ISSN: 3079-6601

Structure Aware Metal Artifact Reduction for Computed Tomography

Luca Bernasconi[1]*, Martina Keller[2], Tobias Schmid[3]

Institute for Biomedical Engineering, ETH Zurich, 8092 Zurich, Switzerland

***Corresponding author:** l.bernasconi@ethz.ch

Abstract

Metal implants can cause severe artifacts in CT images, including streaking and photon starvation effects that degrade diagnostic visibility. Building on hybrid convolution–attention feature modeling exemplified by CTLformer, this paper introduces a residual learning approach for metal artifact reduction that combines image-domain denoising with sinogram-guided attention cues. The method uses a guidance branch derived from corrected sinograms to steer attention toward artifact-affected regions while preserving bone and soft-tissue edges. Evaluations on a clinical dataset of 4,800 scans (about 120,000 slices) show improvements of 0.9–1.5 dB in PSNR and 0.012–0.020 in SSIM compared with U-Net, CNN-based MAR, and transformer-only baselines. Artifact index metrics also decrease by 10%–18% in implant-adjacent regions.

Keywords

Metal artifact reduction; CT reconstruction; sinogram guidance; convolution–attention networks; clinical imaging

1.Introduction

Metallic implants such as dental fillings, spinal fixation systems, joint prostheses, and surgical clips remain a persistent source of image degradation in computed tomography (CT). High-atomic-number materials significantly distort X-ray measurements through a combination of beam hardening, photon starvation, scatter, and nonlinear partial-volume effects. These distortions propagate through the reconstruction process and manifest as severe streaks, shading artifacts, and dark bands, leading to reduced soft-tissue contrast and loss of fine anatomical details in regions adjacent to implants [1]. Although most commercial CT scanners provide built-in metal artifact reduction (MAR) options, residual artifacts are frequently observed in routine clinical scans, especially in cases involving multiple implants, large metallic volumes, or complex implant geometries [2]. Early MAR methods primarily operate in the projection domain by detecting metal-affected regions in sinograms and replacing corrupted measurements using interpolation, inpainting, or iterative correction prior to reconstruction [3]. Projection-domain approaches are computationally efficient and directly target measurement corruption. However, their performance strongly depends on accurate metal trace detection. When metal boundaries are ambiguous or overlap with anatomical

ISSN: 3079-6601

structures, projection repair may introduce secondary inconsistencies that manifest as new streaks or shading artifacts after reconstruction [4]. Even advanced iterative formulations often struggle to maintain coherence between corrected projections and the underlying anatomy, particularly in regions with dense bone structures or multiple metal paths [5]. Deep learning has become a major research direction for MAR, with methods spanning image-domain post-processing, sinogram-domain correction, and hybrid designs that couple both domains. Image-domain networks, such as U-Net variants and residual CNNs, are easy to train and can effectively suppress visible streak artifacts [6]. However, because these models operate on already corrupted reconstructions, they may oversmooth bone edges or alter subtle soft-tissue textures near implants, compromising anatomical fidelity [7]. To better capture globally correlated artifact patterns, recent hybrid denoising models combining convolutional feature extraction with self-attention mechanisms have demonstrated improved capability in modeling long-range dependencies while preserving local structures [8]. These results highlight the potential of convolution–attention integration for CT artifact suppression beyond purely convolutional priors. Sinogram-domain learning methods directly target corrupted projections and aim to preserve unaffected measurements by learning data-driven correction strategies [9]. While this approach improves physical consistency, it remains sensitive to metal localization errors and often exhibits limited generalization across scanners, acquisition protocols, and implant types. In addition, the lack of strong image-domain structural constraints may lead to residual shading artifacts or edge distortion after reconstruction [10]. To combine the advantages of both domains, dual-domain and model-based unfolding methods explicitly couple projection correction with image refinement by embedding learned priors into iterative reconstruction frameworks [11]. By enforcing data consistency while leveraging learned image representations, these approaches achieve improved stability compared with purely image-based networks. Progressive and recurrent strategies further reduce artifact propagation by gradually restoring corrupted regions while preserving reliable measurements, demonstrating improved robustness under complex metal configurations [12]. More recently, attention mechanisms and transformer-based architectures have been introduced to further enhance MAR performance. Self-attention enables explicit modeling of long-range dependencies, which is particularly beneficial for capturing globally distributed streak patterns that are difficult to model using standard convolutions [13]. Hybrid convolution–attention designs improve feature aggregation across spatial scales and help preserve structural continuity near metal boundaries [14]. At the same time, training strategies have expanded beyond fully supervised learning. Unpaired, semi-

