ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA240511

Sinogram-Image Dual-Domain Network for Robust Metal Artifact Reduction in CT Image

Chong Liu[a], Yuhan Huang[a], Bo Li[a] and Hui Ding[a,*]

[a]College of Information Engineering, Capital Normal University, Beijing, China
ORCID (Chong Liu): https://orcid.org/0009-0005-2991-8960, ORCID (Yuhan Huang): https://orcid.org/0009-0001-7671-7687, ORCID (Bo Li): https://orcid.org/0009-0007-5671-1169, ORCID (Hui Ding): https://orcid.org/0000-0002-1920-7613

Abstract. Computed tomography (CT) utilizes X-ray technology for internal body imaging. However, the presence of metal objects often results in artifacts due to their significant absorption and scattering of X-rays, thus obstructing lesion diagnosis, especially in the presence of multiple metals. Existing artifact reduction methods often suffer from deficiencies in completeness and preservation of fine detail. To address this limitation, we propose a novel sinogram and image dual-domain network. Specifically, in the sinogram domain, two enhancement modules are designed: one for extracting information from regions affected by metal traces, and the other for learning to restore the sinogram corresponding to these metal traces. Subsequently, utilizing filtered back projection (FBP), artifact removal images are reconstructed in the image domain. Quantitative and qualitative analyses of synthetic images show our framework's superiority over conventional Metal Artifact Reduction (MAR) methods in both synthetic and clinical settings.

1 Introduction

Metal artifacts present a unique challenge in image restoration compared to tasks like super-resolution [10, 34, 29], compression artifact removal [32, 6], and denoising [3, 19, 11]. Unlike these tasks, metal artifacts often involve structured and non-local distortions, manifesting as solid distortions within the metal region and severely compromising image quality and usability. This leads to the loss or distortion of crucial diagnostic information, hampering physicians' ability to diagnose patients accurately.

The traditional model-based approach uses linear interpolation (LI) [8] and normalized MAR (NMAR) [20] to reconstruct CT images by filling metal-affected regions in the sinogram with different estimation strategies. The rapid advancements in deep learning have significantly enhanced metal artifact reduction (MAR) tasks. Early methods used Convolutional Neural Networks (CNNs) to reduce artifacts and preserve anatomical structures, relying on pre-trained models like CNNMAR [33], which introduced dependencies on model accuracy and stability. To address these issues, novel approaches such as the unsupervised artifact disentanglement network (ADN) [13], deep residual learning for cervical CT images [7], and GANs for direct artifact reduction [5] were developed. Additionally, Wang et al. proposed new optimization algorithms, including DICDNet [24] and the adaptive convolutional dictionary network (ACDNet) [26], to eliminate metal artifacts in CT images.

Although deep learning methods in both sinogram and image domains have shown promising results, they have limitations. To address these, Lin et al. proposed DuDoNet, a dual-domain network that is end-to-end trainable [14]. Wang et al. followed with InDuDoNet [25] and its enhanced version, InDuDoNet+ [27]. Lyn et al. introduced U-DuDoNet [18], a non-paired dual-domain network utilizing metal mask projection encoding. Additionally, Wang et al. proposed DAN-Net [28], a dual-domain adaptive scaling non-local network. Recently, Liu et al. applied the diffusion model in MAR with an approach called Unsupervised CT Metal Artifact Reduction by Plugging Diffusion Priors in Dual Domains [16], demonstrating impressive performance in reducing metal artifacts.

Despite the success of existing methods in mitigating metal artifacts, their effectiveness still needs to be improved. To overcome these challenges, we propose a novel dual-domain network for joint learning in both the image and sinogram domains. Unlike previous approaches, two deep learning networks are trained in the sinogram domain: the first recovers tissue details, and the second treats metal trace regions as missing data. CT images are then reconstructed using the conventional FBP (Filtered Back Projection) algorithm. The images from the sinogram domain are further refined in the image domain through channel concatenation, resulting in the final artifact-reduced image. This framework is trained end-to-end, allowing CT images to mutually benefit from learning in both domains. Evaluation of simulated and real metal artifact data demonstrates our model's superior ability to remove metal artifacts. Our main contributions are summarized as follows:

- We propose a novel end-to-end trainable dual-domain framework that effectively exploits complementary information from the sinogram and image domains, demonstrating superior performance in mitigating metal artifacts in CT imaging.
- In the sinogram domain, two dedicated sinogram-enhancement networks are designed. One prioritizes recovering tissue details, while the other treats metal trace regions as missing data. The synergistic learning in the image domain leverages the strengths of both networks for more comprehensive metal artifact reduction.
- Additionally, a cross-domain collaboration and mutual learning mechanism enhance the efficiency and accuracy of metal artifact removal. The artifact-reduced CT image initially obtained from the sinogram domain is further refined and optimized in the im-

* Corresponding Author. Email: dhui@cnu.edu.cn.

age domain by channel concatenation with conventionally reconstructed FBP images.

2 Backgrounds and Related Works

2.1 Fanbeam CT Geometry

With the advancement of modern medicine, the implantation of prostheses containing metallic materials has become a common treatment method. These prostheses include dentures, pacemakers, joint replacements, and prostheses, among others, and they play an essential role in clinical practice. However, compared to human tissues, these metallic objects have a higher X-ray absorption capacity, leading to complex artifact issues. Metal artifacts are common phenomena in CT images, mainly caused by metallic materials' high X-ray absorption rate. When X-rays pass through metal objects, their energy is absorbed or scattered, leading to abnormally enhanced or weakened signals received by the detector, thus forming artifacts in the image. These artifacts typically appear as streaks or ring-shaped structures, significantly affecting the quality and diagnostic accuracy of the image.

