https://doi.org/10.2352/EI.2025.37.12.HPCI-192

SinoTx: A Transformer-based Model for Sinogram Inpainting

***Jiaze E*[1]*, Zhengchun Liu*[2]*, Tekin Bicer*[2]*, Srutarshi Banerjee*[2]*, Rajkumar Kettimuthu*[2]*, Bin Ren*[1]*, Ian T. Foster*[2]**
[1] ***William & Mary, Williamsburg, VA, USA***
[2] ***Argonne National Laboratory, Lemont, IL, USA***

Abstract

Sinogram inpainting is a critical task in computed tomography (CT) imaging, where missing or incomplete sinograms can significantly decrease image reconstruction quality. High-quality sinogram inpainting is essential for achieving high-quality CT images, enabling better diagnosis and treatment. To address this challenge, we propose SinoTx, a model based on the Transformer architecture specifically designed for sinogram completion. SinoTx leverages the inherent strengths of Transformers in capturing global dependencies, making it well-suited for handling the complex patterns present in sinograms. Our experimental results demonstrate that SinoTx outperforms existing baseline methods, achieving up to a 32.3% improvement in the Structural Similarity Index (SSIM) and a 44.2% increase in Peak Signal-to-Noise Ratio (PSNR).

Introduction

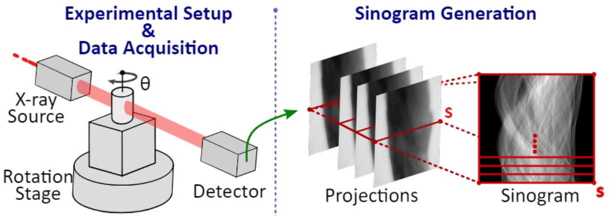

Figure 1: Schematic of CT Imaging and Sinogram Generation: The illustration depicts the CT imaging process, starting with the experimental setup for X-ray projection data acquisition, where a sample rotates to capture projections from multiple angles. These projections are then used to generate a sinogram, a 2D representation that combines all angular projections, critical for reconstructing cross-sectional images of the sample.

Computed tomography (CT) imaging is a cornerstone of both clinical diagnostics and industrial material analysis. At its core, CT involves the collection of X-ray projection data from multiple angles to reconstruct cross-sectional images of an object. These reconstructed images provide detailed insights into internal structures, which are critical for applications ranging from medical diagnosis to material defect detection. Fig 1 illustrates the principle of CT imaging, showing the process of data acquisition through X-ray projections and the subsequent generation of a sinogram from angular projections for image reconstruction.

However, the process of acquiring complete and high-quality sinograms—a 2D representation of these projections—is often hindered by practical constraints such as limited radiation dosage, restricted angular coverage, or mechanical instabilities during data acquisition. These limitations can result in incomplete or corrupted sinograms, significantly degrading the quality of reconstructed images [1].

Addressing the challenges of incomplete sinograms is crucial for advancing CT imaging. Traditional interpolation-based methods for inpainting often fail to capture the complex patterns and dependencies inherent in sinogram data [2]. hese traditional approaches often assume simplistic data distributions and lack the ability to effectively capture global dependencies in sinogram data, which are critical for high-quality reconstructions. Furthermore, such methods frequently require manual parameter tuning and cannot generalize well to diverse imaging conditions [3, 4].

A pressing need exists to address the limitations of traditional methods, which often rely on oversimplified assumptions and struggle with generalization across diverse imaging scenarios. Recent advancements in deep learning techniques provide an opportunity to overcome these challenges effectively. While convolutional neural networks (CNNs) have demonstrated some success in addressing local patterns in sinogram data, they often fail to capture the global dependencies critical for reconstructing missing or irregular sinogram segments. Furthermore, many ML-based approaches rely heavily on large labeled datasets and struggle to generalize well to the high variability and sparsity characteristic of real-world sinogram data. Recent advancements in deep learning, particularly the emergence of Transformer-based architectures, have transformed various domains by effectively modeling complex data dependencies. Originally developed for natural language processing (NLP) tasks, Transformers have demonstrated remarkable success in computer vision applications, leveraging their self-attention mechanisms to capture both local and global relationships in data [5, 6]. Inspired by this paradigm, researchers have begun exploring the potential of Transformer models for scientific imaging tasks, including CT sinogram inpainting. Prior studies on SinoTx show promising results by treating sinograms as analogous to sequential data, where each projection corresponds to a token in a sequence [7, 8, 12]. This approach aligns with the concept of masked modeling, commonly employed in NLP to reconstruct missing tokens, thereby enabling effective inpainting of sinograms.

