
Solving Inverse Problem Using Physics-Constrained Machine Learning

Junbo Peng

Georgia Institute of Technology

junbo.peng@gatech.edu

Abstract

Computed tomography (CT) is extensively employed in medical diagnosis and treatment guidance. However, when patients carry metal implants, the reconstructed CT images often suffer from severe artifacts caused by beam hardening effect. In such cases, the acquired projection data violate Tuy’s data sufficiency condition, making analytical reconstruction infeasible. To address this, a wide range of deep learning-based metal artifact reduction (MAR) methods have been developed, demonstrating impressive performance. Yet, existing approaches are purely data-driven and depend on large collections of reference images for manifold approximation. In this work, we introduce a physics-driven sinogram inpainting approach that leverages the inherent correlations in projection data during CT acquisition for MAR.

1 Introduction

Metal artifact reduction (MAR) in CT reconstruction has long been a challenging problem in medical imaging. Owing to the polychromatic nature of X-ray beams, photons transmitted through metal implants are significantly hardened compared to those traversing soft tissue or bone, resulting in severe artifacts in the reconstructed CT images. Fundamentally, MAR is a data-incomplete reconstruction problem: only a subset of projection data can be reliably acquired, violating Tuy’s data sufficiency condition for analytical reconstruction [1]. Under such circumstances, the linear imaging model becomes ill-posed, and the reconstructed images exhibit pronounced structural distortions and artifacts [2].

Conventional MAR strategies attempt to address this ill-posedness by modifying the corrupted sinogram data. A common approach is interpolation within the metal trace of the sinogram [3], while others attempt to correct for beam hardening using knowledge of the X-ray spectrum and material properties of the implants [4]. However, the success of interpolation-based methods is highly dependent on the sampling pattern of the acquired projections, and beam hardening correction (BHC) becomes ineffective when the attenuation is excessive or photon starvation occurs [5].

In recent years, deep learning-based approaches have gained considerable attention for MAR and other data-incomplete CT problems, such as sparse-view and limited-angle CT. These methods implicitly approximate the target image manifold using neural networks, achieving superior performance compared to traditional approaches[2, 6–10]. Nonetheless, all existing models are purely data-driven, relying heavily on large-scale reference datasets, while the underlying physics of CT acquisition is largely underutilized.

In this work, we introduce a physics-driven sinogram manifold learning framework. Our method explicitly characterizes the physical properties of the sinogram manifold while learning its distribution in an implicit form via a neural network. This constitutes an exact manifold learning strategy that does not require label supervision. We demonstrate the feasibility of the proposed approach in the

MAR task, and emphasize that the framework is readily extensible to other data-incomplete CT reconstruction problems without modification.

2 Methodology and preliminaries

2.1 CT imaging and reconstruction

In general, the CT imaging process can be formulated as a linear system:

$$\vec{b} = F\vec{\mu}, \quad (1)$$

where F denotes the forward projection operator (system matrix), $\vec{\mu}$ is the linear attenuation coefficient (LAC) map, and \vec{b} is the measured sinogram. Ideally, if \vec{b} satisfies Tuy's data sufficiency condition, the image $\vec{\mu}$ can be reconstructed from \vec{b} using the filtered backprojection (FBP) algorithm as [1]

$$\vec{\mu} = FBP(\vec{b}). \quad (2)$$

If Equation 1 admits a unique solution, the FBP operator is equivalent to the exact inverse of the system matrix, i.e.,

$$\vec{\mu} = FBP(\vec{b}) = F^{-1}\vec{b}. \quad (3)$$

In practice, however, Equation 1 is often overdetermined due to noise and thus has no exact solution. In this case, the FBP operator corresponds to the Moore–Penrose pseudoinverse of the system matrix:

$$\vec{\mu} = FBP(\vec{b}) = F^\dagger \vec{b}, \quad (4)$$

which is equivalent to solving a least-squares problem, as commonly done in iterative CT reconstruction methods.