Although some methods can be attempted in the image domain to address metal artifacts, such as filtering or image reconstruction algorithms to suppress or repair artifacts, image domain processing methods have some limitations, such as potentially further compromising image quality and the difficulty of accurately identifying and eliminating the structural distortion caused by artifacts. Therefore, much research has turned to the sinogram domain processing to address metal artifacts more effectively. In the sinogram domain, metal artifacts typically manifest as abnormal signals in projection data, enabling more accurate detection and localization of artifacts. In the projection process of CT imaging, Figure 1(a) shows the general geometric shape of fan-beam CT, with the source and detector rotating relative to the origin at different projection angles. The distance from the source to the origin is D, and β is the projection angle, with the rays emitted by the source being received by the detector. In fan beam CT, the source rotates around the center of rotation, and each projection angle generates a set of projection lines on the detector. Combining the projection lines from all angles results in a sinogram, as shown in Figure 1(b), which contains the projection information of the object at different angles.

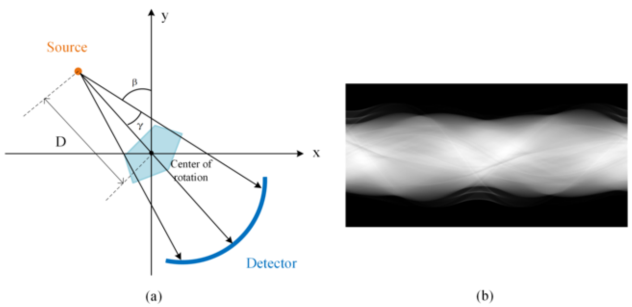


Figure 1. Process of fan beam projection: (a) Fanbeam CT geometry, (b) The combination of projection lines at different angles forms the sinograms.

2.2 The generation process of metal artifacts

FP (Forward Projection): FP refers to the process of generating a transverse view using projection data. In computed tomography (CT) imaging, forward projection is projecting density information inside the object onto the detector. The FP process simulates the interaction of X-rays passing through the object and the detector, generating projection data at different projection angles. These projection data are organized into a two-dimensional dataset called a sinogram, where each row represents one-dimensional projection data at a projection angle. The sinogram reflects the projection information of the object at different angles and serves as the basis for subsequent image reconstruction.

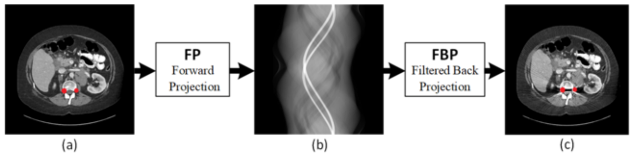


Figure 2. Progress of metal artifact generation: (a) Presence of metal implants (red patches), (b) Effects on the sinogram, and (c) Artifacts in the reconstructed image.

FBP (Filtered Back Projection): FBP is one of the most commonly used methods for CT image reconstruction. It reconstructs the image based on sinogram data through a series of steps.

When a patient undergoes a medical procedure involving a metal implant in the body (as illustrated in Figure 2(a)), the resulting sinogram is inevitably altered following the FP forward projection process (as depicted in Figure 2(b)). Without suitable mitigation strategies, these alterations manifest in the reconstructed image post-FBP, disrupting the natural trajectory of rays and leading to artifacts surrounding the metallic object (as exemplified in Figure 2(c)). This delineates the comprehensive process underlying the generation of metal artifacts.

3 Methodology

3.1 Overview of the Proposed DSI-Net

Our model employs a dual-domain joint learning strategy called DSI-Net to take advantage of the sinogram and image domain information benefits. Figure 3(a) illustrates the overall architecture of DSI-Net, which consists primarily of two components: the sinogram domain and the image domain. In the image domain, we have designed a network, I-Net, outlined in the yellow box in Figure 3(b). In the sinogram domain, we have developed two sinogram enhancement-network modules, the RS-Net (residual-based sinogram domain network) module outlined in the green box in Figure 3(c) and LIS-Net (linear interpolation-based sinogram network) module outlined in the orange box in Figure 3(d). The output image obtained after processing through the two sinogram enhancement networks in the sinogram domain is then further refined and optimized in the image domain using the I-Net module. We will provide more detailed information in the following sections.

3.2 Sinogram Domain Network

Two types of dual-domain networks are used to extract useful information from metal trace regions. The first type uses the difference between original projection data and linear interpolated projection data, along with metal masks, as input to recover more tissue details. Examples of this type include DAN-Net [28] and DSCIP [31]. The second type treats the projection data in metal trace lines as missing data, exemplified by DuDoNet [14], which results in the loss of details near the metal regions in the corrected CT images. We aim