Building upon these foundations, we propose SinoTx, a Transformer-based model specifically designed for sinogram inpainting. SinoTx leverages the inherent strengths of Transformers in capturing global dependencies, making it particularly well-suited for the irregular and sparse patterns characteristic of missing sinogram data. Unlike convolutional neural networks (CNNs), which primarily focus on local features, SinoTx employs self-attention mechanisms to model long-range interactions across the entire sinogram. This capability is critical for accurately reconstructing missing data and preserving structural integrity in CT reconstructions.

Specifically, this work consists of three main contributions as follows:

- Model Architecture: We introduce SinoTx, a specialized Transformer model tailored for sinogram inpainting, incorporating both an encoder and a decoder designed to handle the unique characteristics of sinogram data.
- Scalability and Efficiency: SinoTx integrates Distributed Data Parallelization (DDP) using PyTorch, enabling efficient training and inference across large-scale synchrotron datasets in high-performance computing (HPC) environments [bommasani2021opportunities].
- Empirical Validation: Extensive experiments on simulated and real-world datasets demonstrate that SinoTx achieves superior performance compared to state-of-the-art methods, with significant improvements in Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR) [10, 11].

By addressing the challenges of sinogram inpainting with a Transformer-based approach, this work performs better than baselines and sets a new benchmark for future innovations in the field. The proposed SinoTx model not only enhances the quality of CT reconstructions but also exemplifies the potential of applying advanced deep learning techniques to scientific imaging tasks, bridging the gap between cutting-edge AI and practical applications in synchrotron facilities.

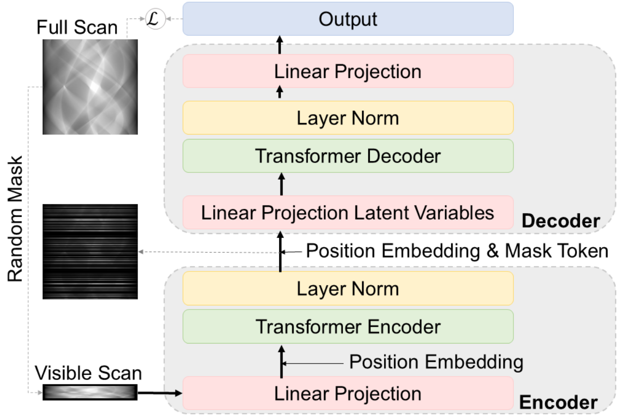

Figure 2: Overview of SinoTx: The overview highlights the encoder-decoder framework for sinogram inpainting. The encoder processes visible sinogram projections through linear projection, position embedding, and transformer layers, generating latent variables. The decoder inpaints the full sinogram by incorporating masked tokens and leveraging transformer blocks, with a final loss computed against the original full scan.

SinoTx Design and Analysis

To tackle the challenges of sinogram inpainting in computed tomography (CT), we propose SinoTx, a Transformer-based architecture designed to capture both local and global dependencies within sinogram data. By leveraging advanced self-attention mechanisms, SinoTx effectively inpaints incomplete or corrupted sinograms, which also enables high-quality CT image reconstructions even under sparse-view or low-dose scenarios. An overview of SinoTx is shown in Fig 2.

Model Design

The SinoTx model consists of two main components: the Encoder and the Decoder, both built upon the Transformer architecture.

