FD-MAR: Fourier Dual-domain Network for CT Metal Artifact Reduction

Zilong Li, Qi Gao, Yaping Wu, Chuang Niu, Junping Zhang, Senior Member, IEEE, Meiyun Wang, Ge Wang, Fellow, IEEE, and Hongming Shan, Senior Member, IEEE

Abstract—The presence of high-density objects such as metal implants and dental fillings can introduce severely streak-like artifacts in computed tomography (CT) images, greatly limiting subsequent diagnosis. Although various deep neural networks-based methods have been proposed for metal artifact reduction (MAR), they usually suffer from poor performance due to limited exploitation of global context in the sinogram domain, secondary artifacts introduced in the image domain, and the requirement of precise metal masks. To address these issues, this paper explores fast Fourier convolution for MAR in both sinogram and image domains, and proposes a Fourier dual-domain network for MAR, termed FD-MAR. Specifically, we first propose a Fourier sinogram restoration network, which can leverage sinogram-wide receptive context to fill in the metal-corrupted region from uncorrupted region and, hence, is robust to the metal trace. Second, we propose a Fourier refinement network in the image domain, which can refine the reconstructed images in a local-to-global manner by exploring image-wide context information. As a result, the proposed FD-MAR can explore the sinogram- and image-wide receptive fields for MAR. By optimizing FD-MAR with a composite loss function, extensive experimental results demonstrate the superiority of the proposed FD-MAR over the state-of-the-art MAR methods in terms of quantitative metrics and visual comparison. Notably, FD-MAR does not require precise metal masks, which is of great importance in clinical routine.

Index Terms—Dual-domain network, metal artifact reduction, Fourier network, interpolation, image post-processing.

I. INTRODUCTION

Computed tomography (CT) is one of major modalities widely used for clinical diagnosis and screening. When the high-density objects such as metal implants and dental fillings present, the raw data from the scanner, a.k.a sinogram, are corrupted due to the resulting photon starvation, beam hardening, and scattering. The resulting reconstructed images present severely streak-like artifacts, greatly limiting subsequent diagnosis. How to effectively and robustly reduce the metal artifacts remains challenging and, hence, is gaining increasing attention with the rapid development of deep neural networks.

To address this issue, metal artifact reduction (MAR) algorithms are being developed in both sinogram and image domains. Working in a single domain cannot effectively recover the true tissues from corrupted data. On the one hand, the metal artifacts on the reconstructed images present globally, even worse when the metal becomes large. On the other hand, although only part of region corrupted in the sinogram, the inpainted raw data cannot guarantee desirable image quality after reconstruction. Current popular way to MAR is to combine the advantages of the both domains through a differential Radon layer [2], [9], significantly improving the performance of MAR. For example, Lin et al. [2] used two U-nets to enhance the sinogram and the reconstructed image to achieve promising result. Nevertheless, the limited information captured by convolution layers hinders the further development of the networks. Thus, extra efforts have been made to exploit other information such as image prior [4], [5], metal mask projection [6], adaptive scale [7], etc.

However, the existing MAR methods suffer from unsatisfied and unstable performance. First, in the sinogram domain, the existing methods do not make full use of global context to perform interpolation, leading to globally secondary artifacts after reconstruction. Second, the correction of severely corrupted CT images by deep neural networks may further introduce global inconsistencies; although some work [7] explores non-local methods to eliminate global errors in the image domain, it may not be helpful when the CT images are globally corrupted. Third, most MAR methods use full window to train and evaluate the CT images, which ignore the contrast information at some clinically important CT windows. Lastly, although existing methods have made great improvements in mining information from the corrupted sinogram, these methods still require the precise metal masks, which may be hard to obtain in clinical scenarios. As a result, these methods may become unstable in clinical practice.

Inspired by the success of fast Fourier convolution [8]–[10], we explore the Fourier convolution networks for MAR in both sinogram and image domains, termed Fourier dual-domain network MAR or FD-MAR. More specifically, to accurately fill in corrupted region from the uncorrupted region in the sinogram domain, we propose a novel sinogram restoration network based on fast Fourier convolution, which can leverage sinogram-wide receptive field to perform global interpolation, as demonstrated in Fig. 1. Such a network is robust to the quality of the metal tasks, which is of great importance in clinical scenarios where precise metal masks are hard to obtain due to the metal materials and low image quality. To better
refine the reconstructed images in the image domain, we also propose a novel Fourier refinement network based on fast Fourier convolution, which refines the images in a local-to-global manner.

In addition, due to the high numerical range of CT images, a full-range window may be too wide to emphasize some clinically important windows. To help the network sense accurate feedback at different dynamic ranges, we optimize FD-MAR with a composite loss function extended from the multi-window network [11] for better image quality. Experimental results demonstrate that the proposed method achieves better performance and robustness than the state-of-the-art methods in terms of quantitative metrics and visual comparison.

The contributions of this work are summarized as follows.

1) This paper makes the first attempt at exploiting the fast Fourier convolution to explore the sinogram-wide and image-wide receptive fields for MAR, termed FD-MAR.
2) We propose a Fourier restoration network in the sinogram domain to leverage sinogram-wide context to fill in the metal-corrupted region from uncorrupted region.
3) We propose a Fourier refinement network in the image domain, which can refine the CT images in a local-to-global manner by exploring image-wide context information.
4) Extensive experimental results demonstrate the superiority of FD-MAR over the state-of-the-art MAR methods in terms of quantitative metrics and visual comparison. Notably, FD-MAR does not require precise metal masks, which is of great importance in clinical scenarios.

II. RELATED WORK

A. Deep Learning-based MAR

Due to the complexity of MAR in the image domain, it is difficult to model the metal artifact directly by a mathematical model. Thus, researchers turn to deep neural networks for their excellent expressive capability in modeling complicated artifacts. Supervised networks including convolutional neural networks (CNN) [3], [12] and U-Net [13] are widely used for MAR. Besides, unsupervised methods [14]–[17] are also introduced to address the artifact in a latent space. Generally, processing the corrupted sinogram can address the diverging artifact caused by metal because projection data outside the metal trace can be regarded as clean data [18], [19]. Some neural networks are proposed to generate the corrupted part locally instead of interpolation [20]. One of the challenges for sinogram restoration is the secondary artifact caused by the discontinuities after the restoration process. Since each layer of CNN only has limited respective field, for example a kernel of $3 \times 3$, the long-range discontinuities can only be partially captured by deep layers. To fully explore the local information of surrounded angle and detector, extra prior information are used for further improvement [5], [21]. For example, metal mask projection and adaptive scales [6], [7] can mine more information in the corrupted regions for better performance.