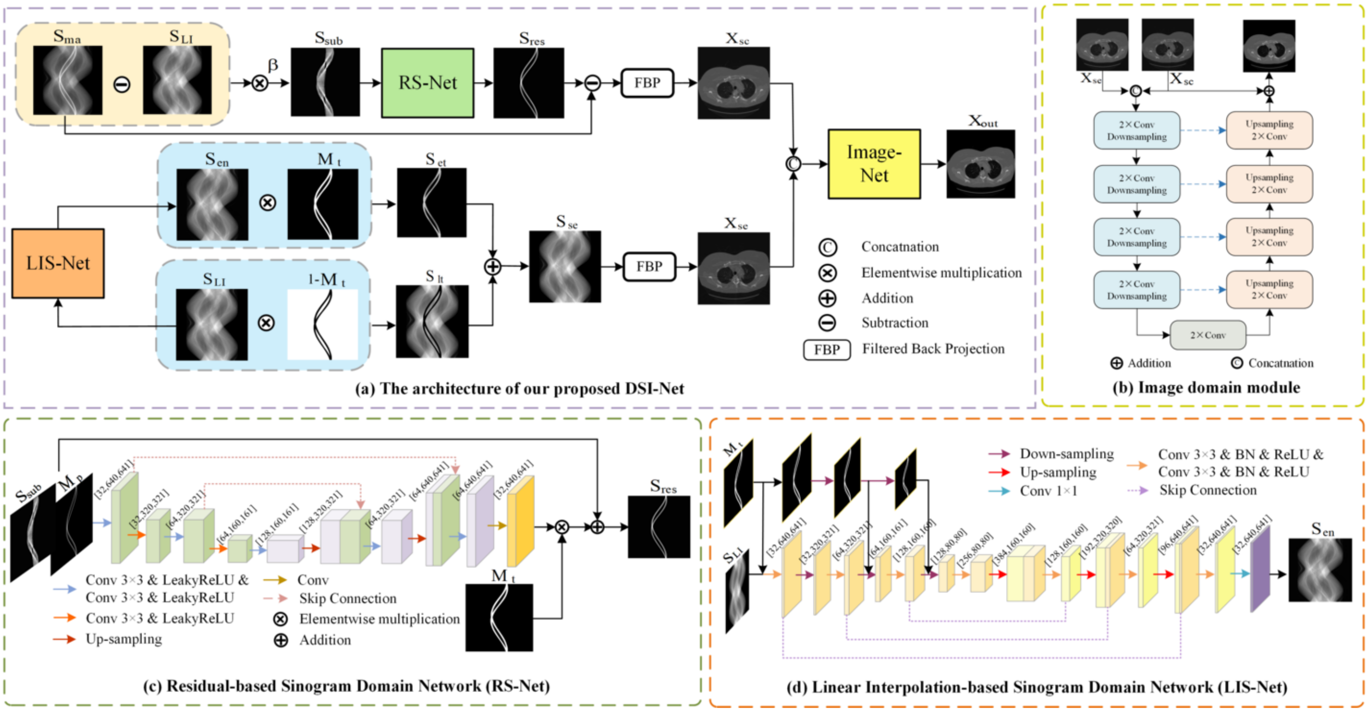


Figure 3. The architecture of DSI-Net and detailed modules.

to combine the strengths of both types by proposing two sinogram enhancement networks in the sinogram domain.

3.2.1 RS-Net Module

We have proposed a sinogram learning strategy for RS-Net that can effectively restore tissue details while removing metal artifacts. We draw inspiration from prior work [17], which suggests that the residual, denoted as S_{sub}, between the original projection data, denoted as S_{ma}, and the linear interpolation projection data, denoted as S_{LI}, retains valuable information. To restore the data in the metal trace, [2] adopted a linear attenuation operation, selecting a parameter β, the value of β is set to between 0.3 and 0.5, to control the trade-off between reducing metal artifacts and preserving details of the surrounding implanted tissues. S_{sub} can be described as:

$$S_{sub} = (S_{ma} - S_{LI}) * \beta. \tag{1}$$

Meanwhile, we propose an architecture RS-Net (as shown in Figure 3(c)), comprising multiple layers of convolution and upsampling operations, learning intricate representations of metal artifacts. This network architecture includes convolution layers with Rectified Linear Unit (ReLU) activation, downsampling layers, and upsampling layers. However, due to the current network containing downsampling operations, metal trace information will be lost. Therefore, we need to use the knowledge of metal mask M_p, denoted as:

$$M_p = FP(X_{metal}). \tag{2}$$

where X_{metal} represents the metal mask in the image domain, and FP is forward projection. Then, We concatenate S_{sub} and M_p to form the input of RS-Net. Through RS-Net, we derive the image S_{res}, denoted as:

$$S_{res} = RS\text{-}Net([S_{sub}, M_p]). \tag{3}$$

where $[S_{sub}, M_p]$ represents the concatenation operation of image S_{sub} and M_p. Using S_{sub} as input can enhance the smoothness of preprocessed projection data, obtain useful information from the metal trace area, and improve the continuity of the metal trace boundary. The residual is obtained by subtracting S_{res} from the original metal projection image S_{ma}, denoted as:

$$S_{sc} = S_{ma} - S_{res}. \tag{4}$$

Subsequently, a filtered back projection (FBP) operation is performed on S_{sc} to obtain the image X_{sc}, denoted as:

$$X_{sc} = FBP(S_{sc}). \tag{5}$$

where FBP is filtered back projection. We use an L1 loss to train RS-Net, denoted as:

$$L_{RS} = \|S_{sc} - S_{gt}\|_1 . \tag{6}$$

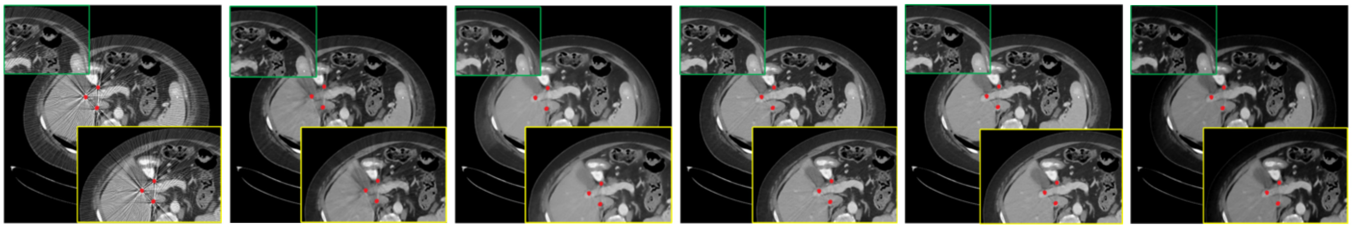
where S_{gt} is the sinogram of the clean image.