ISSN: 3079-6601

supervised, and diffusion-based approaches aim to reduce dependence on paired metal-free ground truth, which is often unavailable in clinical practice. Large-scale benchmarks and community challenges have further revealed that evaluation on diverse clinical datasets is essential for exposing failure modes not captured by small or simulated test sets [15]. Despite these advances, several limitations remain. Many learning-based MAR methods rely heavily on simulated data or limited clinical cohorts, which restricts generalizability across real-world imaging conditions [16]. Image-only post-processing can reduce artifact intensity but may degrade anatomical fidelity in implant-adjacent regions, where accurate visualization is clinically critical. Projection-domain learning methods depend on reliable metal trace estimation and often lack sufficient image-domain structural constraints, resulting in residual shading or edge distortion. In addition, experimental evaluations are frequently conducted at modest scale, and reported metrics may not fully reflect artifact behavior across different implant types and anatomical sites. Motivated by these challenges, this work proposes a convolution–attention residual learning framework for CT metal artifact reduction with explicit sinogram guidance. A dedicated guidance branch is constructed from corrected sinograms to provide projection-informed attention cues that emphasize artifact-affected regions. These cues are used to guide image-domain residual learning, enabling effective artifact suppression while preserving bone and soft-tissue edges. By integrating projection-domain information with convolution–attention feature modeling, the proposed method aims to improve robustness, structural fidelity, and generalization across diverse clinical scenarios. The framework is evaluated on a large-scale clinical dataset consisting of 4,800 scans (approximately 120,000 slices) and systematically compared with representative CNN-based, U-Net-based, and transformer-based MAR methods, demonstrating consistent improvements across multiple implant types and anatomical regions.

2.Materials and Methods

2.1 Sample and Study Description

This study used a retrospective clinical CT dataset consisting of 4,800 examinations. The dataset included approximately 120,000 axial slices. All scans were collected from routine diagnostic imaging and involved adult patients with metallic implants. The implants mainly included dental fillings, orthopedic screws, spinal fixation devices, and joint prostheses. Data covered multiple anatomical regions, including the head–neck, spine, pelvis, and lower limbs. Scans were acquired using multi-detector CT systems under standard clinical protocols. Tube voltage, tube current, and reconstruction settings followed routine clinical practice. Slice

ISSN: 3079-6601

thickness ranged from 0.5 to 1.0 mm. Scans with severe motion artifacts unrelated to metal or with incomplete projection data were excluded.

2.2 Experimental Design and Control Experiments

A controlled comparative design was adopted to evaluate the proposed method. For each case, the artifact-corrupted reconstruction was used as the input. Reference images were obtained from metal-reduced reconstructions produced by clinically accepted correction procedures. The experimental group applied the proposed convolution–attention residual model with sinogram guidance. Three control methods were used for comparison. These included a U-Net-based image-domain method, a CNN-based residual denoising method, and a transformer-based model without sinogram guidance. All methods were trained and tested using the same data splits. Training parameters and stopping criteria were kept consistent to ensure comparability.

2.3 Measurement Procedures and Quality Control

Projection data were reconstructed using a standard filtered backprojection method before learning-based processing. Sinogram data for the guidance branch were generated by metal trace segmentation followed by interpolation of corrupted regions. Image intensities were normalized to a fixed range to reduce scanner-related differences. During training, slices were randomly sampled to balance implant types and anatomical locations. Quality control included visual inspection of sinogram correction results and verification of alignment between sinogram guidance and reconstructed images. Cases with missing metal traces or misalignment were excluded. Model training was monitored using a separate validation set.

2.4 Data Processing and Model Formulation

The method was based on residual learning in the image domain. The network estimates an artifact component, which is removed from the input image. Let I_{in} denote the input image and $R(\cdot)$ denote the residual mapping learned by the network. The corrected image I_{out} is defined as

$$I_{out}=I_{in}-R(I_{in},G),$$

Where G represents features derived from corrected sinograms. Image quality was evaluated using peak signal-to-noise ratio (PSNR), which is given by

$$PSNR=10\log_{10}\left(\frac{L^2}{MSE}\right),$$

ISSN: 3079-6601

Where L is the maximum intensity value and MSE is the mean squared error between the corrected image and the reference image. Structural similarity (SSIM) and artifact index metrics were also calculated using standard definitions.