Recently, dual-domain networks are explored to further address the secondary artifact [2], [6], [7], [22]. With two or more coarse enhanced images, dual-domain architecture largely improves the performance over those single domain methods. However, the fundamental problem of lacking global information remains unsolved. This paper explores Fourier neural networks that can easily capture the sinogram-and image-wide context information for global restoration.

B. Fourier Neural Network

Fourier transform has been widely used in digital image processing [23]. Since Fourier transform provides frequency information that is hard to be captured by networks, researchers start to incorporate Fourier transform in neural networks to improve performance [24]–[26]. Thanks to global respective field in frequency domain provided by Fourier transform, researchers leverage this property to help the networks make better use of long-range information [8]–[10] that is difficult to be captured by traditional architecture. For example, [9] designed a Fast Fourier convolution block to replace the vanilla convolution with a fast Fourier unit and perform convolution in frequency domain for global attention. FFC achieves great success in the field of image inpainting [8] with promising results even the masks are large and complex. [10] designed a novel Fourier filter block to replace the existing MLP-mixer block [27] in vision transformer architecture for better computational capacity. This architecture is also shown to achieve good results at different resolution, demonstrating the advantage of Fourier network in capturing global features.

This paper makes the first attempt at exploiting the advantages of Fourier neural networks to perform global interpolation in the sinogram domain and local-to-global refinement in the image domain, achieving better performance than previous state-of-the-art methods.

III. METHOD

A. Problem Formulation

A 2D CT slice of human body shows the distribution of the attenuation coefficients, $X = \mu(i,j)$, where $(i,j)$ indicates the 2D coordinate. Let $X(E)$ be a 2D attenuation coefficient...
image at energy level $E$, the ideal sinogram $S$ can be expressed via Lambert-Beer Law \[28\]:
\[
S = -\ln \int \eta(E) e^{-P(X(E))} dE = P(X),
\]
where $P$ and $\eta(E)$ represent the projection generation process and the energy distribution at $E$, respectively. For the normal tissues, $X(E)$ is almost constant with respect to $E$, i.e. $X = X(E)$. The sinogram $S$ and CT image $X$ can be produced from each other by forward projection, $S = P(X)$, and back projection, $X = F(S)$, respectively; note that $F$ is the inverse operation of $P$.

When metallic implant $X_m(E)$ presents, the metal-corrupted sinogram can be written as:
\[
S_{mc} = P(X) - \ln \int \eta(E) e^{-P(X_{mc}(E))} dE = S \odot (1 - M) + S_{mc} \odot M,
\]
where $S_{mc}$, $S$, $M$ represent the metal-corrupted sinogram, metal-free sinogram, and binary metal trace, respectively. The symbol $\odot$ represents the point-wise multiplication. Here, $M \in \{0,1\}^{H_s \times W_s}$, where $M = 1$ indicates the metal-corrupted region in the sinogram; $H_s$ and $W_s$ are the detector size and the number of projection angles, respectively.

MAR can be achieved in the sinogram domain and image domain.

1) **Sinogram domain:** In the sinogram domain, the goal of MAR is to infer the metal-free sinogram $S$ from the metal-corrupted one $S_{mc}$. There are two different types: sinogram completion and sinogram enhancement.

- **Sinogram Completion** is to fill in the metal-corrupted region, specified by the metal trace, from the non-metal-region through interpolation. In other words, the problem is to infer $S \odot M$ from $S \odot (1 - M)$.

- **Sinogram Enhancement** is to enhance the metal-corrupted region, specified by the metal trace, from the metal-corrupted sinogram \[6, 7\]. In other words, the problem is to infer $S \odot M$ from $S_{mc}$ directly.

In this work, we implement sinogram complete as default FD-MAR, and sinogram enhancement as FD-MAR$_E$.

2) **Image domain:** After the CT image reconstruction by, for example, filtered back projection (FBP), the resulting image $X_{mc} = F(S_{mc})$ suffers from strong metal artifacts. Removing metal artifacts in the image domain is challenging as the metal causes global artifacts, which are mixed with normal tissues.

Algorithm 1: Fast Fourier convolution

| Input: $X \in \mathbb{R}^{B \times H \times W \times C}$ |
| Output: $Y$ |
| 1: $X_{id} = \text{rfft}(X)$ $\quad \triangleright X_{id} \in \mathbb{C}^{B \times H \times \frac{W}{2} \times C}$ |
| 2: $X_d = \text{complex2real}(X_{id})$ $\quad \triangleright X_{id} \in \mathbb{R}^{B \times H \times \frac{W}{2} \times 2}$ |
| 3: $X_t = \text{ReLU}(\text{BN}(\text{Conv}(X_d)))$ |
| 4: $X_d = \text{real2complex}(X_t)$ $\quad \triangleright X_{id} \in \mathbb{C}^{B \times H \times \frac{W}{2} \times C}$ |
| 5: $Y = \text{irfft}(X_{id})$ |

B. Overview of the Proposed FD-MAR

Fig. 2 shows the proposed FD-MAR, which consists of three modules: (a) Fourier sinogram restoration network, termed FS-Net; (b) local U-net restoration network, called LU-Net; and (c) Fourier refinement network, named FR-Net. Each module plays a significantly different role in MAR. FS-Net aims to recover the metal-corrupted region from the non-metal-region. Although the directly reconstructed image from metal-corrupted sinogram presents strong metal artifacts, it also maximally preserves the structural details. Therefore, we use LU-Net, a vanilla U-net, to locally post-process the images for better preserving local details, which serves as a stable base image for refinement. Finally, by combining the base image and reconstructed image from restored sinogram, FR-Net refines the base image in a local to global manner.

The key technique behind FS-Net and FR-Net is the fast Fourier convolution (FFC) \[9\]. By performing convolution in Fourier domain, fast Fourier convolution gains an image-wide respective field for a given image. Since image only contains real numbers, real fast Fourier transform (FFT) and inverse real FFT are widely used in computer vision. With Fourier transform, convolution layer, and inverse Fourier transform, fast Fourier convolution is summarized in Algorithm 1.