3.2.2 LIS-Net Module

Enhancing sinograms holds promise in mitigating artifacts induced by metal objects. Metal objects present in the subject during CT imaging disrupt X-ray absorption, leading to artifacts that degrade image quality and diagnostic precision. To address this, we introduce a sinogram learning strategy for LIS-Net, illustrated in Figure 3(d), which is a convolutional neural network (CNN)-based architecture. LIS-Net is designed to learn the recovery of the S_{LI} region of M_t=1. Due to the down-sampling operations in the network, this will result in the loss of metal trace information [21, 4]. To retain sufficient information on metal traces, a mask pyramid network (MPN) [12] is introduced to explicitly feed the mask information into each layer [31], utilizing the linear interpolated projection data S_{LI} and metal

Table 1. Average PSNR (dB)/SSIM of different ablation methods on the synthesized DeepLesion data.

Methods	Large Metal		→ Small Metal		Average
Input	25.90/0.6159	29.53/0.6109	30.56/0.6878	32.28/0.7184	29.57/0.6582
RS-Net	37.14/0.9598	38.91/0.9590	40.12/0.9849	41.85/0.9758	39.51/0.9699
LIS-Net	39.60/0.9740	40.07/0.9759	41.07/0.9665	42.04/0.9818	40.70/0.9746
RS-Net + I-Net	40.35/0.9854	41.58/0.9873	43.69/0.9888	46.16/0.9903	42.94/0.9880
LIS-Net + I-Net	40.46/0.9787	41.56/0.9781	44.45/0.9845	**47.87**/0.9901	43.59/0.9829
RS-Net + LIS-Net + I-Net (Ours)	**40.70/0.9870**	**41.98/0.9854**	**45.28/0.9913**	47.44/**0.9912**	**43.85/0.9889**



(a) Input (a) RS-Net (b) LIS-Net (c) RS-Net+I-Net (d) LIS-Net+I-Net (e) Ours

Figure 4. Visual comparisons of ablation variants of DSI-Net. The red patches indicate metallic implants. The display window is [-175, 275] HU.

trace M_t as joint inputs for feature extraction and down-sampling operations to extract higher-level feature representations. After being processed by LIS-Net, the image is element-wise multiplied with the metal trace M_t to remove partial metal artifacts effectively. To further enhance the entire image, a reverse mask is obtained using 1-M_t, which is then element-wise multiplied with S_{LI}. Linearinterpolated projection data is used in non-metallic areas to fill or repair possible artifacts or missing information. They were finally, adding these two parts results in the optimized image S_{se}. Subsequently, a filtered back projection (FBP) operation is performed on S_{se} to obtain the image X_{se}. The above description can be obtained:

$$S_{se} = M_t \odot LIS\text{-}Net([S_{LI}, M_t]) + (1 - M_t) \odot S_{LI}. \quad (7)$$

$$X_{se} = FBP(S_{se}). \quad (8)$$

where $\odot$ denotes element-wise multiplication, [S_{LI}, M_t] represents the concatenation operation of image S_{LI} and M_t, and FBP is filtered back projection. We use an L1 loss to train LIS-Net, denoted as:

$$L_{LIS} = \|S_{se} - S_{gt}\|_1 . \quad (9)$$

where S_{gt} is the sinogram of the clean image.

3.3 Image Domain Network

After filtering back projections, secondary projections will occur. An image enhancement network, termed I-Net, is incorporated into our framework to address the remaining metal artifacts and further refine the image. The I-Net module, depicted in Figure 3(b), employs a U-Net architecture [23] with a depth of 4 for residual learning. It takes as input the images obtained from the previous sinogram domain processing and generates an enhanced output image X_{out} through the following residual learning formulation:

$$X_{out} = X_{sc} + I\text{-}Net([X_{sc}, X_{se}]). \quad (10)$$

where X_{sc} is the image obtained from RS-Net, X_{se} is the image obtained from LIS-Net, and [X_{sc}, X_{se}] represents the concatenation operation of image X_{sc} and X_{se}. We use an L1 loss to train I-Net, denoted as:

$$L_I = \|(X_{out} - X_{gt}) \odot (1 - Mask)\|_1 . \quad (11)$$

where $\odot$ denotes element-wise multiplication, and Mask denotes metal mask.

3.4 Overall Loss Function

The L1 above loss employed for optimization in the RS-Net and LIS-Net modules penalizes inconsistencies in individual projection values within the sinogram domain. However, it does not account for geometric consistency. To mitigate potential new artifacts in the reconstructed CT images arising from such inconsistencies, we further incorporate a filtered back-projection (FBP) loss term, formulated as:

$$L_{FBP\_RS} = \|(X_{sc} - X_{gt}) \odot (1 - Mask)\|_1 . \quad (12)$$

$$L_{FBP\_LIS} = \|(X_{se} - X_{gt}) \odot (1 - Mask)\|_1 . \quad (13)$$

$$L_{FBP} = L_{FBP\_RS} + L_{FBP\_LIS}. \quad (14)$$

where Mask is the metal mask and $\odot$ denotes element-wise multiplication. Our total model loss L includes the sinogram enhancement loss L_{RS} and L_{LIS}, image enhancement loss L_I, and filtered back projection loss L_{FBP}, denoted as:

$$L = L_{RS} + L_{LIS} + L_{FBP} + L_I. \quad (15)$$

4 Experiments

4.1 Datasets & Experimental Setting

Synthesized Dataset: Following the guidelines outlined in reference [31], we selected 1200 CT images randomly from the publicly available DeepLesion dataset [30] and paired them with 100 metal masks from reference [33]. Our training set consisted of 1000 clean CT images and 90 metal images, resulting in 90,000 combinations. The remaining 200 clean CT images and 10 metal images were utilized for testing, creating 2000 combinations for network evaluation. Our metal synthesis process took into account various factors such as beam hardening, Poisson noise, and multi-colored X-rays. We utilized 640 evenly spaced projection images between 0-360 degrees, resulting in a synthesized CT image size of 416x416 and a sinogram size of 641x640.

