is compared against all image/sinogram pairs (i, j) in samples supporting metric learning.

2.2 Denoising network

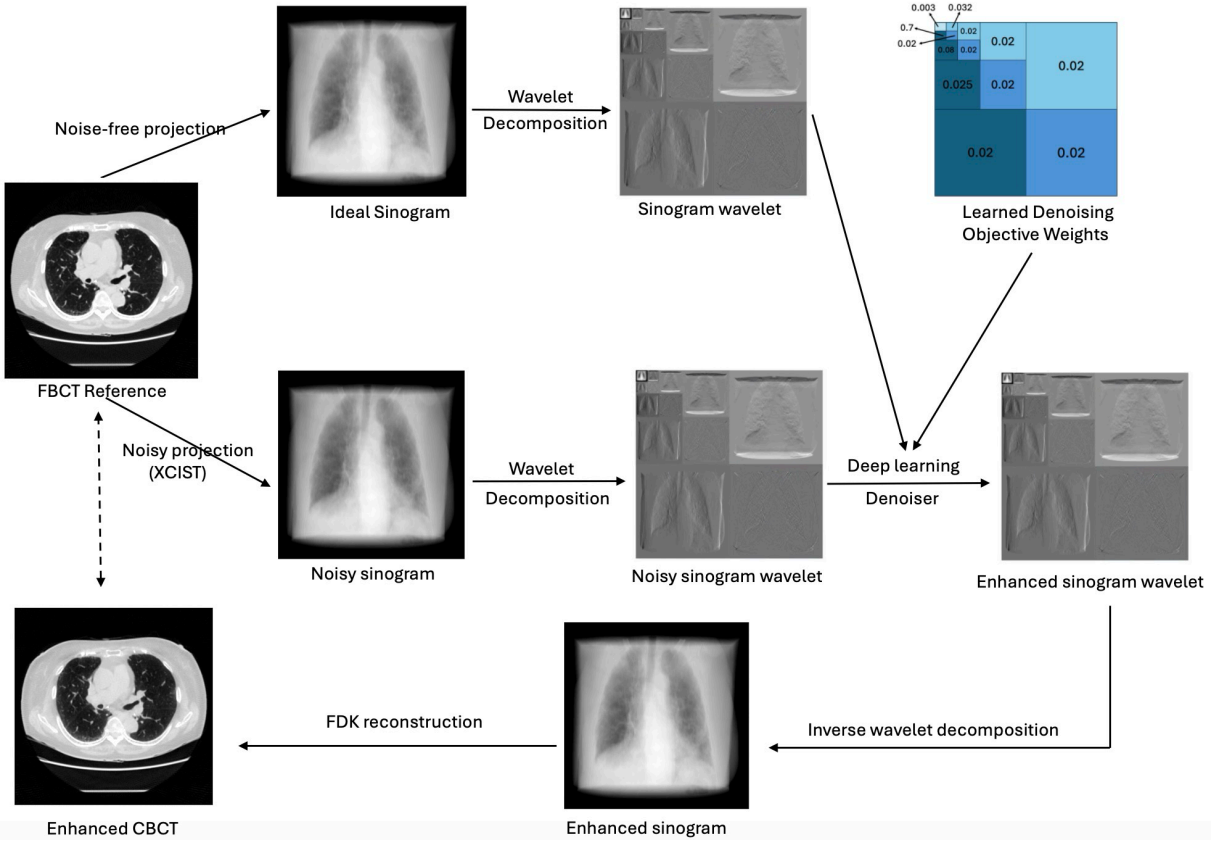

Figure 2. Schema of the workflow of MEGATRON. Our presenting workflow consists of two major modules, one is the metric learning module, the other is the denoising module.

As demonstrated in **Figure 2**, the learnt metric is utilized to train for the denoising network, and subsequently for denoising and reconstructing the image volume. The denoising network with a Dense Residual 3D U-Net backbone[9], is used to map wavelet coefficients of the corrupted sinogram data to its ideal counterpart. The network is trained in a supervised fashion to seek the optimal network parameters Θ. **Equation 2** presents the corresponding loss function, where Φ is our denoiser and w is the result from metric learning.

$$\langle \| \Phi(\text{Wavelet}(\text{Sino})_{\text{corrupt}};\ \Theta) - \text{Wavelet}(\text{Sino})_{\text{ideal}} \|_2 , \text{w} \rangle \qquad (2)$$

2.3 Comparison benchmarks

To understand and appreciate the contribution of each design component, namely the metric learning and wavelet representation, we assessed two denoising benchmarks for ablation analysis and performance comprehension. The first one directly uses sinogram data and trains a

denoising network to minimize the conventional L_2 difference between denoised data and its ideal counterpart, abbreviated as “Denoise in sino”. The second one trains the network on the wavelet-transformed representation, also driven by L_2 cost, noted as “Denoise in wl-sino”. For fair comparison, data usage and network structure are kept the same across the benchmarks and the proposed work.