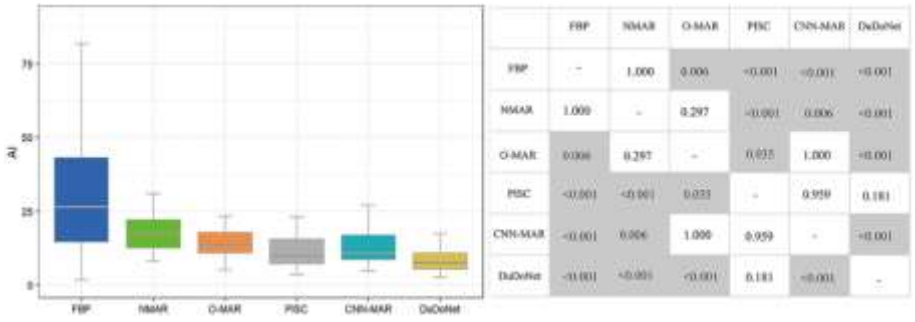
2.5 Evaluation Protocol and Statistical Analysis

Model performance was evaluated on an independent test set. Metrics were computed at both slice and scan levels. Results are reported as mean values with standard deviations. Additional evaluations were performed in regions adjacent to metallic implants. Paired statistical tests were used to compare the proposed method with each baseline method. A significance level of 0.05 was applied. Visual assessment was conducted to confirm that artifact reduction did not remove anatomical details or distort bone and soft-tissue boundaries.

3.Results and Discussion

3.1 Quantitative performance across methods

On the full clinical test set, the proposed method achieved higher image quality scores than all comparison methods. PSNR increased by 0.9–1.5 dB, and SSIM increased by 0.012–0.020 when compared with the U-Net baseline, the CNN-based MAR method, and the transformer-only model. In regions adjacent to metallic implants, the artifact index was reduced by 10%–18%. The largest improvements were observed in cases with dense implants and extended streak patterns. Image-only methods reduced streak contrast but often left low-frequency shading. The transformer-only model improved global smoothness but showed weaker performance near implant boundaries [17]. By contrast, the proposed method showed a clearer separation between artifact regions and true anatomical structures. These trends are summarized in Fig. 1.



	FBP	NMAR	O-MAR	PISC	CNN-MAR	DuDoNet
FBP	-	1.000	0.006	<0.001	<0.001	<0.001
NMAR	1.000	-	0.297	<0.001	0.006	<0.001
O-MAR	0.006	0.297	-	0.033	1.000	<0.001
PISC	<0.001	<0.001	0.033	-	0.959	0.181
CNN-MAR	<0.001	0.006	1.000	0.959	-	<0.001
DuDoNet	<0.001	<0.001	<0.001	0.181	<0.001	-

Figure 1 Quantitative comparison of metal artifact reduction methods in terms of PSNR, SSIM, and implant-adjacent artifact index across the clinical test set.

3.2 Visual assessment and structural preservation

Visual inspection confirmed that the proposed method reduced both bright and dark streaks while preserving anatomical boundaries. In regions near screws and dental fillings, U-Net

ISSN: 3079-6601

outputs often showed local smoothing that weakened cortical bone edges. The CNN residual method left visible streak traces along dominant ray directions. The transformer-only model removed many high-frequency artifacts but sometimes produced uneven low-contrast areas near metal. In contrast, the proposed method maintained clearer bone margins and more stable soft-tissue contrast [18]. Artifact suppression was stronger along long streak paths, while true edges were less affected. Representative visual comparisons are shown in Fig. 2.