Encoder

The encoder processes the input sinogram data by first embedding it into a high-dimensional space using a convolutional layer:

$$\mathbf{E} = ProjEmb(\mathbf{S}), \tag{1}$$

where **S** denotes the input sinogram and **E** represents the embedded projections. This embedding is then augmented with positional encodings **P** to preserve the sequential structure of the sinogram projections:

$$\mathbf{E}' = \mathbf{E} + \mathbf{P}. \tag{2}$$

A series of self-attention layers within the encoder then extracts meaningful global and local dependencies, producing a robust representation of the sinogram data:

$$\mathbf{Z} = \mathscr{E}(\mathbf{E}'), \tag{3}$$

where **Z** is the encoded representation. During training, a random subset of sinogram projections is masked to simulate missing data, and the encoder is tasked with learning representations that are invariant to these missing elements.

Decoder

The decoder reconstructs the missing projections by leveraging the encoded representations and filling in the masked positions. To achieve this, mask tokens are appended to the encoded features at the positions corresponding to the missing sinogram data. The Transformer decoder, composed of multiple attention layers, learns to recover the missing values by modeling the relationships between the observed and masked projections. Finally, the reconstructed sinogram is projected back to its original resolution using a linear transformation:

$$\mathbf{S}_{\mathbf{inpaint}} = \mathscr{D}(\mathbf{Z}). \tag{4}$$

Loss Function

The model is trained using a Mean Squared Error (MSE) loss function to minimize the difference between the reconstructed and ground truth sinograms:

$$\mathscr{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N} ||\mathbf{S}_{\mathbf{inpaint},\mathbf{i}} - \mathbf{S}_{\mathbf{fullscan},\mathbf{i}}||_2^2, \tag{5}$$

where N is the number of samples in the training set.

Training

Graphics and equations should fit within one column (3.38 inches wide), but full width (7 inch) figures are also acceptable. Equations, figures and figure captions each have their own style tags. Equations are numbered using parentheses flushed right as shown below.

During training, 80% of the sinogram projections are randomly masked to mimic real-world scenarios of incomplete data.

Table 1: Quantitative Results: The table presents quantitative accuracy results for the SinoTx model across three datasets, showcasing the model's performance in various inpainting scenarios. The values in parentheses correspond to the results of the reconstructed images.

(a) `Shepp2D`: This subtable shows results for `Shepp2D`, which evaluates the model's performance on geometric phantoms.

Methods	SSIM			PSNR		
	Mask Ratio			Mask Ratio		
	0.4	0.6	0.8	0.4	0.6	0.8
SinoTx	**0.967** (0.938)	**0.968** (0.926)	**0.962** (0.906)	**37.2** (35.8)	**37.5** (34.4)	**36.3** (33.9)
UsiNet [20]	0.494 (0.447)	0.499 (0.462)	0.492 (0.452)	13.1 (12.1)	15.0 (13.2)	15.6 (14.5)
StrDiffusion [21]	0.516 (0.497)	0.554 (0.505)	0.562 (0.538)	15.5 (12.5)	16.3 (14.9)	16.7 (15.3)
CMT [22]	0.504 (0.457)	0.507 (0.477)	0.518 (0.485)	20.8 (19.3)	22.6 (19.3)	22.7 (19.5)
MISF [23]	0.731 (0.702)	0.749 (0.713)	0.743 (0.720)	25.8 (24.1)	26.5 (23.7)	26.9 (25.4)
LaMa [11]	0.698 (0.663)	0.708 (0.686)	0.720 (0.692)	25.0 (23.6)	25.8 (23.0)	26.6 (23.6)

(b) `Shape`: The Shape dataset highlights the model's ability to inpaint sinograms containing complex geometric structures.