2.4 Implementation and Experimental details

To perform metric optimization, 50 chest volume data from the Open-Source Imaging Consortium (OSIC) Pulmonary Fibrosis Progression challenge were used[10]. Each sinogram image was decomposed into 13 wavelet subbands using open-source package pywt[11], where decomposition level is set to 4. For N sample pairs, the Mean Squared Error (MSE) for each subband was calculated, forming an Nx13 matrix. In the image domain, the pairwise difference was recorded in an Nx1 matrix. The optimization in **Equation 1** specialized to estimating the 13x1 weighting and was solved with a sequential least squares programming approach 'SLSQP' using the open-source library Scipy.

In denoiser training process, a different subset of 66 chest CT volumes was used. The whole cohort was split into 50 for training and 16 for validation. The Xcist open-source library was used to simulate the primary only (ideal) sinogram, and its corresponding scattered and Poisson noise corrupted the measurement. The training was performed using 32 projection angles with 512x512 pixel detector arrays. The denoising network adopted a Dense Residual 3D U-Net backbone[10] comprising five encoding and decoding levels respectively. Within the network, skip connections were established from each encoder layer to the corresponding decoder feature layers. In each layer, residual blocks were used to maintain the information across convolutional operations and support better back propagation. Our network was built with PyTorch version 2.0.1 and CUDA 11.7, operating on an NVIDIA GeForce RTX 1080Ti GPU, with the Adam optimizer with a learning rate of 0.0001. The testing for sinogram denoising and FDK reconstruction were performed on a separate holdoff set of 7 subjects.

3. Results

3.1 Cone-beam projection Simulation

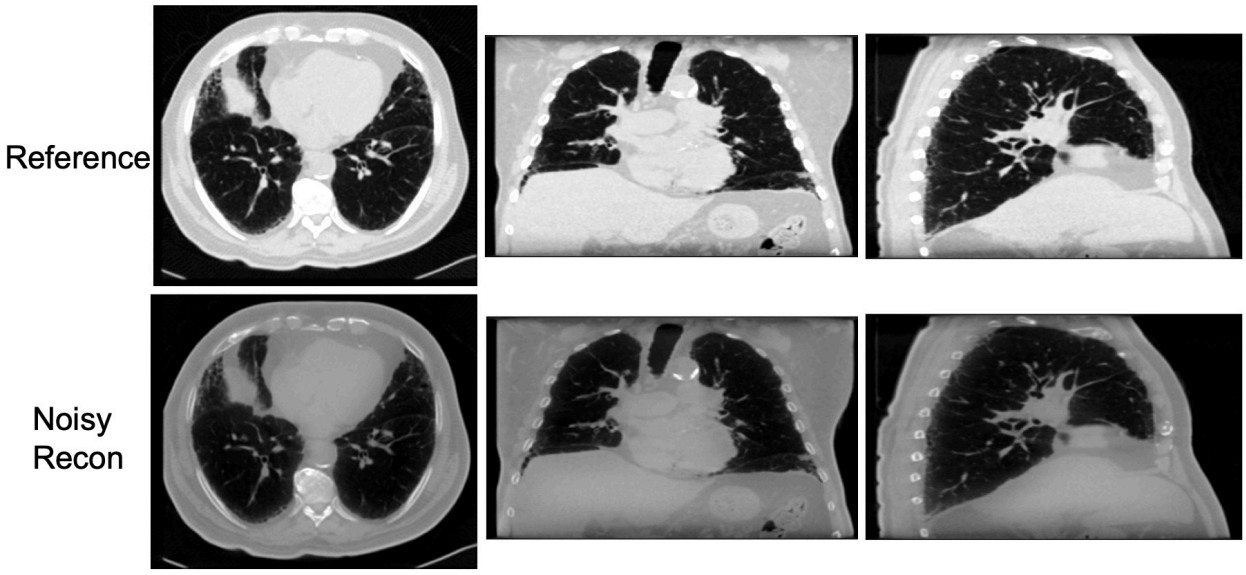

Figure 3. Visualization of the model target and model input (from scatter-Poisson corrupted sinogram simulation with SPR = 2.9). From left to right are axial view, coronal view, and sagittal view of the object. All images are displayed with window level [-600,1500].