FFC is differential and can replace the vanilla convolution block in any network. As shown in Fig. 1, vanilla convolution block is hard to model the artifact in sinogram due to the limited respective field. By transforming the corrupted image and metal trace to Fourier domain, the global bright artifact is evenly spread across the image. Thus, metal artifact is easier
to be removed in Fourier domain. Similarly, the global artifact in image domain can be better modeled by network in Fourier domain.

C. Fourier Sinogram Restoration Network (FS-Net)

To better explore the global context for sinogram inpainting, we propose a novel Fourier sinogram restoration network (FS-Net) based on FFC, which can leverage the sinogram-wide receptive field to accurately restore the metal-corrupted sinogram. As shown in Fig. 2 the input tensor is first split by channel into two branches named local branch and global branch. Here, 3/4 of the channel are split to the global branch with fast Fourier convolution. Three vanilla convolution blocks are first used to capture the multi-scale information. In the meantime, a fast Fourier convolution block restores the sinogram in Fourier domain. With the global context provided by FFC, FS-Net can take full advantage of long-range information to recover the sinogram.

The output of FS-Net is written as:

\[ S_r = \text{FS-Net}(S_{mc}, M), \]

where the inputs include the corrupted sinogram \( S_{mc} \in \mathbb{R}^{H_r \times W_r} \) and the corresponding binary metal trace \( M \) highlighting the metal-corrupted regions in the sinogram. In practice, the binary metal trace can be obtained through thresholding.

Since the non-metal regions are not affected by the metal, we replace the metal-corrupted region with the restored one for reconstruction. By applying a differential Radon layer to the replaced sinogram \([29]\), we obtain a CT image \( X_s \):

\[ X_s = \text{Radon}(S_t \odot M + S_{mc} \odot (1 - M)). \]

We use the \( \ell_1 \) loss to measure the difference in the sinogram domain and the smooth \( \ell_1 \) loss \([30]\) to measure the difference in the image domain to avoid the overfitting and gradient exploding. Therefore, the loss to optimize FS-Net is defined as follows:

\[ L_{FS} = \| S_r - S_{gt} \|_1 + \| X_s - X_{gt} \|_{\text{smth}1}, \]

where \( S_{gt} \) and \( X_{gt} \) are the metal-free ground truth sinogram and image, respectively.

D. Fourier Refinement Network (FR-Net)

Previous work usually uses the coarse refined CT images, such as linear interpolated (LI) image \([20]\), as a stable initial value for refinement \([2, 5, 7]\). Since the LI image is locally interpolated and reconstructed from the corrupted sinogram, the structural errors introduced by it is hard to be eliminated in subsequent processes, \textit{a.k.a.} the secondary artifacts. In order to keep the detail and possible lesion while avoiding secondary artifacts, we use a vanilla U-net to locally process the image reconstructed from the metal-corrupted sinogram, which is termed LU-Net. The process is described as:

\[ X_u = \text{LU-Net}(\text{Radon}(S_{mc})). \]

The loss function to optimize LU-net is the \( \ell_1 \) loss:

\[ L_{LU} = \| X_u - X_{gt} \|_1. \]

Now, we have two different reconstructed images: one reconstructed from restored sinogram, \( X_s \), and one reconstructed from metal-corrupted sinogram that was processed by a LU-Net, \( X_u \). We propose a Fourier refinement network (FR-Net) to refine the base image \( X_u \) with the information from both images. The reason why we select \( X_u \) instead of \( X_s \) is that the \( X_u \) is more stable with only metal artifacts and without potential secondary artifacts.

The proposed FR-Net uses the U-net architecture but in a local-to-global manner. The key novelty lies in the Fourier-based skip connection. The input tensor from encoder is then sent into two branch for local-global refinement. For local branch, we use a network structure of \text{ReLU} \circ \text{BN} \circ \text{Conv} to process the image locally. For the global branch, we apply \text{ReLU} \circ \text{Conv} in Fourier domain to eliminate secondary artifacts globally.

To make the refinement easier, we use residual learning strategy to build the refinement network so that the output can be written as:

\[ X_r = \text{FR-Net}(X_s, X_u) + X_u. \]

The final difference between \( X_r \) and \( X_{gt} \) is measured by a composite loss function to ensure the high fidelity refinement of structural details, including three important image quality metrics: pixel-wise loss, edge loss, and perceptual loss \([31]\).

Pixel-wise loss: We use \( \ell_1 \) loss to measure the pixel-wise difference, which is written as:

\[ L_1 = \| X_r - X_{gt} \|_1. \]

Edge loss: The edge information is critical for clinical diagnosis to measure the boundary information. We used Sobel filter \([32, 33]\) to extract the gradient information for better comparison, which is defined as

\[ L_{el} = \| \text{Sobel}(X_r) - \text{Sobel}(X_{gt}) \|_1. \]

Perceptual loss: To ensure the output of network has the similar texture to that of the metal-free images, we leverage the perceptual loss to extract high-level features for comparison \([34]\). We used a pre-trained VGG-16 network to form a feature extractor \( \phi \). The perceptual loss is defined as

\[ L_{pl} = \| \phi(X_r) - \phi(X_{gt}) \|_F^2, \]

where \( \| \cdot \|_F \) denotes the Frobenius norm. In our experiment, we use the layer 2, 4 and 7 of VGG-16 for calculation.