2.2 Incomplete CT data acquisition

In incomplete CT data acquisition, Equation 1 becomes

$$\vec{b} = \begin{bmatrix} \vec{b}_M \\ \vec{b}_\emptyset \end{bmatrix} = \begin{bmatrix} (F\vec{\mu})_M \\ (F\vec{\mu})_\emptyset \end{bmatrix} = \begin{bmatrix} F_M \vec{\mu} \\ F_\emptyset \vec{\mu} \end{bmatrix} \quad (5)$$

where $\vec{b}_{M/\emptyset}$ denotes the measured or missing sinogram data, and $F_{M/\emptyset}$ represents the corresponding row blocks of the system matrix. In other words, applying the operator F_M to the CT image $\vec{\mu}$ yields the incomplete sinogram \vec{b}_M , while the residual sinogram \vec{b}_\emptyset is zero in the acquired data. It should be noted that the missing data \vec{b}_\emptyset do not necessarily form a contiguous block in the complete sinogram \vec{b} . However, by reordering the rows of F and \vec{b} , the formulation in Equation 5 can be applied, after which the rows must be reordered back before performing the FBP operator. In the context of MAR, the missing data \vec{b}_\emptyset correspond to the metal trace in the sinogram, as indicated by the red arrow in Figure 1.

By directly applying the FBP operator to the acquired sinogram, one obtains the reconstructed image as

$$\vec{\mu}^* = FBP\left(\begin{bmatrix} \vec{b}_M \\ \vec{0} \end{bmatrix}\right) = F^\dagger\left(\begin{bmatrix} \vec{b}_M \\ \vec{0} \end{bmatrix}\right) \quad (6)$$

where the reconstructed image $\vec{\mu}^*$ suffers from severe structural distortions and artifacts, since there does not exist an image $\vec{\mu}$ whose sinogram equals $[\vec{b}_M \quad \vec{0}]^T$. In other words, $[\vec{b}_M \quad \vec{0}]^T$ is not a valid sinogram and therefore does not belong to the sinogram manifold.

2.3 Physics constraint of the sinogram manifold

A natural and fundamental question arises: what type of data constitutes a sinogram? In other words, how can one determine whether a given image is a valid sinogram or not? This question is essentially equivalent to characterizing the manifold of sinogram data.

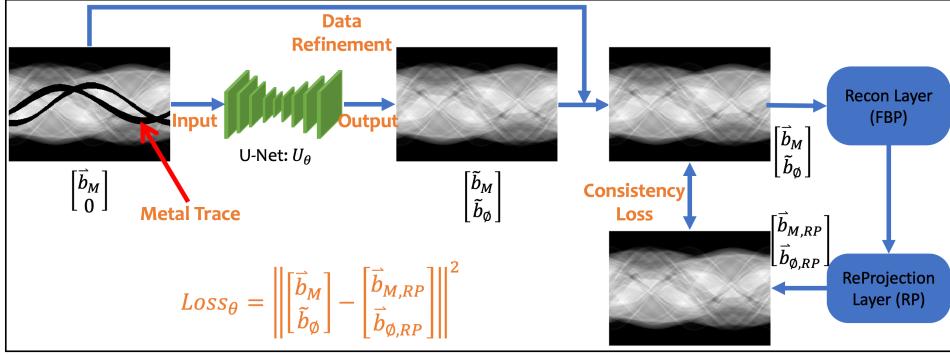


Figure 1: The proposed physics-constrained sinogram inpainting for MAR.

The answer follows from the properties of the system matrix F and its Moore–Penrose pseudoinverse F^\dagger . In practice, the columns of F are linearly independent, since each column corresponds to a distinct pixel in $\vec{\mu}$ that is theoretically independent. Under this condition, the pseudoinverse F^\dagger acts as a left inverse of F , such that

$$FF^\dagger = I, \quad (7)$$

while in general

$$F^\dagger F \neq I. \quad (8)$$

This non-commutative property explains why FBP reconstruction succeeds with complete data (F followed by F^\dagger), but fails when applied to incomplete data (F^\dagger followed by F). Exploiting this property, one can verify whether an estimated \tilde{b} is a valid sinogram by checking whether the cascade operator FF^\dagger leaves it unchanged, i.e.,

$$FF^\dagger \tilde{b} = \tilde{b} \quad \text{if } \exists \tilde{\mu} \text{ such that } F\tilde{\mu} = \tilde{b}. \quad (9)$$

Indeed,

$$FF^\dagger \tilde{b} = FF^\dagger F\tilde{\mu} = F\tilde{\mu} = \tilde{b}, \quad (10)$$

where $\tilde{\mu}$ denotes the CT image corresponding to \tilde{b} , provided \tilde{b} is a sinogram. If \tilde{b} is not a sinogram, then $FF^\dagger \tilde{b}$ will deviate significantly from \tilde{b} .