Sinograms with a Scatter to Primer Ratio (SPR) of 2.9 were simulated. When reconstructed with FDK, it resulted in Normalized Mean Square Error (NMSE), structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) were 105.13±11.79, 75.06±4.51 and 22.48±1.68. Illustrative examples and their line profile details are presented in **Figures 3** and **4**.

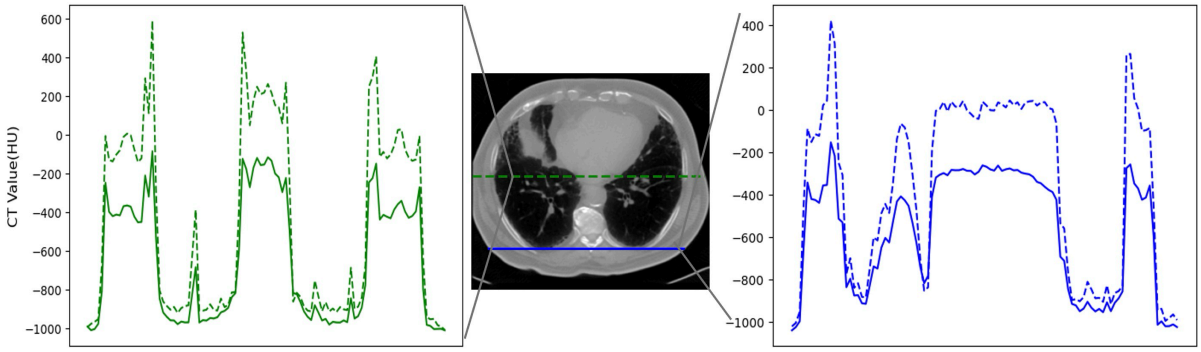

Figure 4. Line profile view for the reconstructions from reference (primary only) vs. scatter-Poisson corrupted sinogram simulation with SPR = 2.9. Solid lines indicate the reference and dashed lines indicate reconstruction from the simulated noisy measurements.

3.2 Wavelet band weights from metric Learning.

The optimized subband weights are reported in **Figure 5**, with its impact on the correlation between the sinogram and image domains shown in **Figure 6**.