However, a typical full range window of over [−1000, 2000] HU usually used for training may be too wide to emphasize some clinically important windows. Considering the clinical routine, radiologists usually use different CT window for specific clinical tasks. Therefore, we further extend the three losses above into a multi-window setting \([11]\). In this paper, we use three commonly-used windows: full-range window [−1000, 2000] HU; lung window [−1000, 200] HU; and soft tissue window [−200, 300] HU. We use \( L_{pl}^{W_u} \) and \( L_{pl}^{W_s} \) to represent the summation of pixel-wise loss and edge loss over these three
windows, respectively. For perceptual loss, we concatenate the images under these three windows to form a RGB-like image as the input to VGG network, resulting in a new loss $L^W$. Therefore, the final loss function to optimize the FR-Net is defined as:

$$L_{FR} = L^W_1 + L^W_{cl} + 0.1L^W_{pl}. \quad (13)$$

E. Total Objective Function

The total objective function to optimize our FD-MAR is defined as:

$$L = L_{FS} + L_{LU} + L_{FR}. \quad (14)$$

Here, although we empirically set these three loss functions with equal weights, we find that the results of trained model outperform the existing methods.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

1) Datasets: Following [6], we use DeepLesion [35] to generate the dataset due to its high quality. We generate 360,000 cases and 2,000 cases for training and testing, respectively. All the CT images are resized to $208 \times 208$. 90 metal shapes are used for the training while another 10 metal shapes are used for testing set. The sizes of 10 metal implants in the testing set are: [9, 13, 27, 28, 30, 59, 108, 220, 223, 515]. We use the procedures of [6] to synthesize metal-corrupted sinograms and CT images. We simulate the forward projection and back projection using fan-beam transform. The distance from X-ray source to rotation center is set to 59.5 cm, while 320 projection angles are spaced uniformly between 360 degrees. The number of detector is set to 321. Thus, the sinogram is of size $321 \times 320$, each pixel of sinogram represent the intensity received by the detector on its corresponding angle.

2) Evaluation metrics: We use peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [56] and root mean square error (RMSE) for quantitative evaluations. All these three metrics are widely used for image quality evaluation. Unless noted otherwise, we report the average results of these three windows, i.e. full-range window, lung window, and soft-tissue window.

3) Implementation details: Our model is implemented in PyTorch\(^1\). We use the Adam optimizer [37] with $(\beta_1, \beta_2) = (0.5, 0.999)$ to train the model. To accelerate the training, we pre-train each module for 20 epochs. In the fine-tuning stage, the learning rate starts from $5 \times 10^{-4}$ and is halved for every 6 epoch. The model is trained on NVIDIA 2080Ti GPU for 30 epochs with a batch size of 4.

B. Ablation Study

We first evaluate the effectiveness of each component in FD-MAR. When evaluating the network on sinogram, we only use the reconstructed CT images for fair comparison. The configurations involved in ablation study are as follows:

- a) S-Net: The sinogram restoration network;
- b) FS-Net: S-Net with Fourier blocks that has the same parameters and architecture as S-Net;
- c) LU-Net: Local U-Net for image enhancement;
- d) FR-Net: Fourier refinement network;
- e) LU-Net+FR-Net: MAR in image domain with LU-Net and FR-Net;
- f) FS-Net+LI+FR-Net: Dual-domain method with LI image for residual learning;
- g) FD-MAR w/o m.w.: Our full network (FS-Net+LU-Net+FR-Net) without a multi-window setting;
- h) FD-MAR: FS-Net+LU-Net+FR-Net with multi-window setting;
- i) FD-MAR$E$: FD-MAR with sinogram enhancement while other settings are the same as h.

Table I summarizes the performance of each configuration. By comparing a-e and f-i, dual-domain methods greatly improve the performance. Although e reduces the artifacts twice in the image domain, there was no significant improvement, highlighting the importance of sinogram domain.

1) Effect of Fourier convolutions in sinogram: By comparing a and b in Table I Fourier convolution can better restore the corrupted sinogram. With the same parameters, FS-Net gained a great improvement on all metrics. Detailed comparison between FS-Net of different settings can be found in Sec. IV-D

\(^1\)Sourcecode will be made available upon the acceptance of this work.

### Table I

Quantitative evaluation in the form of \([\text{RMSE} \times 10^{-2}] / \text{PSNR (dB)} / \text{SSIM (%)}\) for different components in FD-MAR. The best two results are highlighted in **BOLD** and **UNDERSCORE**.

| Configuration | Sinogram | Small Metal | Medium Metal | Large Metal | Average |
|---------------|----------|-------------|--------------|-------------|---------|
| S-Net         | 2.78/32.36/92.08 | 2.80/32.32/92.12 | 2.98/31.77/91.83 | 3.72/29.96/90.48 | 4.31/28.74/89.15 |
| FS-Net        | 2.70/32.61/92.44 | 2.73/32.52/92.36 | 2.86/32.10/92.19 | 3.44/30.57/91.13 | 3.99/29.34/90.02 |
| LU-Net        | 1.14/40.41/98.44 | 1.06/41.06/98.54 | 1.22/40.09/98.39 | 2.02/35.55/97.78 | 1.78/36.95/97.39 |
| FR-Net        | 1.11/40.67/98.44 | 1.01/41.42/98.54 | 1.21/40.79/98.40 | 1.98/36.10/97.78 | 1.76/37.08/97.39 |
| LU-Net+FR-Net | 1.11/40.68/98.50 | 1.02/41.47/98.65 | 1.21/40.86/98.55 | 1.99/36.18/97.94 | 1.77/37.39/97.25 |
| FS-Net+LI+FR-Net | 0.82/43.64/99.15 | 0.80/43.92/99.25 | 0.97/42.22/99.19 | 1.44/38.89/98.75 | 1.90/36.56/98.19 |
| FS-Net+LI+FR-Net w/o m.w. | 0.82/43.63/99.14 | 0.73/44.45/99.28 | 0.92/42.71/99.14 | 1.72/37.19/98.67 | 1.43/39.12/98.55 |
| LU-Net+FR-Net | 0.85/43.34/99.14 | 0.72/44.50/99.29 | 0.89/43.09/99.18 | 1.70/37.39/98.73 | 1.38/39.57/98.60 |
| FD-MAR (ours) | 0.76/44.01/99.20 | 0.70/44.47/99.28 | 0.81/43.61/99.19 | 1.47/38.44/98.72 | 1.20/40.31/98.71 |
| FD-MAR$E$ (ours) | 0.84/44.01/99.20 | 0.70/44.47/99.28 | 0.81/43.61/99.19 | 1.47/38.44/98.72 | 0.99/42.17/99.02 |

\(\text{Configuration Small Metal Medium Metal Large Metal Average}\)

\(\text{Configuration Small Metal Medium Metal Large Metal Average}\)
Fig. 3. Visual results of the processing flow by FD-MAR: (a) Metal-corrupted images (inputs); (b) FS-Net; (c) LU-Net; (d) FR-Net and (e) Ground Truth. The display window ranges from [-200, 300] HU for the first row and [-1000, -200] HU for the second row.