2.4 Proposed physics-constrained sinogram inpainting

With the sinogram manifold property explicitly formulated in Eq. 9, we propose an unsupervised manifold learning framework for sinogram completion in data-incomplete CT imaging. The consistency loss between the refined estimated sinogram $[\tilde{b}_M | \tilde{b}_\emptyset]^T$ and the corresponding re-projected (RP) sinogram $[\tilde{b}_M^{RP} | \tilde{b}_\emptyset^{RP}]^T$ serves as the sole loss function, enforcing the manifold constraint on the model output. This guarantees that the generated sinogram is a valid data point on the sinogram manifold. During the data refinement process, the measured entries are preserved by replacing the estimated values with the acquired data, thereby enforcing measurement consistency throughout the generation process, as illustrated in Figure 1. Ideally, the model completes the sinogram data with a perfect estimation of \tilde{b}_\emptyset , such that there exists an image $\tilde{\mu}$ whose sinogram is $[\tilde{b}_M | \tilde{b}_\emptyset]^T$. The image $\tilde{\mu}$ may be unique or non-unique, depending on the specific problem.

3 Experiments

3.1 Datasets and experimental settings

Following prior studies in MAR, metal implants were simulated on real patient CT images to evaluate the proposed sinogram manifold learning method. Patient CT scans and corresponding metal implant masks were obtained from the AAPM CT-MAR Challenge dataset [11]. For the brain study, 1,300

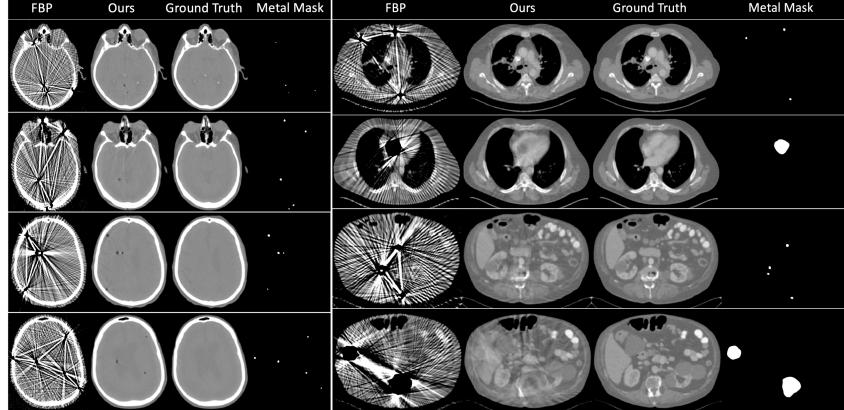


Figure 2: Results of the study of brain and abdominal patients. Display window: [-500 500] HU.

slices were used for training and 227 slices for testing. For the lung and abdomen study, 2,268 slices were used for training and 494 slices for testing.

All experiments were implemented in the PyTorch framework and conducted on a 40 GB NVIDIA A100 GPU. The sinograms had a resolution of 256×216 , while the reconstructed CT images were of size 512×512 . A residual U-Net with attention blocks was adopted for sinogram inpainting. The Adam optimizer was employed with a learning rate of 1×10^{-5} and β parameters set to 0.5 and 0.999, respectively. The batch size was fixed at 2 for all experiments. In addition, customized differentiable forward projection (F) and FBP reconstruction (F^\dagger) layers were integrated into the training pipeline.

For quantitative evaluation, mean absolute error (MAE), peak signal-to-noise ratio (PSNR), and normalized correlation coefficient (NCC) were computed. The dynamic range for PSNR calculation was defined based on the ground-truth CT images, restricted to regions where tissue structures were visible.