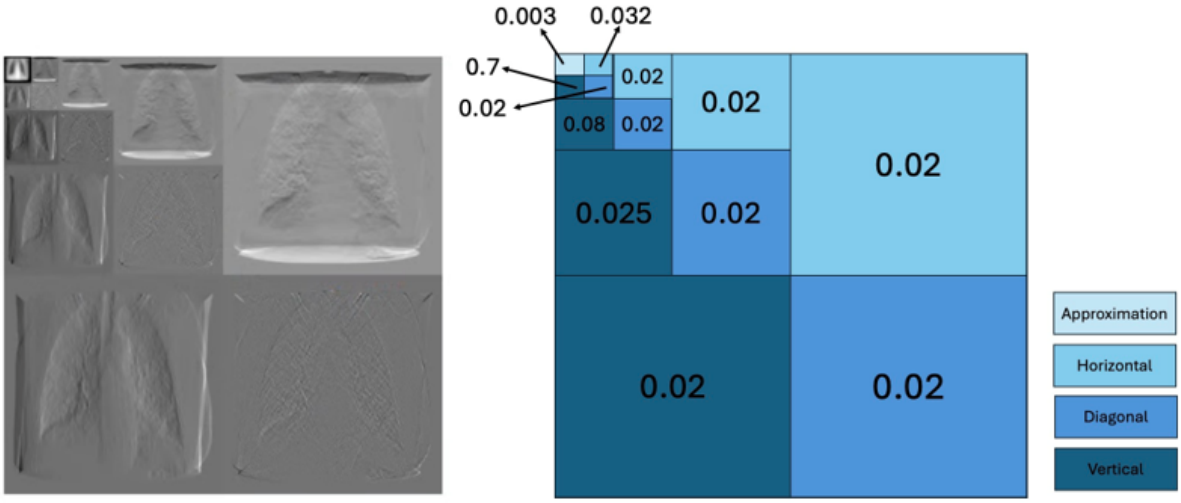

Figure 5. Wavelet decomposition of a sinogram and the optimized band weighting

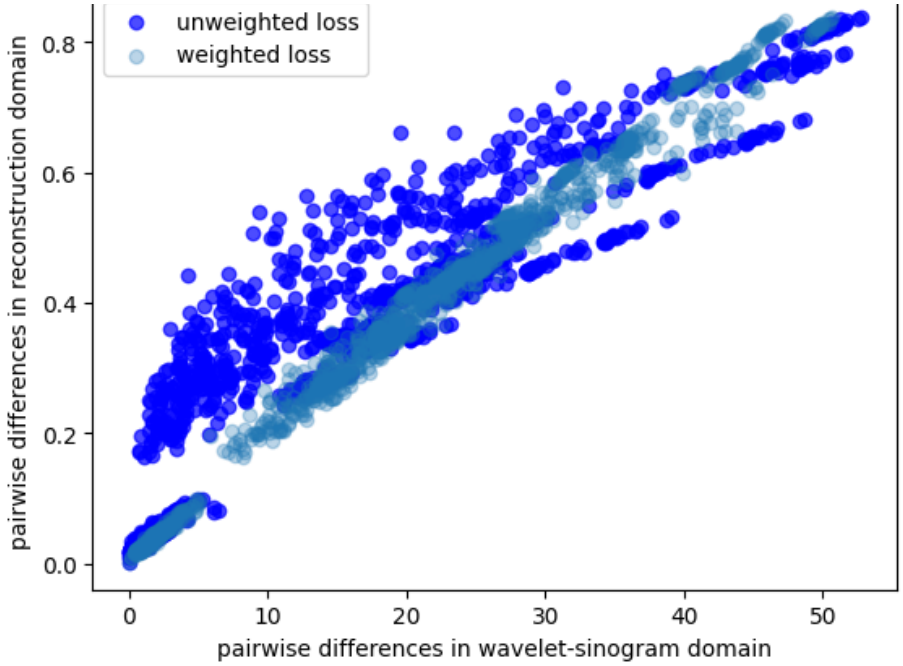

Figure 6. Scatter plot of the L2 metrics in the wavelet-sinogram domain and image domain. The coordinates of each sample correpond to the "distance" between a pair of samples, with the x-coordinate representing the pairwise distance in sinogram domain, and the y-coordinate representing the distance in image domain. With the introduction of wavelet band weighting, the correlation increases from 0.94 to 0.98, with a corresponding R2 increasing from 0.88 to 0.96.

The wavelet weighting shows preferrable emphasis on superior-inferior oriented features on various scales but discounts the global low frequency component. The improvement in cross-domain metric alignment is notable from **Figure 6**. The much tighter spread with an R2 going from

0.88 to 0.96 shows that the adjustment makes the two metrics much better aligned. In other words, weighted wavelet-sinogram metric is an improved surrogate than the unweighted version for the image domain discrepancy metric. This is in alignment with our optimization intention. The ultimate impact on reconstruction is further validated and reported in **Section 3.3.**

3.3 Comparison with Different methods

Table 1. Quantitative evaluation of different methods in Reconstruction and sinogram domain, Mean ± std.

	Reconstruction Domain			Sinogram Domain		
	RMSE	SSIM	PSNR	NMSE	SSIM	PSNR
Noisy	173.83±16.83	75.36±4.23	22.43±1.57	$2.5x10^{-1}$ ±$2.0x10^{-2}$	92.42±1.98	16.52±0.89
Denoise in sino	56.28± 7.10	85.44±3.55	32.27±2.12	**$2.0x10^{-4}$ ± $7.1x10^{-5}$**	**99.41±0.14**	**47.42± 1.76**
Denoise in wl-sino	39.81±4.86	90.55±1.72	35.25±1.72	$5.0x10^{-4}$ ± $9.5x10^{-5}$	98.03±0.23	43.44± 0.88
MEGATRON	**30.97±4.25***	**93.23±1.62***	**37.45±1.78***	5.0 $x10^{-4}$± 1.0 $x10^{-4}$	97.42± 0.39	43.21± 0.88

Table 1 reveals the importance of aligning denoising intention to the use of objectives and the corresponding configurations. When direct sinogram data was used as the input and output, and the objective was also defined directly on the paired difference for supervised learning, as in the "Denoise in sino" configuration, it yields the best denoising performance in the sinogram domain. However, with the intention being performance in reconstruction domain, the proposed MEGATRON excels in all assessment metrics, improving RMSE, SSIM, and PSNR. Paired t-test resulted in p-values of 0.003, 0.011, 0.036 respectively for these three metrics, compared with the close runner-up of "Denoise in wl-sino," indicating statistically significant superiority of MEGATRON. **Figures 7** and **8** show the reconstruction results from sinogram denoising when we used different objectives to drive the same 3D ResUnet denoising network structure.