### Table II

| Window   | Small Metal | Large Metal | mean ± std  |
|----------|-------------|-------------|-------------|
| g) S.T.  | 1.48        | 1.29        | 2.96        | 2.40        | 1.96 ± 0.70 |
| h) S.T.  | 1.55        | 1.28        | 2.92        | 2.35        | 1.95 ± 0.67 |
| g) L.    | 0.65        | 0.59        | 0.70        | 1.44        | 1.26        | 0.93 ± 0.39 |
| h) L.    | 0.66        | 0.59        | 0.66        | 1.42        | 1.18        | 0.90 ± 0.37 |

2) **Effect of local-global architecture:** Since LI image is reconstructed from LI sinogram, f lacks local restoration while e lacks global information provided by sinogram restoration network. The absence of global or local restoration limits the performance of the architecture especially when the metal is large. We found that when metal is of large size, f is even worse than the single domain network e, mainly due to the secondary artifacts introduced by LI image. By replacing LI image with LU-Net, g has a great improvement in general. Fig. 3 shows the processing flow of FD-MAR. By comparing the results of FS-Net and LU-Net, we observe that with the help of global interpolation, FS-Net successfully recovers the accurate structure in the severely corrupted regions indicated by blue arrows. While LU-Net, which undertakes local information recovery, does not recover the accurate structural information in these areas. However, compared with the outputs of FS-Net, the outputs of local restoration by LU-Net have a higher fidelity with sharper edges obviously, especially in those areas where the artifacts are not severe. Combining the advantages of these two coarse outputs, the final results generated by FR-Net from local to global refinement preserves both the accurate structure and clear details at the same time as shown in Fig. 3(d). Thus, the local-global manner of FD-MAR can lead to better results and interpretability in recovering structure and details.

3) **Effect of the multi-window setting:** By comparing g and h in Table II, h has higher SSIM and PSNR value than g especially when large metals present. In the two groups of the largest metal, PSNR has a 0.2db and 0.45db improvement. Table III shows the RMSE of g and h in soft-tissue window and lung window. Interestingly, the model optimized with a multi-window setting is more robust for metals of different sizes. The effectiveness of multi-window setting is consistent with [11]. In practice, task-specific windows could be selected for optimization.

4) **Effect of sinogram restoration:** By comparing FD-MAR with FD-MAR$_E$, we found that FD-MAR$_E$ largely performs better than FD-MAR on different sizes of metals. However, we emphasize that FD-MAR$_E$ requires precise metal masks, otherwise those methods with sinogram enhancement including FD-MAR$_E$ would become unstable as shown in Sec. IV-D.

### C. Comparison with State-of-the-Art Methods

We compare our model with the following state-of-the-art methods: LI [20], NMAR [21], CNNMAR [3], DuDoNet [2], DSCNet [5], DuDoNet++ [6] and DANNet [7]. LI and NMAR are the widely-used baseline methods in MAR. CNNMAR is an image domain deep-learning-based method. DuDoNet is a state-of-the-art dual domain method using two U-Nets. DSCNet is also a dual-domain method using PriorNet to generate prior image first to guide sinogram learning. Both DuDoNet++ and DANNet are dual-domain networks with sinogram enhancement to better suppress the secondary artifacts. DuDoNet++ uses metal mask projection instead of metal trace to provide more accurate position information in sinogram, while DANNet uses adaptive scale to obtain extra information from the corrupted sinogram. All these models are trained and tested on the same dataset for fair comparison.

1) **Quantitative comparison:** Table III shows the quantitative comparison. Dual-domain methods perform much better than single domain methods. Methods with sinogram enhancement are significantly superior to those with sinogram completion, especially on small metals. Although previous methods on small metals have achieved excellent result, we further demonstrate the potential of Fourier network for MAR. FD-MAR not only significantly improves the performance over those methods with sinogram completion, but also is superior to state-of-the-art methods with sinogram enhancement such as DuDoNet++ and DANNet on large metals and average performance. In addition, FD-MAR$_E$ also performs better than FD-MAR by mining information from the corrupted sinogram while achieving the best results among all methods.

2) **Visual comparison:** Fig. 4 shows the two images with severe artifacts caused by three small implants and two spinal rods. Since the metal blocks most of the scanning angle, model has to learn to make full use of the limited information to recover the corrupted region. Due to the high amount and the large metal size, conventional interpolation-based methods including LI and NMAR fail to remove the artifacts due to the strong structural error introduced by local interpolation in sinogram. We observe similar problem in DuDoNet: in Fig. 4(e), the area near the metal implants are over-smoothed by local pixels. As a result, these images with strong structural errors lost its medical value. DSCNet can suppress the structural errors compared to DuDoNet by reversing the domain order. However, the secondary artifacts are not eliminated as well, which compromise the image quality. By introducing metal mask projection and adaptive scale for extra information, DuDoNet++ and DANNet mitigate the secondary artifacts to some extent in Fig. 4(g) and (h). Among all these state-of-the-art methods, FD-MAR is good at recovering missing details
TABLE III
QUANTITATIVE EVALUATION IN THE FORM OF [RMSE ($\times 10^{-2}$) / PSNR (dB) / SSIM (%)] FOR STATE-OF-THE-ART METHODS. THE BEST RESULTS AMONG DUAL DOMAIN METHODS WITH SINOGRAM COMPLETION AND ENHANCEMENT METHODS ARE HIGHLIGHTED IN UNDERSCORE RESPECTIVELY WHILE THE BEST RESULTS AMONG ALL THE METHODS ARE HIGHLIGHTED IN BOLD.