3.2 Results

The effect of the proposed method on the reduction of metal artifacts is presented in Figure 2 for the study of brain, lung, and abdomen patients. The first column lists directly reconstructed CT images from metal-corrupted sinograms using the FBP algorithm. The second and third columns are the MAR results by the proposed method and the ground-truth CT images. Masks of metal implants are attached in the last column for reference. Impressively, the proposed physics-constrained sinogram manifold learning method achieves superior performance in MAR without any reference data for the manifold approximation. In the brain study, the proposed method reduced the HU error from an MAE of 157.89 HU with FBP reconstruction to 7.62 HU with sinogram restoration. Correspondingly, PSNR and NCC improved from 10.06 dB and 0.55 (FBP) to 43.34 dB and 0.99, respectively. For the lung and abdomen study, the MAE decreased from 182.29 HU (FBP) to 10.96 HU, while PSNR and NCC increased to 40.51 dB and 0.99.

4 Conclusion

In this work, we presented a physics-constrained sinogram inpainting method for MAR. The key contribution lies in introducing a physics-based constraint to determine whether a generated image is a valid sinogram, by exploiting the non-commutative property between projection and reconstruction in CT imaging. Unlike previously reported approaches, the proposed method is entirely physics-driven and does not require labeled complete sinograms for model training. The feasibility of the method was verified through studies on brain as well as lung and abdomen patients, demonstrating its effectiveness in MAR tasks. One limitation of this work is that image-domain information was not incorporated into the sinogram inpainting process, which will be the focus of our future research.

Acknowledgments and Disclosure of Funding

References

- [1] Frank Natterer. *The mathematics of computerized tomography*. SIAM, 2001.
- [2] Wei-An Lin, Haofu Liao, Cheng Peng, Xiaohang Sun, Jingdan Zhang, Jiebo Luo, Rama Chellappa, and Shaohua Kevin Zhou. Dudonet: Dual domain network for ct metal artifact reduction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10512–10521, 2019.
- [3] Willi A Kalender, Robert Hebel, and Johannes Ebersberger. Reduction of ct artifacts caused by metallic implants. *Radiology*, 164(2):576–577, 1987.
- [4] Hyoung Suk Park, Dosik Hwang, and Jin Keun Seo. Metal artifact reduction for polychromatic x-ray ct based on a beam-hardening corrector. *IEEE transactions on medical imaging*, 35(2):480–487, 2015.
- [5] Lequan Yu, Zhicheng Zhang, Xiaomeng Li, and Lei Xing. Deep sinogram completion with image prior for metal artifact reduction in ct images. *IEEE transactions on medical imaging*, 40(1):228–238, 2020.
- [6] Hyoung Suk Park, Sung Min Lee, Hwa Pyung Kim, Jin Keun Seo, and Yong Eun Chung. Ct sinogram-consistency learning for metal-induced beam hardening correction. *Medical physics*, 45(12):5376–5384, 2018.
- [7] Xia Huang, Jian Wang, Fan Tang, Tao Zhong, and Yu Zhang. Metal artifact reduction on cervical ct images by deep residual learning. *Biomedical engineering online*, 17:1–15, 2018.
- [8] Yanbo Zhang and Hengyong Yu. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE transactions on medical imaging*, 37(6):1370–1381, 2018.
- [9] Haofu Liao, Wei-An Lin, Zhimin Huo, Levon Vogelsang, William J Sehnert, S Kevin Zhou, and Jiebo Luo. Generative mask pyramid network for ct/cbct metal artifact reduction with joint projection-sinogram correction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, pages 77–85. Springer, 2019.
- [10] Hong Wang, Yuexiang Li, Haimiao Zhang, Jiawei Chen, Kai Ma, Deyu Meng, and Yefeng Zheng. Indudonet: An interpretable dual domain network for ct metal artifact reduction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pages 107–118. Springer, 2021.
- [11] American Association of Physicists in Medicine. Ct metal artifact reduction (ct–mar). <https://www.aapm.org/GrandChallenge/CT-MAR/>, 2024. Accessed: 2025-08-29.