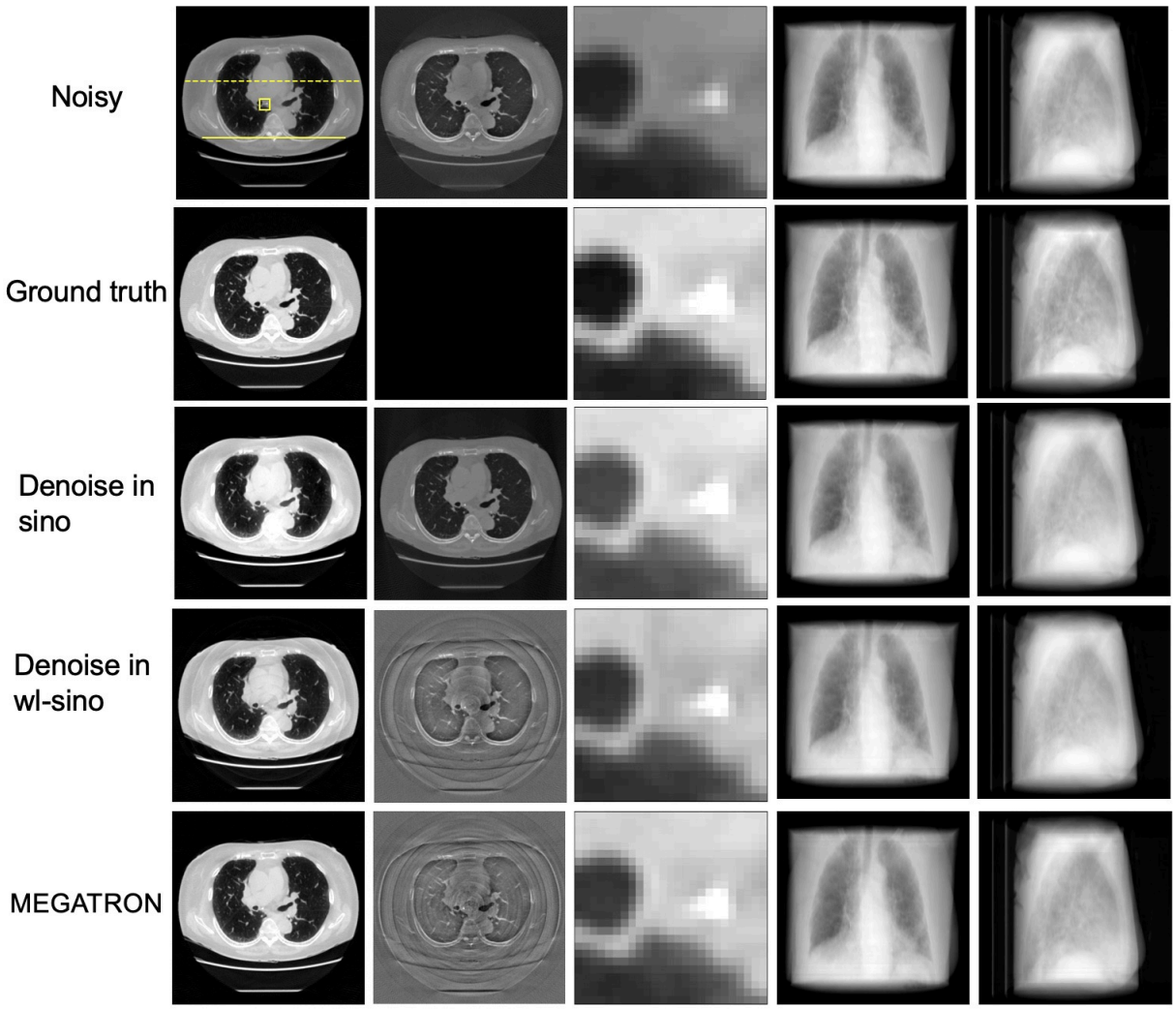

Figure 7 Visualization of an example of denoising results. Columns (left to right): an axial view, axis view of the difference image, a zoom-in at the indicated box region, coronal and sagittal sinogram views. Rows (top to bottom): reference truth, noisy simulation, denoised with L2 cost in sinogram, denoised with L2 cost in wavelet sinogram, denoised with learnt wavelet-band weighted L2 cost in sinogram. All reconstructed images are displayed with window level [-600, 1500].