| Methods                  | Small Metal | Large Metal | Average |
|--------------------------|-------------|-------------|---------|
| Single Domain            |             |             |         |
| Input                    | 3.70/29.12/82.54 | 3.99/28.61/81.88 | 4.13/28.28/81.23 | 5.25/25.66/77.97 | 7.71/22.85/71.34 | 5.03/26.90/78.99 |
| LI [20]                  | 3.49/30.31/89.65 | 3.72/29.83/88.92 | 3.96/29.29/88.61 | 5.22/27.06/86.24 | 6.23/25.48/83.97 | 4.54/28.39/87.40 |
| NMAR [21]                | 3.45/30.58/89.91 | 4.65/29.60/88.46 | 4.55/29.08/88.29 | 5.20/27.34/86.51 | 6.39/25.61/83.95 | 4.82/28.48/87.43 |
| Dual Domain w/ Sinogram  |             |             |         |
| CNNMAR [7]               | 3.14/31.19/91.35 | 3.21/31.01/91.11 | 3.34/30.69/90.88 | 4.13/28.92/89.54 | 5.01/27.41/88.21 | 3.77/28.38/87.14 |
| DuDoNet                  | 0.94/42.34/99.07 | 0.94/42.38/99.05 | 1.19/40.33/98.87 | 1.89/36.51/98.23 | 2.47/34.29/97.46 | 1.49/39.17/98.54 |
| DANNet [7]               | 0.91/42.34/99.07 | 0.94/42.38/99.05 | 1.19/40.33/98.87 | 1.89/36.51/98.23 | 2.47/34.29/97.46 | 1.49/39.17/98.54 |
| DSCNet [5]               | 0.83/43.34/99.14 | 0.72/44.50/99.29 | 0.80/43.99/99.18 | 1.70/37.39/98.73 | 1.38/39.57/98.60 | 1.11/41.58/98.99 |
| FD-MAR [ours]            | 0.63/43.87/99.24 | 0.68/44.41/99.19 | 0.83/43.26/99.06 | 2.02/35.57/98.03 | 1.91/36.27/97.65 | 1.21/40.88/98.85 |
| Dual Domain w/ Sinogram  |             |             |         |
| CNNMAR [7]               | 0.94/42.34/99.07 | 0.94/42.38/99.05 | 1.19/40.33/98.87 | 1.89/36.51/98.23 | 2.47/34.29/97.46 | 1.49/39.17/98.54 |
| DuDoNet++ [6]            | 0.91/42.34/99.07 | 0.94/42.38/99.05 | 1.19/40.33/98.87 | 1.89/36.51/98.23 | 2.47/34.29/97.46 | 1.49/39.17/98.54 |
| DANNet [7]               | 0.94/42.34/99.07 | 0.94/42.38/99.05 | 1.19/40.33/98.87 | 1.89/36.51/98.23 | 2.47/34.29/97.46 | 1.49/39.17/98.54 |
| DSCNet [5]               | 0.83/43.34/99.14 | 0.72/44.50/99.29 | 0.80/43.99/99.18 | 1.70/37.39/98.73 | 1.38/39.57/98.60 | 1.11/41.58/98.99 |
| FD-MAR [ours]            | 0.76/44.01/99.20 | 0.70/44.47/99.28 | 0.81/43.61/99.19 | 1.70/37.39/98.73 | 1.20/40.31/98.71 | 0.99/42.17/99.02 |

Fig. 4. Visual comparison of state-of-the-art methods: (a) Metal-corrupted images (inputs); (b) LI; (c) NMAR; (d) CNNMAR; (e) DuDoNet; (f) DSCNet; (g) DuDoNet++; (h) DANNet (i) FD-MAR (ours) and (j) Ground Truth. The display window ranges from [-200, 300] HU. The metal masks are colored in red.

and removing metal artifacts. Notably, the details in heavily corrupted regions are also correctly reconstructed by FD-MAR in Fig. 4(i), showing the superiority of Fourier network for MAR by fully exploring the global information in dual domains.

D. Robustness to Inaccurate Metal Traces and Masks

In the experimental settings, precise masks and metal traces are used for training and testing. However, in clinical scenarios, the precise metal masks on both sinogram and image are hard to obtain due to the severe artifacts around the metal and the complex composition of the metal implants, and this may lead to an additional error [5]. One simple way is to enlarge the size of metal masks obtained. Therefore, the performance of MAR model under a slightly larger metal mask has important clinical significance. To further investigate the robustness with respect to the size of metal traces and metal masks on both sinogram and the final MAR results, we test different methods with dilated metal trace and metal mask directly to evaluate the robustness of each method. We apply a series of dilation operators to generate larger metal traces and masks gradually. The dilation operator is implemented using the MaxPool2d function in PyTorch with 3 × 3, 5 × 5, and 7 × 7 as the kernel sizes to expand the metal trace and mask; the zero-padding was set to 1, 2 and 3 to maintain the original shape. Together with the original accurate metal trace, we name them Trace0 (accurate trace), Trace3, Trace5, and Trace7, respectively, and so are masks.

TABLE IV
RMSE ($\times 10^{-2}$) FOR DIFFERENT SINOGRAM NETWORKS EVALUATED UNDER DILATED METAL TRACES. THE BEST AND SECOND BEST RESULTS ARE HIGHLIGHTED IN BOLD AND UNDERSCORE, RESPECTIVELY.

| Methods                  | Trace0 mean±std | Trace5 mean±std | Trace7 mean±std |
|--------------------------|-----------------|-----------------|-----------------|
| Mask-U                   | 3.25/42.22/55.53 | 3.35/43.06/55.27 | 3.45/44.05/55.39 |
| Mask-U++                 | 2.17/60.70/73.67 | 2.15/61.49/74.13 | 2.15/61.49/74.13 |
| FS-Net w/o Fourier       | 2.76/35.57/44.33 | 2.75/35.57/44.33 | 2.75/35.57/44.33 |
| FS-Net w 1/4 Fourier     | 2.32/41.01/56.32 | 2.31/41.01/56.32 | 2.31/41.01/56.32 |
| FS-Net (ours)            | 2.14/66.67/77.97 | 2.13/66.67/77.97 | 2.13/66.67/77.97 |
| FS-Net (ours)            | 2.15/70.00/82.37 | 2.14/70.00/82.37 | 2.14/70.00/82.37 |

1) Dilated metal trace: Table IV shows the results of six sinogram restoration networks evaluated on the dilated metal traces:

- **Mask-U**: The widely-used mask U-net [2] with LI sinogram as input;
- **Mask-U++**: mask U-net with metal corrupted sinogram and metal mask projection as input [6]. The metal mask projection is projected from dilated metal mask;
- **FS-Net w/o Fourier blocks**: FS-Net without fast Fourier convolution for global branch in which the fast Fourier convolution reduced to the vanilla convolution;
- **FS-Net w 1/4 Fourier block**: FS-Net with 1/4 channels split to the global branch in the Fast Fourier convolution;
- **FS-Net (ours)**: FS-Net of FD-MAR with 3/4 channels split to the global branch;
- **FS-Net (ours)**: FS-Net of FD-MAR with 3/4 channel split to the global branch.