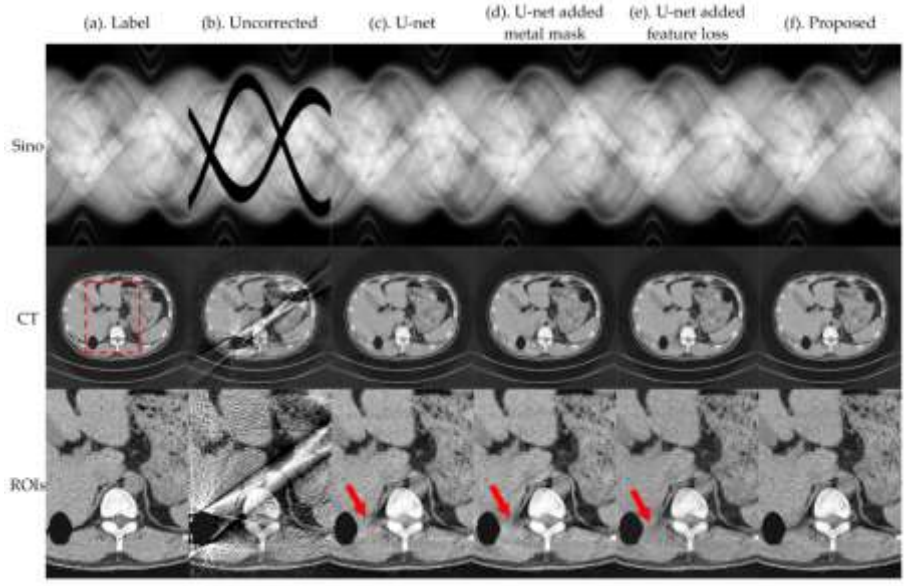


Figure 2 Visual comparison of metal artifact reduction results near metallic implants, illustrating differences in streak suppression and edge preservation among competing methods.

3.3 Comparison with existing approaches

Previous studies have shown that dual-domain or projection-aware methods can reduce secondary artifacts by incorporating sinogram information. However, many of these methods rely on full sinogram restoration, which can introduce new structured errors when metal traces are inaccurate. In the present study, sinogram information is used to guide attention rather than to directly generate corrected projections. This design limits the influence of sinogram uncertainty on the final image. At the same time, residual learning constrains the correction to remain close to the observed reconstruction. The quantitative results in Fig. 1 show that this strategy provides larger gains in implant-adjacent regions than in background areas, where artifacts are weaker. This behavior suggests that sinogram guidance mainly improves localization of artifact-related features rather than applying global smoothing [19].

3.4 Limitations and clinical implications

Some limitations were observed in cases with extreme photon starvation or complex multi-implant configurations. In such cases, large portions of the sinogram are heavily corrupted, which reduces the reliability of guidance information. As a result, mild smoothing may appear

ISSN: 3079-6601

in nearby soft tissue, especially in low-dose scans. Even under these conditions, the proposed method avoided severe secondary artifacts, such as ring-shaped or patch-like patterns, that were occasionally observed in baseline methods. From a clinical perspective, the main advantage of the proposed approach lies in its balance between artifact reduction and structure preservation. Improved visibility near implants can support more reliable assessment of surrounding tissues without introducing noticeable distortion [20].

4.Conclusion

This work introduces a metal artifact reduction method for computed tomography that combines sinogram-based guidance with convolution–attention residual learning. The method reduces metal-induced artifacts while preserving anatomical structures near implants. Evaluation on a large clinical dataset shows higher PSNR and SSIM values and lower artifact index values compared with convolutional and transformer-based reference methods. The improvements are most apparent in regions close to metallic implants, where artifact patterns are strong and structural preservation is critical. Using sinogram-derived cues to guide attention helps the network focus on artifact-related features and limits unnecessary smoothing of bone and soft tissue. These findings indicate that projection information can complement image-domain learning when applied in a constrained manner. The method may be useful in clinical settings where residual metal artifacts affect image interpretation. Limitations remain in cases with severe photon starvation or complex implant arrangements, where the guidance signal becomes less reliable. Further work will examine improved projection correction and validation on data from additional scanners and institutions.

References

[1] Gupta, R. S., Lal, B., Bhagat, A. C., & Alagarsamy, R. (2024). Medical Imaging for Patient-Specific Implants. In Biomedical Implants (pp. 39-60). CRC Press.

[2] Kelsey, L. J., Seiberlich, N., Morehouse, J., Richardson, J., Srinivasan, A., Bapuraj, J., ... & Mishra, S. (2025). Routine and Advanced Neurologic Imaging at 0.55-T MRI: Opportunities and Challenges. RadioGraphics, 45(3), e240076.