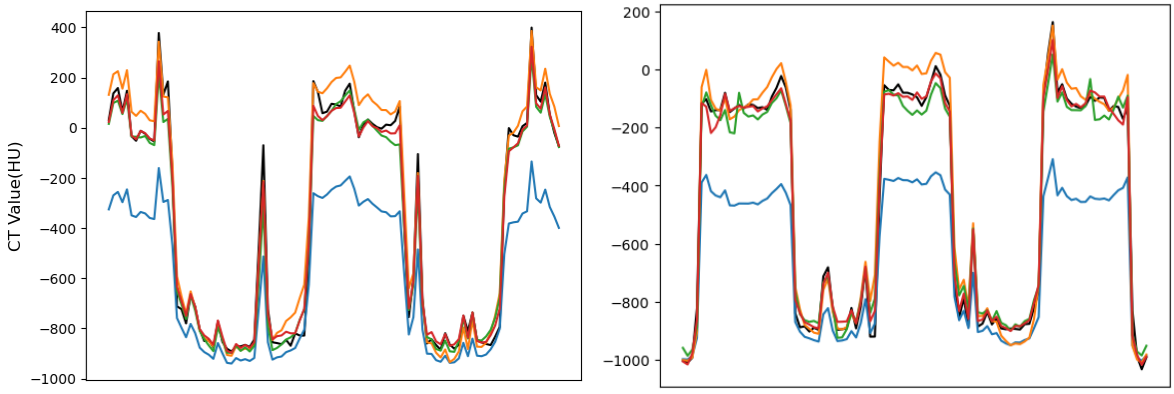

Figure 8. Line profile comparison among various methods. The left and right panels correspond to the upper dashed and lower solid line locations in Figure 7. Ground truth (-), noisy (-), denoising on sinogram (-), denoising on wavelet (-) and the proposed MEGATRON(-) are plotted in black, blue, orange, green, and red, respectively.

4. Discussions and Conclusion

We have proposed a simple approach to perform denoising in the sinogram domain with awareness of quantitative reconstruction goals in CBCT reconstruction. The major contribution is that a cross-domain objective can be applied with any objective driven or guided denoisers in the sinogram domain including BM4D or various networks[12,13]. It fits in well with a sequential denoising-reconstruction pipeline that is friendly with clinical use. The benefit is insensitive to the choice of reconstruction module and can be determined based on user access or preference. We used simple ResUnet for the denoiser and standard FDK for the latter, to demonstrate the contribution from the transfer metric. It is expected that further optimization of either one or both components can further improve the absolute performance further.

To provide a good quantitative reference from FBCT, XCIST was used to simulate scatter, spectrum integration, scatter factor, Poisson/Gaussian noise, and detector characteristics[14]. We used the default parameters in the XCIST which was calibrated against known phantoms and clinical data to ensure realism in our simulations. With XCIST using a convolution kernel approach to simulate scatter, and handling attenuation in the line integral fashion (post logarithm transfer of count ratio), negativity arose in the air and was converted to zero in post-processing. This could lead to an overall bias towards lower attenuation in the simulation. It was not a problem with network denoiser but could cause potential issues with model-driven unbiased estimator as denoisers.

Wavelet transformation involves a set of unitary operators and are not expected to change the behavior with respect to metric, in principle. The positive difference in performance introduced by using wavelet may be attributed to its *explicit* help with feature encoding for the denoising network. Explicit guidance on structure or feature design has shown benefit in deep learning performance. Study has proven the importance of effectively incorporating prior knowledge and feature fusion techniques to enhance the encoding capabilities of deep learning models, making them more robust and capable of handling diverse and complex tasks[15,16].

While sophisticated CBCT reconstruction methods, such as dual-domain regularized reconstruction offer potentially higher accuracy by leveraging both image and sinogram domains, they typically demand considerably higher computational resources and more complex tuning. These methods, although robust, can pose significant barriers to clinical deployment due to their complexity. Our approach prioritizes ease of use and modularity. This alignment with the practical needs of clinical environments ensures that it fits seamlessly into existing workflows. As a result, it offers an excellent balance of performance and usability, making it a commendable choice for clinical settings. This makes it an attractive option for clinical efficiency and effectiveness.