Methods	SSIM			PSNR		
	Mask Ratio			Mask Ratio		
	0.4	0.6	0.8	0.4	0.6	0.8
SinoTx	0.746 (0.717)	0.743 (0.696)	**0.796** (0.750)	**26.5** (25.1)	26.9 (25.2)	27.1 (25.1)
UsiNet [20]	0.509 (0.480)	0.549 (0.511)	0.546 (0.512)	14.1 (10.9)	13.7 (12.6)	15.3 (12.8)
StrDiffusion [21]	0.524 (0.497)	0.552 (0.506)	0.600 (0.572)	14.3 (12.1)	15.1 (13.5)	15.9 (13.5)
CMT [22]	0.454 (0.413)	0.469 (0.427)	0.477 (0.431)	19.9 (17.1)	20.2 (17.4)	21.0 (18.1)
MISF [23]	0.716 (0.692)	0.739 (0.717)	0.753 (0.735)	26.4 (25.3)	27.6 (26.5)	27.8 (25.0)
LaMa [11]	**0.747** (0.718)	**0.762** (0.730)	0.779 (0.726)	26.1 (23.9)	**27.8** (25.8)	**28.6** (27.8)

(c) `Real-world`: `Real-world` captures the complexities of real CT imaging scenarios. The results suggest potential areas for improvement in real-world applicability.

Methods	SSIM			PSNR		
	Mask Ratio			Mask Ratio		
	0.4	0.6	0.8	0.4	0.6	0.8
SinoTx	0.572 (0.556)	0.597 (0.544)	0.643 (0.609)	20.4 (18.2)	21.7 (19.5)	21.9 (20.2)
UsiNet [20]	0.460 (0.416)	0.482 (0.431)	0.494 (0.448)	12.8 (10.7)	14.5 (11.5)	14.9 (12.6)
StrDiffusion [21]	0.462 (0.432)	0.508 (0.465)	0.507 (0.474)	13.7 (10.8)	15.7 (14.2)	16.2 (12.4)
CMT [22]	0.436 (0.392)	0.440 (0.407)	0.453 (0.406)	16.1 (14.8)	16.2 (13.2)	16.1 (14.9)
MISF [23]	**0.706** (0.671)	**0.717** (0.664)	**0.723** (0.699)	**25.1** (23.3)	**25.2** (23.5)	**26.4** (25.1)
LaMa [11]	**0.706** (0.672)	0.716 (0.683)	0.715 (0.666)	22.1 (20.6)	22.9 (20.3)	24.6 (22.1)

The input sinogram data is first preprocessed and embedded into a high-dimensional space using a linear projection layer. Positional encodings are then added to the embeddings to retain structural information. The masked sinogram is passed through the Transformer encoder, where both local and global features are extracted. The decoder takes these encoded representations, fills in the missing projections using mask tokens, and reconstructs the full sinogram.

SinoTx is pretrained on a large-scale simulated dataset to capture generalizable features and patterns in sinogram data. After pretraining, the model can be fine-tuned for specific tasks, such as sparse-view or low-dose inpainting, to ensure task-specific performance. The optimization is performed using the Adam optimizer with a learning rate scheduler, promoting stable and efficient convergence. To address the computational demands of handling large-scale datasets, SinoTx is implemented with Distributed Data Parallelism (DDP) in PyTorch, allowing efficient training across multiple GPUs. Finally, the reconstructed sinogram is used to produce CT images via the inverse Radon [13] and Gridrec [14].

Analysis

SinoTx offers several advantages over traditional methods. Its self-attention mechanisms enable the model to capture long-range dependencies, making it particularly effective for inpainting irregular or sparse patterns in sinogram data. Additionally, the Transformer architecture is highly scalable, allowing SinoTx to handle large datasets efficiently using parallel computing frameworks. The flexibility of the model makes it adaptable to diverse scenarios, including sparse-view and low-dose inpaintings, achieving high levels of robustness and accuracy. Moreover, the pretraining-finetuning paradigm ensures that the model learns both general and task-specific features, enhancing its versatility in various CT related tasks.

Evaluation

To evaluate the performance of SinoTx, we conduct extensive experiments on both simulated and real-world datasets, comparing our method against several baselines. The evaluation metrics include Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR), which quantitatively and qualitatively assess the inpainting of sinograms and the reconstruction of final

CT images.