Obviously, the performances of all six models drop when
Metal traces are dilated. The proposed FS-Net retains both accuracy and robustness. By comparing FS-Net with FS-Net w/o Fourier, and FS-Net w 1/4 Fourier in Table V we found that the overall performance across three dilated metal traces improves significantly as the channels of global branch increase and the standard deviation, showing the robustness with different sizes, is greatly reduced from 0.83 of FS-Net w/o global branch to 0.62 of FS-Net with 3/4 channels split to the global branch. This suggests that, compared to the local interpolation by vanilla convolution, the global interpolation by Fourier convolution can greatly improve the accuracy and stability in sinogram interpolation.

Notably, although FS-Net and Mask-U++ also perform well under the precise metal trace, FS-Net w/o Fourier and FS-Net w 1/4 Fourier in Table V we found that the overall performance across three dilated metal traces improves significantly as the channels of global branch increase and the standard deviation, showing the robustness with different sizes, is greatly reduced from 0.83 of FS-Net w/o global branch to 0.62 of FS-Net with 3/4 channels split to the global branch. This suggests that, compared to the local interpolation by vanilla convolution, the global interpolation by Fourier convolution can greatly improve the accuracy and stability in sinogram interpolation.

![Image](image.png)

**Fig. 5.** Visual comparison of sinogram restoration by Mask-U, Mask-U++ and FS-Net. (a)-(d) represent the results under Trace0, Trace3, Trace5, and Trace7, respectively, and (e) is the ground truth. The blue boxes point out the intersection of two metal traces in sinogram.

### TABLE V
**QUANTITATIVE EVALUATION IN THE FORM OF [PSNR (dB) / SSIM (%)] OF STATE-OF-THE-ART METHODS UNDER DILATED METAL MASKS. THE BEST TWO RESULTS ARE HIGHLIGHTED IN UNDERSCORE AND BOLD, RESPECTIVELY.**

|                  | Mask0          | Mask3          | Mask5          | Mask7          | mean±std       |
|------------------|----------------|----------------|----------------|----------------|----------------|
| DuDoNet          | 39.17/98.54    | 38.00/98.24    | 38.29/98.35    | 38.24/98.34    | 38.43±0.51/98.37±0.13 |
| DSCNet           | 38.79/97.61    | 38.73/97.59    | 38.69/97.59    | 38.66/97.58    | 38.72±0.06/97.59±0.01 |
| DUDENet++        | 40.88/98.63    | 31.00/96.55    | 28.46/94.40    | 26.45/91.49    | 31.70±6.40/95.27±3.05 |
| DANNet           | 41.13/98.88    | 27.89/95.61    | 25.47/94.30    | 24.14/93.58    | 29.66±7.80/95.59±2.35 |
| FD-MAR (ours)    | **42.28/99.02**| **38.77/98.62**| **37.76/98.52**| **36.77/98.36**| **38.90±2.40/98.63±0.28** |
| FD-MAR (ours)    | 41.65/98.99    | **40.69/98.68**| **41.14/98.89**| **41.08/98.89**| **41.14±0.39/98.86±0.13** |

2) Dilated metal mask: In Table V we show the comparison between our method and the state-of-the-art dual-domain methods under dilated masks. By comparing DuDoNet, DUDENet++, and DANNet, we can see that both DUDENet++ and DANNet can obtain higher accuracy than DuDoNet with the help of sinogram enhancement methods when the mask is precise (Mask0). However, as the size of the dilated mask increases, the inaccurate information provided by mask projection and scaled metal projection leads to a rapid performance drop in terms of PSNR and SSIM. In contrast, DuDoNet has significantly higher stability though it does not perform as well as DUDENet++ and DANNet under Mask0. Since DSCNet eliminates artifacts in the image domain first, a slightly inaccurate mask will not affect the sinogram network because the generated prior image remains stable. The problem is the stubborn structural errors introduced by the generated prior image and the remaining secondary artifacts, which is why DSCNet performs poorly in PSNR. In the comparison
between FD-MAR$_E$ and FD-MAR, FD-MAR$_E$ is better than FD-MAR under Mask0 similar to DuDoNet and DANNet. However, when the size of the imprecise mask increases, the PSNR of FD-MAR$_E$ also decreases significantly, but is still better than DuDoNet++ and DANNet because of the global information from uncorrupted regions. Among all the methods, FD-MAR maintains the excellent performance from Mask0 to Mask7, which has a great potential to reduce the errors caused by imprecise metal masks for clinical applications.

Fig. 6 shows the visualization results of state-of-the-art methods under inaccurate masks. (a)-(d) show the results under Mask0, Mask3, Mask5, and Mask7, respectively. (e) is the ground truth. The display window ranges from [-1000, -200] HU. The (dilated) metal masks are marked in red.

Fig. 6. Visual comparison of different methods with inaccurate masks. (a)-(d) show the results under Mask0, Mask3, Mask5, and Mask7, respectively. (e) is the ground truth. The display window ranges from [-1000, -200] HU. The (dilated) metal masks are marked in red.

E. Experimental Results on Clinical Data

We further transfer the model trained on simulated data to clinical images and evaluate the performance of the proposed method. In this retrospective study, four patients with metal implants in the abdomen were scanned at Henan Provincial People’s Hospital (Zhengzhou, Henan, China) using a UIH uCT960+ CT scanner with 100 kVp and 370 mA. The clinical data were de-identified. We highlight that directly transferring the model from simulated data to clinical data is challenging due to the different scanning and reconstruction methods, different components and density of the metal implants, and no precise metal masks available. Since sinogram data is not accessible in clinical applications, we follow the previous work [2] to test our method on the clinical CT images. The metal mask was segmented by a threshold of over 2000HU.