The general rationale of metric learning followed by a denoiser network is applicable towards various CBCT enhancement applications. However, the specific optimal weights for metric transformation and denoiser network coefficient can be site specific.

In summary, the proposed MEGATRON makes the conscious compromise to use a simple, lightweight, modular, and clinically friendly pipeline, in lieu of much more sophisticated dual domain regularized reconstruction or networks. It allows for easy offline appreciation of the metric relation and allows for plug-and-play of various modular components. Within the realm of clinical utility, we feel such decision and pipelined design could be appealing.

Acknowledgement

The authors would like to thank Drs. Bayat and Altunbas at the University of Colorado for motivating this development, constructive discussions, and continuous support. We thank Dr. Yuanwei He for help with proofreading.

References

1. Zhu L, Wang J, Xing L. Noise suppression in scatter correction for cone-beam CT. *Med Phys*. Mar 2009;36(3):741-52. doi:10.1118/1.3063001
2. Maggioni M, Katkovnik V, Egiazarian K, Foi A. Nonlocal Transform-Domain Filter for Volumetric Data Denoising and Reconstruction. *IEEE Transactions on Image Processing*. 2013;22(1):119-133. doi:10.1109/TIP.2012.2210725
3. Zhao S, Wang G. Feldkamp-type cone-beam tomography in the wavelet framework. *IEEE Trans Med Imaging*. Sep 2000;19(9):922-9. doi:10.1109/42.887839
4. Jia X, Dong B, Lou Y, Jiang SB. GPU-based iterative cone-beam CT reconstruction using tight frame regularization. *Physics in Medicine & Biology*. 2011;56(13):3787.
5. Cruz-Bastida JP, Moncada F, Martinez-Davalos A, Rodriguez-Villafuerte M. Task-based transferable deep-learning scatter correction in cone beam computed tomography: a simulation study. *J Med Imaging (Bellingham)*. Mar 2024;11(2):024006. doi:10.1117/1.JMI.11.2.024006
6. Piao Z, Deng W, Huang S, et al. Adaptive scatter kernel deconvolution modeling for cone-beam CT scatter correction via deep reinforcement learning. *Med Phys*. Feb 2024;51(2):1163-1177. doi:10.1002/mp.16618
7. Xu C, Yang B, Guo F, Zheng W, Poignet P. Sparse-view CBCT reconstruction via weighted Schatten p-norm minimization. *Optics Express*. 2020;28(24):35469-35482.
8. Bayat F, Ruan D, Miften M, Altunbas C. A quantitative CBCT pipeline based on 2D antiscatter grid and grid‐based scatter sampling for image‐guided radiation therapy. *Medical Physics*. 2023;50(12):7980-7995.
9. Joseph J, Biji I, Babu N, et al. Fan beam CT image synthesis from cone beam CT image using nested residual UNet based conditional generative adversarial network. *Phys Eng Sci Med*. 2023;46(2):703-717.
10. Shahin A, Wegworth C, Estes E, et al. Data from: OSIC Pulmonary Fibrosis Progression. 2020. doi:Kaggle. https://kaggle.com/competitions/osic-pulmonary-fibrosis-progression
11. Lee G, Gommers R, Waselewski F, Wohlfahrt K, O'Leary A. PyWavelets: A Python package for wavelet analysis. *Journal of Open Source Software*. 2019;4(36):1237.
12. Maggioni M, Katkovnik V, Egiazarian K, Foi A. Nonlocal transform-domain filter for volumetric data denoising and reconstruction. *IEEE transactions on image processing*. 2012;22(1):119-133.
13. Liu D, Wen B, Liu X, Wang Z, Huang TS. When image denoising meets high-level vision tasks: A deep learning approach. *arXiv preprint arXiv:170604284*. 2017;
14. Wu M, FitzGerald P, Zhang J, et al. XCIST—an open access x-ray/CT simulation toolkit. *Physics in Medicine & Biology*. 2022;67(19):194002.
15. Mungoli N. Adaptive feature fusion: enhancing generalization in deep learning models. *arXiv preprint arXiv:230403290*. 2023;
16. Chen S, Guo W. Auto-encoders in deep learning—a review with new perspectives. *Mathematics*. 2023;11(8):1777.