Clinical Dataset: We evaluated our method using the CTPelvic1K [15] dataset, a comprehensive CT dataset tailored for pelvic segmentation tasks. This dataset encompasses 1184 CT scans sourced from 7 distinct origins, among which 75 scans exhibit metal artifacts. Five sources are publicly accessible datasets, while the remaining two were recently compiled. The dataset is meticulously annotated to segment critical anatomical structures, including the lumbar spine, sacrum, left hip, and right hip. Additionally, the CLINIC-metal subset within the dataset comprises 14 volumes specifically characterized by the presence of metal artifacts alongside pixel-level annotations. Metal implants are delineated using threshold segmentation at 2500 Hounsfield Units (HU) to facilitate accurate identification within the scans.

Evaluation Metrics: Consistent with other metal artifact removal methods, we have also adopted the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as this study's primary quantitative evaluation metrics. These two indicators are widely used in image processing and can objectively reflect the improvement in image quality. By calculating the PSNR and SSIM values of images before and after processing, we can accurately assess the effectiveness of our proposed method in removing metal artifacts, thus enabling a scientific and fair comparison with other methods.

4.2 Implementation Details

Our model is implemented using the PyTorch framework [22], and our network is trained end-to-end. We applied differential FP and FBP operations as suggested by the ODL library [1] to train our RS-Net and LIS-Net. To train the model, we utilized the Adam [9] optimizer with parameters (β_1, β_2) = (0.5, 0.999), a batch size of 1, and a learning rate starting from 2x10^{-4}, halved every 30 epochs. The model was trained for 300 epochs on an NVIDIA TITAN RTX. Within each training iteration, we randomly selected a synthetic metal artifact CT image from a pool of 90 different metal mask pairs and used different CT images as a small batch input for computing the total objective function.

4.3 Ablation Study

In this section, we evaluate the effectiveness of different components in the proposed method. Experiments were conducted on the following configurations:

1. RS-Net: Using only the RS-Net for the sinogram network
2. LIS-Net: Using only the LIS-Net for the sinogram network
3. RS-Net + I-Net: Employing both RS-Net and I-Net for the dual-domain learning network.
4. LIS-Net + I-Net: Utilizing both LIS-Net and I-Net for the dual-domain learning network.
5. RS-Net + LIS-Net + I-Net: Using all three networks for the dual-domain learning network.

The quantitative results of the ablation study are detailed in Table 1, while the visual results are presented in Figure 4. Compared to LIS-Net, RS-Net shows a slight deficiency in reducing artifacts but still significantly preserves the details of the tissue of the image, maintaining good clarity and integrity even around the artifacts. In contrast, although LIS-Net effectively reduces interference from artifacts, this process sacrifices some clarity in organizational structure, resulting in slightly blurred edges. Further analysis of the data in Table 1 demonstrates that using the sinogram enhancement network alone does not yield the best results, and both PSNR and SSIM fail to reach peak levels. Combining various technologies or network architectures to improve image quality is further confirmed.

RS-Net + I-Net significantly reduced metal artifacts and notably improved PSNR and SSIM metrics. Although some stripe artifacts are still visible, the details of the tissues surrounding the metal are better preserved. According to the data in Table 1, LIS-Net + I-Net outperforms RS-Net + I-Net by increasing the average PSNR by 0.65 dB. Visually, LIS-Net + I-Net demonstrates superior artifact removal, eliminating dark bands in the corrected image. Our model RS-Net + LIS-Net + I-Net integrates the advantages of RS-Net + I-Net and LIS-Net + I-Net in the sinogram domain. Compared to LIS-Net + I-Net, Our model increases the average PSNR by 0.26 dB, albeit with a slight decrease in SSIM by 0.006. However, not only does the data support this finding, but Our model also exhibits the best performance in artifact removal and tissue restoration, both in the regions surrounding the metal and those further away. This discovery provides robust quantitative support for our research and is visually validated through the comparison of images, thus demonstrating the need for three networks.

4.4 Comparision with State-of-the-Art Methods

Quantitative analysis: We have conducted a thorough comparison between our model and several other existing methods, including widely used linear interpolation (LI) [8], deep learning-based methods such as DICDNet [24] and ACDNet [26], an interpretable dual-domain network InDuDoNet+ [27] and end-to-end trainable dual-domain MAR method DuDoNet [14]. To evaluate the performance of each method, we have used publicly available codes and models to run a series of rigorous experiments. In Table 2, we have summarized the comparison results of our method with other metal artifact reduction (MAR) techniques on the DeepLesion dataset, using SSIM and PSNR as evaluation metrics. Our analysis shows that while LI improves SSIM and PSNR over uncorrected CT images, deep learning methods like DICDNet perform better. Dual-domain networks like DuDoNet and InDuDoNet+ further enhance PSNR and SSIM due to integrated sinogram enhancement. ACDNet leverages the prior structure of metal artifacts and demonstrates better MAR performance and generalization capability than DuDoNet. However, our proposed method outperforms ACDNet in terms of SSIM and PSNR. Our proposed method achieves the best performance in terms of PSNR and SSIM, demonstrating its effectiveness in reducing metal artifacts.