Implementation Details

We train our model using the Polaris supercomputer at Argonne Leadership Computing Facility (ALCF) resources at Argonne National Laboratory (ANL)[1]. Polaris supercomputer consists of 560 compute nodes, each of which is equipped with 4 NVIDIA A100 GPUs (connected via NVLink). We use PyTorch version 2.1.0 and CUDA version 12.2. The dimensions of the input images are 512x512.

Dataset

Three datasets are used in our experiments, each containing 100k samples with a resolution of 512×512:

`Shepp2d` dataset, generated using the TomoPy library [15], comprises simulated Shepp-Logan phantoms. These phantoms are widely recognized as a standard benchmark in tomographic imaging, providing a controlled environment for evaluating the accuracy and stability of reconstruction algorithms. The dataset's characterized by well-defined geometric structures, makes it ideal for assessing the model's fundamental inpainting capabilities under varying masking ratios.

`Shape` dataset, created using the scikit-image library [16], contains a diverse collection of simulated geometric shapes, including circles, rectangles, triangles, and ellipses. By introducing a range of geometric complexities, such as sharp edges and varying curvatures, this dataset challenges the model to accurately reconstruct intricate structural details. Its diversity allows for a comprehensive evaluation of the model's ability to handle different spatial patterns and topological variations.

`Real-world` dataset [17] consists of sinograms derived from actual synchrotron radiation CT scans, including data collected at facilities such as the Advanced Photon Source (APS[2]) at Argonne National Laboratory. These sinograms capture the complexities of real-world imaging, encompassing diverse materials and varying noise levels. Additionally, this dataset incorporates dynamic [18] and in situ [19] systems, which represent the challenges associated with real-time and experimental imaging scenarios. Its inherent complexity and variability provide a robust benchmark for evaluating the model's performance in practical sinogram inpainting tasks.

Quantitative Analysis

The performance of SinoTx was first assessed on the `Shape2d`, a standard benchmark for tomographic imaging. SinoTx demonstrated significant improvements over baseline methods across all mask ratios. Specifically, SinoTx achieved up to a 32.3% improvement in SSIM and a 44.2% increase in PSNR compared to baselines. As shown in Table 1a, SinoTx consistently outperformed all baselines under varying mask ratios, highlighting its robustness and effectiveness.

On the `Shape` dataset, SinoTx delivered competitive results, often comparable or superior to baselines, particularly at higher mask ratios. This dataset, with its diverse geometric patterns, allowed SinoTx to demonstrate its ability to handle sharp edges and complex curvatures effectively. The results, summarized in Table 1b, indicate that SinoTx maintains strong performances for `Shape`.

For the `Real-world` dataset, SinoTx exhibited reasonable generalization capabilities but slightly lagged behind methods like LaMa and MISF, especially when addressing highly complex and noisy sinogram structures. As shown in Table 1c, SinoTx's performance in real-world scenarios is promising but suggests the need for further enhancements to better adapt to the variability and challenges inherent in real-world data.

Qualitative Analysis

In addition to quantitative metrics, we present the inpainted images generated by SinoTx compared to other baselines. The results shwon in Fig 3 demonstrate that SinoTx can effectively recover fine details in the inpainted sinograms in regions with high sparsity or irregular patterns. For example, SinoTx's inpaintings closely resemble the ground truth.

Conclusion

In this study, we introduce SinoTx, a Transformer-based model specifically designed for sinogram inpainting in computed tomography. By leveraging the inherent strengths of Transformers in capturing global dependencies, SinoTx demonstrate significant improvements over baselines in both SSIM and PSNR across multiple datasets. Its ability to handle missing and sparse sinogram data is validated on simulated datasets, and its scalability is enabled through Distributed Data Parallelism in PyTorch, making it well-suited for large-scale HPC environments. While SinoTx exhibits strong performance in synthetic settings, further optimization is needed to enhance its generalization to real-world data. These findings establish SinoTx as a robust framework for sinogram inpainting and provide a foundation for future advancements in CT imaging and high-performance computing applications.