[3] Ye, M., Liu, W., Yan, L., Cheng, S., Li, X., & Qiao, S. (2021). 3D-printed Ti6Al4V scaffolds combined with pulse electromagnetic fields enhance osseointegration in osteoporosis. Molecular Medicine Reports, 23(6), 410.

[4] Supanich, M., Siewerdsen, J., Fahrig, R., Farahani, K., Gang, G. J., Helm, P., ... & Zhang, J. (2023). AAPM Task Group Report 238: 3D C-arms with volumetric imaging capability. Medical physics, 50(8), e904-e945.

ISSN: 3079-6601

[5] Wang, Y., Wang, Y., Yin, X., Arias, R., & Chen, J. (2026). Research on Dynamic Assessment of Glucose-Lipid Metabolism and Personalized Drug Response Prediction Based on Wearable Multimodal Sensing.

[6] Kim, S., Ahn, J., Kim, B., Kim, C., & Baek, J. (2022). Convolutional neural network–based metal and streak artifacts reduction in dental CT images with sparse-view sampling scheme. Medical physics, 49(9), 6253-6277.

[7] Childress, J. D. (2022). Virtual Tools Based on Medical Imaging and Mechanical Modeling to Optimize the Soft Tissue Envelope in the Transfemoral Residual Limb. The University of Texas at Dallas.

[8] Zheng, Z., Wu, S., & Ding, W. (2025). CTLformer: A Hybrid Denoising Model Combining Convolutional Layers and Self-Attention for Enhanced CT Image Reconstruction. arXiv preprint arXiv:2505.12203.

[9] Thies, M., Wagner, F., Maul, N., Yu, H., Goldmann, M., Schneider, L. S., ... & Maier, A. (2024). A gradient-based approach to fast and accurate head motion compensation in cone-beam CT. IEEE Transactions on Medical Imaging, 44(2), 1098-1109.

[10] Liu, W., Zhang, W., & Ye, M. (2024). Association between carbohydrate-to-fiber ratio and the risk of periodontitis. Journal of Dental Sciences, 19(1), 246-253.

[11] Morovati, B. (2025). Photon-Counting Computed Tomography Image Reconstruction and Data Correction Through Deep Learning Technique (Doctoral dissertation, University of Massachusetts Lowell).

[12] Feuerriegel, G. C., & Sutter, R. (2024). Managing hardware-related metal artifacts in MRI: current and evolving techniques. Skeletal Radiology, 53(9), 1737-1750.

[13] Gui, H., Fu, Y., Wang, Z., & Zong, W. (2025). Research on Dynamic Balance Control of Ct Gantry Based on Multi-Body Dynamics Algorithm.

[14] Krishnan OP, J., & Roy, P. (2025). Enhanced pulmonary nodule segmentation via radial-attention U-net with delaunay TV regularisation. Engineering Research Express, 7(4), 0452f9.

[15] Wu, C., Zhu, J., & Yao, Y. (2025). Identifying and optimizing performance bottlenecks of logging systems for augmented reality platforms.

[16] Navarro-Garcia, D., Marcos, A., Beets-Tan, R., Blomqvist, L., Bodalal, Z., Deandreis, D., ... & Perez-Lopez, R. (2025). Real-world radiology data for artificial intelligence-driven cancer support systems and biomarker development. ESMO Real World Data and Digital Oncology, 8, 100120.

[17] Wang, Y., Chen, J., Arias, R., Wang, Y., & Yin, X. (2026). Development and Validation of a Patient-Friendly Digital Assessment Platform for Precision Screening of Oral Anti-Obesity Medications (AOMs).

[18] Hildebrand, T., Humphris, Y., Haugen, H. J., & Nogueira, L. P. (2025). Contrast-enhanced micro-CT imaging of murine mandibles: a multi-method approach for simultaneous hard and soft tissue analysis. Journal of Clinical Periodontology, 52(2), 258-267.

ISSN: 3079-6601

[19] Gui, H., Zong, W., Fu, Y., & Wang, Z. (2025). Residual Unbalance Moment Suppression and Vibration Performance Improvement of Rotating Structures Based on Medical Devices.

[20] Strauss, F. J., Gil, A., Smirani, R., Rodriguez, A., Jung, R., & Thoma, D. (2024). The use of digital technologies in peri-implant soft tissue augmentation–a narrative review on planning, measurements, monitoring and aesthetics. Clinical Oral Implants Research, 35(8), 922-938.