Qualitative analysis: Figure 5 visually compares our method with others using three examples from the DeepLesion dataset, featuring different sizes of metal implants. The figure includes the reference non-metal image (Figure 5(A1-A3)), simulated metal artifact image (Figure 5(B1-B3)), and results from various MAR methods. Metal is highlighted in red, and the yellow box magnifies the effect on metal and surrounding tissues. Traditional LI (Figure 5(C1-C3)) leaves many artifacts, while deep learning methods perform better. However, DuDoNet (Figure 5(F1-F3)) overly smooths surrounding tissues, and DICDNet (Figure 5(D1-D3)) shows radial artifacts. In contrast, our method (Figure 5(H1-H3)) effectively suppresses artifacts while preserving tissue details, achieving the best SSIM and

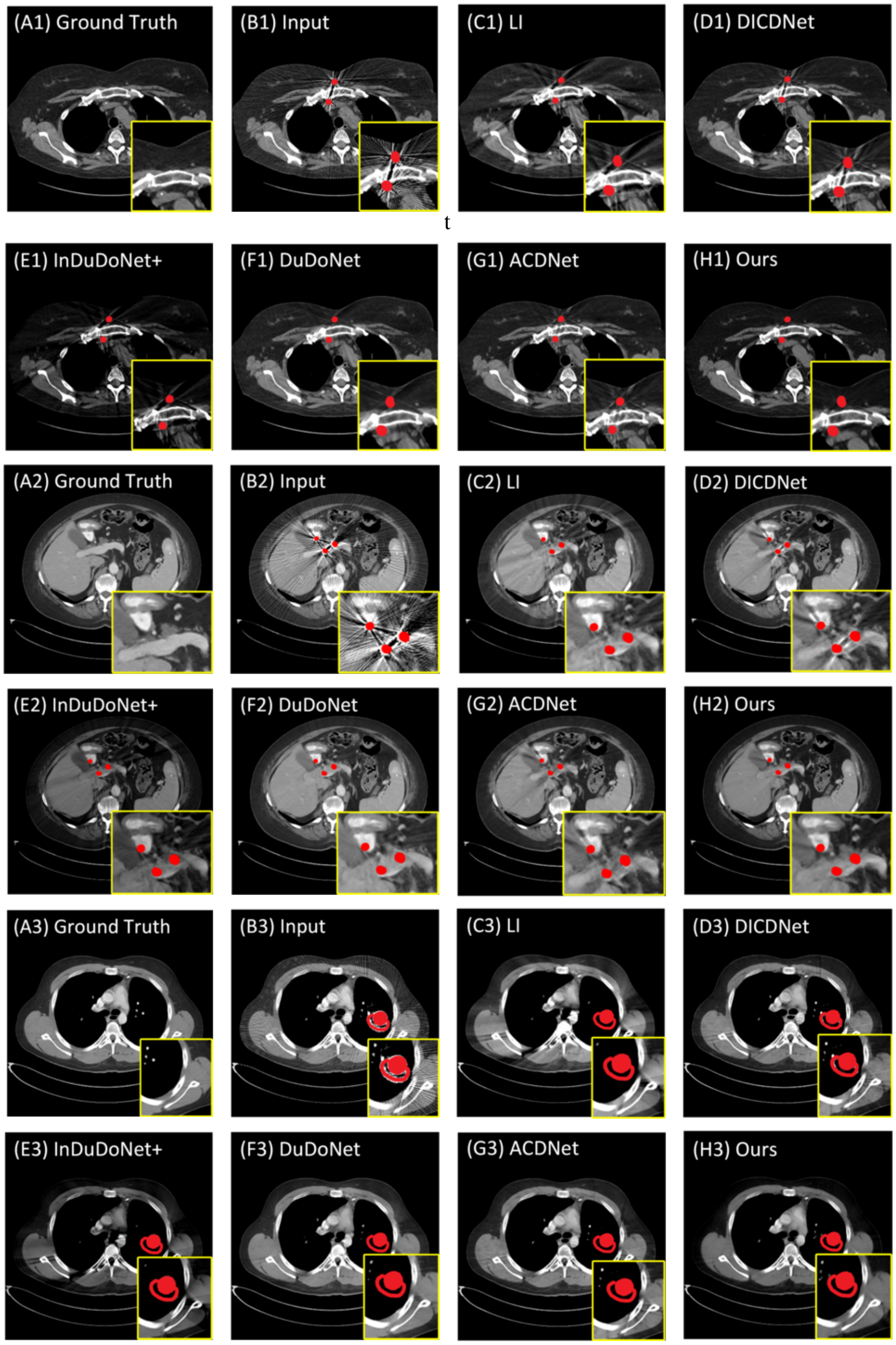


Figure 5. Visual comparison of different MAR methods for different metal sizes using the DeepLesion dataset. The red parts indicate metallic implants. The display window is [-175, 275] HU.

Table 2. Average PSNR (dB)/SSIM of different MAR methods on the synthesized DeepLesion data.

Methods	Large Metal → Small Metal				Average
Input	25.90/0.6159	29.53/0.6109	30.56/0.6878	32.28/0.7184	29.57/0.6582
LI [8]	27.37/0.8148	31.16/0.8593	36.72/0.8770	39.11/0.8916	33.58/0.8607
DICDNet [24]	35.62/0.9598	39.56/0.9810	42.81/0.9813	45.63/0.9870	40.90/0.9772
InDuDoNet+ [27]	36.76/0.9479	39.58/0.9646	43.03/0.9825	45.71/0.9856	41.27/0.9701
DuDoNet [14]	39.95/0.9823	40.44/0.9408	42.05/0.9636	46.28/0.9898	42.13/0.9691
ACDNet [26]	40.28/0.9812	41.70/0.9809	43.69/0.9840	46.52/0.9908	43.05/0.9842
DSI-Net (Ours)	**40.70/0.9870**	**41.98/0.9854**	**45.28/0.9913**	**47.44/0.9912**	**43.85/0.9889**