References

[1] J. Radon, "On the determination of functions from their integral values along certain manifolds," *IEEE Transactions on Medical Imaging*, vol. 5, no. 4, pp. 170–176, 1986.

[2] Z. Liu, T. Bicer, R. Kettimuthu, D. Gursoy, F. De Carlo, and I. Foster, "Tomogan: low-dose synchrotron x-ray tomography with generative adversarial networks: discussion," *JOSA A*, vol. 37, no. 3, pp. 422–434, 2020.

[3] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16 000–16 009.

[4] D. M. Pelt, K. J. Batenburg, and J. A. Sethian, "Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks," *Journal of Imaging*, vol. 4, no. 11, p. 128, 2018.

[5] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems*, 2017.

[6] A. Dosovitskiy, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[7] M. Raghu and E. Schmidt, "A survey of deep learning for scientific discovery," *arXiv preprint arXiv:2003.11755*, 2020.

[8] Z. Liu, T. Bicer, R. Kettimuthu, and I. Foster, "Deep learning accelerated light source experiments," in *2019 IEEE/ACM Third Workshop*

[1] Polaris Cluster at ALCF: `https://www.alcf.anl.gov/polaris`

[2] Advanced Photon Source: `https://www.aps.anl.gov`

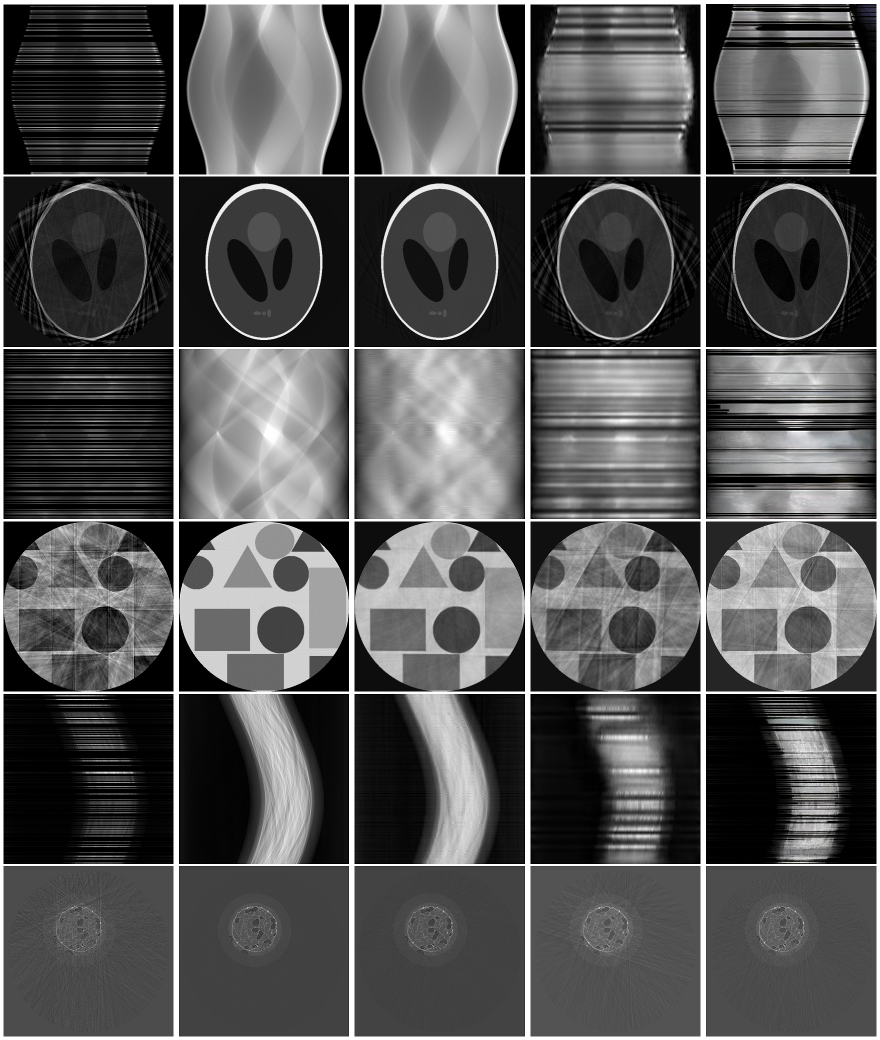

Figure 3: Qualitative Results: Lines 1, 3, 5 show visual results of sinogram inpainting of `Shepp2d`, `Shapes`, and `Real-world`, respectively. Lines 2, 4, 6 show the reconstructed images of the sinograms.