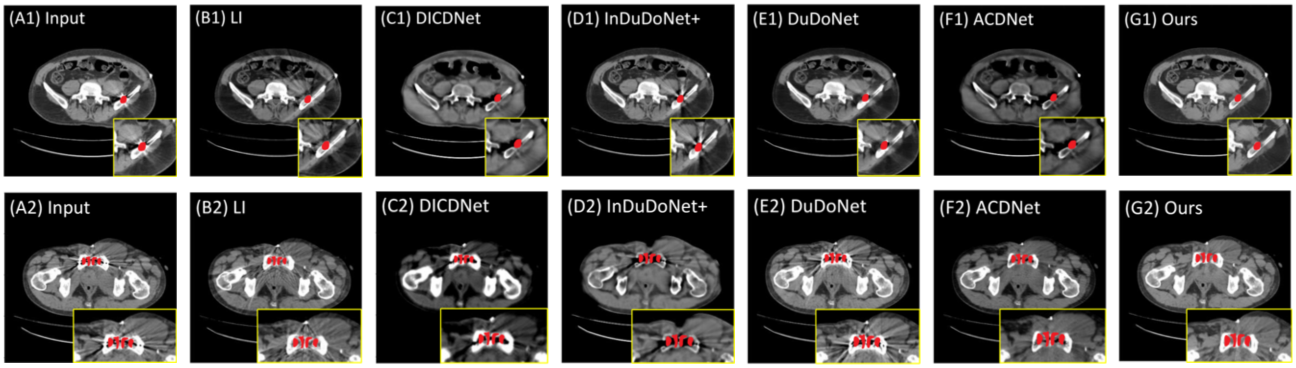


Figure 6. Visual comparison of different MAR methods for different metal sizes using the CTPelvic1K dataset. The red parts indicate metallic implants. The display window is [-175, 275] HU.

PSNR metrics, and demonstrating its effectiveness.

4.5 *Comparison on Clinical Dataset*

To evaluate the efficacy of our proposed method in clinical settings, we conducted experiments using two examples on the CTPelvic1K dataset on CT images with metal artifacts common in clinical practice. Figure 6 shows the MAR results from different methods, with metal regions highlighted in red. We segmented metal artifacts (Figure 6(A1-A2)) and corrected them using LI (Figure 6(B1-B2)), DICDNet (Figure 6(C1-C2)), InDuDoNet+ (Figure 6(D1-D2)), DuDoNet (Figure 6(E1-E2)), ACDNet (Figure 6(F1-F2)) and our proposed approach (Figure 6(G1-G2)). LI only partially mitigates artifacts and may introduce new ones. DICDNet and ACDNet improve MAR but tend to blur surrounding tissue, while InDuDoNet+ avoids blurring but has streak artifacts.

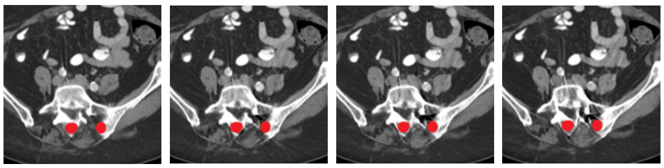
DICDNet and ACDNet perform well on simulated datasets but less on clinical data. DuDoNet effectively removes most metal artifacts in clinical data. However, our proposed method surpasses DuDoNet by better suppressing metal artifacts and restoring tissue structures, providing consistent results across simulated and clinical datasets. These findings highlight the potential of our method for clinical applications.

4.6 *Robustness to Inaccurate Masks*

Severe artifacts around metals complicate the precise acquisition of metal masks on sinograms, leading to increased errors. A straightforward approach is to enlarge the obtained metal mask size. We implemented a series of dilation operators to generate larger metal masks and obtained the corresponding metal mask projections. Using the MaxPool2d function in PyTorch, we realized dilation operators, employing kernel sizes of 3×3, 5×5, and 7×7 for metal mask expansion. We set zero padding to 1, 2, and 3 to maintain the original shape. The expanded masks and their respective metal traces are designated as Mask0 (the precise metal mask), Mask3, Mask5, and Mask7, with corresponding metal traces named Trace0 (the precise metal trace), Trace3, Trace5, and Trace7, as shown in Figure 7. Figure 8 illustrates the visual outcomes of our proposed method, DSI-Net, under imprecise masking conditions. The data presented in Figure 8 indicates that our methodology, DSI-Net, demonstrates remarkable efficacy in mitigating metal artifacts across varying mask sizes, particularly exhibiting favorable performance under the exact mask (Mask 0). However, additional artifact phenomena are observed as the mask size increases (Mask3, Mask5, Mask7). This observation underscores the detrimental impact of inaccurate metal masks on experimental outcomes, indicating that imprecise artificial information can degrade processing performance and potentially introduce new metallic artifact issues.

Trace0 Trace3 Trace5 Trace7

Figure 7. The precise metal trace (Trace 0) and dilated metal traces (Trace3, Trace5, and Trace7).



(a) Mask0 (b) Mask3 (c) Mask5 (d) Mask7

Figure 8. Visualization results using DSI-Net under Mask0, Mask3, Mask5, and Mask7. The red patches indicate metallic implants. The display window is [-175, 275] HU.

5 Conclusion and Future Work

Metal implants in CT imaging often lead to severe artifacts, degrading image quality. Over time, dual-domain metal artifact reduction (MAR) techniques have shown promise. This paper introduces an innovative dual-domain framework that effectively mitigates metal artifacts while preserving crucial tissue details. Our novel method synergistically combines two sinogram enhancement networks - one dedicated to recovering tissue information and the other specialized in artifact removal. Through fostering mutual learning between the sinogram and image domains, our approach has demonstrated impressive performance in significantly reducing metal artifacts while meticulously preserving the tissue structures surrounding the metal regions. Our method outperforms metal artifact reduction, yet enhancements are needed, especially for large or multiple metal objects. To refine and strengthen our approach, we plan to conduct extensive clinical studies and rigorously evaluate its performance on authentic patient data across diverse clinical settings, thereby assisting its future adoption and facilitating improved clinical applications.

References

[1] J. Adler, H. Kohr, and O. Öktem. Operator discretization library (odl). *Zenodo*, 2017.

[2] L. M. Chen, Y. Liang, G. A. Sandison, and J. Rydberg. Novel method for reducing high-attenuation object artifacts in ct reconstructions. In *Medical Imaging 2002: Image Processing*, volume 4684, pages 841–850. SPIE, 2002.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.

[4] M. U. Ghani and W. C. Karl. Fast enhanced ct metal artifact reduction using data domain deep learning. *IEEE Transactions on Computational Imaging*, 6:181–193, 2019.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[6] J. Guo and H. Chao. Building dual-domain representations for compression artifacts reduction. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 628–644. Springer, 2016.

[7] X. Huang, J. Wang, F. Tang, T. Zhong, and Y. Zhang. Metal artifact reduction on cervical ct images by deep residual learning. *Biomedical engineering online*, 17:1–15, 2018.

[8] W. A. Kalender, R. Hebel, and J. Ebersberger. Reduction of ct artifacts caused by metallic implants. *Radiology*, 164(2):576–577, 1987.

[9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[10] 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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[11] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2noise: Learning image restoration without clean data. In *International Conference on Machine Learning*, pages 2965–2974. PMLR, 2018.

[12] H. Liao, W.-A. Lin, Z. Huo, L. Vogelsang, W. J. Sehnert, S. K. Zhou, and J. 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.

[13] H. Liao, W.-A. Lin, S. K. Zhou, and J. Luo. Adn: artifact disentanglement network for unsupervised metal artifact reduction. *IEEE Transactions on Medical Imaging*, 39(3):634–643, 2019.

[14] W.-A. Lin, H. Liao, C. Peng, X. Sun, J. Zhang, J. Luo, R. Chellappa, and S. K. 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.

[15] P. Liu, H. Han, Y. Du, H. Zhu, Y. Li, F. Gu, H. Xiao, J. Li, C. Zhao, L. Xiao, et al. Deep learning to segment pelvic bones: large-scale ct datasets and baseline models. *International Journal of Computer Assisted Radiology and Surgery*, 16:749–756, 2021.

[16] X. Liu, Y. Xie, S. Diao, S. Tan, and X. Liang. Unsupervised ct metal artifact reduction by plugging diffusion priors in dual domains. *IEEE Transactions on Medical Imaging*, 2024.

[17] Y. Lyu, W.-A. Lin, H. Liao, J. Lu, and S. K. Zhou. Encoding metal mask projection for metal artifact reduction in computed tomography. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23*, pages 147–157. Springer, 2020.

[18] Y. Lyu, J. Fu, C. Peng, and S. K. Zhou. U-dudonet: unpaired 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 296–306. Springer, 2021.

[19] M. Makitalo and A. Foi. Optimal inversion of the anscombe transformation in low-count poisson image denoising. *IEEE transactions on Image Processing*, 20(1):99–109, 2010.

[20] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß. Normalized metal artifact reduction (nmar) in computed tomography. *Medical physics*, 37(10):5482–5493, 2010.

[21] H. S. Park, S. M. Lee, H. P. Kim, J. K. Seo, and Y. E. Chung. Ct sinogram-consistency learning for metal-induced beam hardening correction. *Medical physics*, 45(12):5376–5384, 2018.

[22] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. 2017.

[23] 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*, pages 234–241. Springer, 2015.

[24] H. Wang, Y. Li, N. He, K. Ma, D. Meng, and Y. Zheng. Dicdnet: deep interpretable convolutional dictionary network for metal artifact reduction in ct images. *IEEE Transactions on Medical Imaging*, 41(4):869–880, 2021.

[25] H. Wang, Y. Li, H. Zhang, J. Chen, K. Ma, D. Meng, and Y. 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.

[26] H. Wang, Y. Li, D. Meng, and Y. Zheng. Adaptive convolutional dictionary network for ct metal artifact reduction. *IEEE The 31st International Joint Conference on Artificial Intelligence*, 2022.

[27] H. Wang, Y. Li, H. Zhang, D. Meng, and Y. Zheng. Indudonet+: A deep unfolding dual domain network for metal artifact reduction in ct images. *Medical Image Analysis*, 85:102729, 2023.

[28] T. Wang, W. Xia, Y. Huang, H. Sun, Y. Liu, H. Chen, J. Zhou, and Y. Zhang. Dan-net: Dual-domain adaptive-scaling non-local network for ct metal artifact reduction. *Physics in Medicine & Biology*, 66(15): 155009, 2021.

[29] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

[30] K. Yan, X. Wang, L. Lu, L. Zhang, A. P. Harrison, M. Bagheri, and R. M. Summers. Deep lesion graphs in the wild: relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9261–9270, 2018.

[31] L. Yu, Z. Zhang, X. Li, and L. Xing. Deep sinogram completion with image prior for metal artifact reduction in ct images. *IEEE transactions on medical imaging*, 40(1):228–238, 2020.

[32] X. Zhang, W. Yang, Y. Hu, and J. Liu. Dmcnn: Dual-domain multi-scale convolutional neural network for compression artifacts removal. In *2018 25th IEEE international conference on image processing (icip)*, pages 390–394. IEEE, 2018.

[33] Y. Zhang and H. Yu. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE transactions on medical imaging*, 37(6):1370–1381, 2018.

[34] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018.