on Deep Learning on Supercomputers (DLS). IEEE, 2019, pp. 20–28.

[9] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill *et al.*, "On the opportunities and risks of foundation models," *arXiv preprint arXiv:2108.07258*, 2021.

[10] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.

[11] R. Suvorov, E. Logacheva, A. Mashikhin, A. Remizova, A. Ashukha, A. Silvestrov, N. Kong, H. Goka, K. Park, and V. Lempitsky, "Resolution-robust large mask inpainting with fourier convolutions," in *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 2022, pp. 2149–2159.

[12] Z. Liu, R. Kettimuthu, and I. Foster, "Masked sinogram model with transformer for ill-posed computed tomography reconstruction: a preliminary study," *arXiv preprint arXiv:2209.01356*, 2022.

[13] J. Radon, "1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten," *Classic papers in modern diagnostic radiology*, vol. 5, no. 21, p. 124, 2005.

[14] B. A. Dowd, G. H. Campbell, R. B. Marr, V. V. Nagarkar, S. V. Tipnis, L. Axe, and D. P. Siddons, "Developments in synchrotron x-ray computed microtomography at the national synchrotron light source," in *Developments in X-ray Tomography II*, vol. 3772. SPIE, 1999, pp. 224–236.

[15] D. Gürsoy, F. De Carlo, X. Xiao, and C. Jacobsen, "Tomopy: a framework for the analysis of synchrotron tomographic data," *Journal of synchrotron radiation*, vol. 21, no. 5, pp. 1188–1193, 2014.

[16] 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.

[17] F. De Carlo, D. Gürsoy, D. J. Ching, K. J. Batenburg, W. Ludwig, L. Mancini, F. Marone, R. Mokso, D. M. Pelt, J. Sijbers *et al.*, "Tomobank: a tomographic data repository for computational x-ray science," *Measurement Science and Technology*, vol. 29, no. 3, p. 034004, 2018.

[18] K. A. Mohan, S. Venkatakrishnan, J. W. Gibbs, E. B. Gulsoy, X. Xiao, M. De Graef, P. W. Voorhees, and C. A. Bouman, "Timbir: A method for time-space reconstruction from interlaced views," *IEEE Transactions on Computational Imaging*, vol. 1, no. 2, pp. 96–111, 2015.

[19] D. M. Pelt and K. J. Batenburg, "Fast tomographic reconstruction

from limited data using artificial neural networks," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5238–5251, 2013.

[20] L. Yao, Z. Lyu, J. Li, and Q. Chen, "No ground truth needed: unsupervised sinogram inpainting for nanoparticle electron tomography (usinet) to correct missing wedges," *npj Computational Materials*, vol. 10, no. 1, p. 28, 2024.

[21] H. Liu, Y. Wang, B. Qian, M. Wang, and Y. Rui, "Structure matters: Tackling the semantic discrepancy in diffusion models for image inpainting," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024, pp. 8038–8047.

[22] K. Ko and C.-S. Kim, "Continuously masked transformer for image inpainting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 13 169–13 178.

[23] X. Li, Q. Guo, D. Lin, P. Li, W. Feng, and S. Wang, "Misf: Multi-level interactive siamese filtering for high-fidelity image inpainting," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 1869–1878.

Author Biography

Jiaze E received her BE in Artificial Intelligence Science and Technology from Nanjing University of Science and Technology (2019) and her MS in Computer Science from The George Washington University (2021). She is now a PhD candidate majoring in Computer Science from William & Mary.