As shown in Fig. 7, LI eliminates the metal artifacts in the four cases but leads to additional global errors. DuDoNet++ and DANNet still contain noticeable artifacts. We noticed that sinogram enhancement based methods may lead to erroneous images possibly due to the sensitivity to metallic materials and masks. In comparison, the proposed FD-MAR successfully recovers the structural details from the artifact in both local and global aspects. Notably, FD-MAR maintains the performance to a large extent under a dilated mask in Fig. 7(b). We highlight that our method is robust to a low quality of masks resulting from the thresholding in the third row; this, however, typically happens in clinical routine. The results suggest that imprecise (dilated) metal masks or coarse metal masks by thresholding can be used in clinical applications with satisfactory results.

V. DISCUSSION AND CONCLUSION

Although many methods make progress in extracting the information from the corrupted region in the sinogram domain, we emphasize that these methods may suffer from potential limitation in clinical scenarios, where the composition of metal implants is complicated and the precise masks are hard to obtain. By reducing metal artifacts in a local-global aspect by fast Fourier convolution, our method can avoid the potential risk while keeping the clinical details to a large extent.

We acknowledge some limitations. One limitation is the additional computational cost introduced, especially by Fourier transform and that is why we apply it on small-size feature maps rather than the original input. Moreover, it is also worth studying how to incorporate the geometric information for global interpolation in the sinogram domain, in which such domain knowledge may give more interpretability in clinical applications.

In conclusion, this paper explored fast Fourier convolution for metal artifact reduction, which can leverage sinogram- and
image-wide receptive fields to perform global interpolation in the sinogram domain and refine images in the image domain, respectively. The proposed FD-MAR achieved better performance and stability than state-of-the-art baseline methods. This study could shed new light on the importance of Fourier convolution for MAR.

REFERENCES

[1] G. Wang, J. C. Ye, and B. De Man, “Deep learning for tomo- graphic image reconstruction,” Nat. Mach. Intell., vol. 2, no. 12, pp. 737–748, 2020.
[2] W.-A. Lin et al., “DuDoNet: Dual domain network for CT metal artifact reduction,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2019, pp. 10 512–10 521.
[3] Y. Chang and H. Yu, “Convolutional neural network based metal artifact reduction in x-ray computed tomography,” IEEE Trans. Med. Imaging, vol. 37, no. 6, pp. 1370–1381, 2018.
[4] Z. Chang, D. H. Ye, S. Srivastava, J.-B. Thibault, K. Sauer, and C. Bouman, “Prior-guided metal artifact reduction for iterative x-ray computed tomography,” IEEE Trans. Med. Imaging, vol. 58, no. 6, pp. 1532–1542, 2018.
[5] L. Yu, Z. Zhang, X. Li, and L. Xing, “Deep sinogram completion with image prior for metal artifact reduction in CT images,” IEEE Trans. Med. Imaging, vol. 40, no. 1, pp. 228–238, 2020.
[6] 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 Int. Conf. Med. Image Comput. Comput. Assist. Interv., 2020, pp. 147–157.
[7] T. Wang et al., “Dual-domain adaptive-scaling non-local network for CT metal artifact reduction,” in Int. Conf. Med. Image Comput. Comput. Assist. Interv. Springer, 2021, pp. 243–253.
[8] R. Suvorov et al., “Resolution-robust large mask inpainting with Fourier convolutions,” in Proc. IEEE Winter Conf. Appl. Comput. Vis., 2022, pp. 2149–2159.
[9] L. Chi, B. Jiang, and Y. Mu, “Fast Fourier convolution,” in Adv. Neural Inf. Process. Syst., vol. 33, 2020, pp. 4479–4488.
[10] Y. Rao, W. Zhao, Z. Zhu, J. Lu, and J. Zhou, “Global filter networks for image classification,” Adv. Neural Inf. Process. Syst., vol. 34, 2021.
[11] C. Niu, M. Li, and G. Wang, “Multi-window learning for metal artifact reduction,” in Developments in X-Ray Tomography XIII, vol. 11840. International Society for Optics and Photonics, 2021, p. 1184015.
[12] X. Huang, J. Wang, F. Tang, T. Zhong, and Y. Zhang, “Metal artifact reduction on cervical CT images by deep residual learning,” Biomed. Eng. Online, vol. 17, no. 1, pp. 1–15, 2018.
[13] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional networks for biomedical image segmentation,” in Int. Conf. Med. Image Comput. Comput. Assist. Interv. Springer, 2015, pp. 234–241.
[14] H. Liao, Z. Huo, W. J. Sehnert, S. K. Zhou, and J. Luo, “Adversarial sparse-view CBCT artifact reduction,” in Int. Conf. Med. Image Comput. Comput. Assist. Interv., 2018, pp. 154–162.
[15] H. Liao, W. Lin, S. K. Zhou, and J. Luo, “ADN: artifact disentanglement network for unsupervised metal artifact reduction,” IEEE Trans. Med. Imaging, vol. 39, no. 3, pp. 634–643, 2019.
[16] C. Niu et al., “Low-dimensional manifold constrained disentanglement network for metal artifact reduction,” IEEE Trans. Radiat. Plasma Med. Sci., vol. 6, no. 6, pp. 656–666, 2022.
[17] J. Lee, J. Gu, and J. C. Ye, “Unsupervised ct metal artifact learning using attention-guided β-CycleGAN,” IEEE Trans. Med. Imaging, vol. 40, no. 12, pp. 3932–3944, 2021.
[18] G. Wang, D. L. Snyder, J. A. O’Sullivan, and M. W. Vannier, “Iterative deblurring for CT metal artifact reduction,” IEEE Trans. Med. Imaging, vol. 15, no. 5, pp. 657–664, 1996.
[19] G. Wang, T. Frei, and M. W. Vannier, “Fast iterative algorithm for metal artifact reduction in X-ray CT,” Acad. Radiol., vol. 7, no. 8, pp. 607–614, 2000.
[20] W. A. Kalender, R. Hebel, and J. Ebersberger, “Reduction of CT artifacts caused by metallic implants,” Radiology, vol. 164, no. 2, pp. 576–577, 1987.
[21] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, “Normalized metal artifact reduction NMAR in computed tomography,” Med. Phys., vol. 37, no. 10, pp. 5482–5493, 2010.
[22] H. Wang et al., “InDuDoNet: An interpretable dual domain network for CT metal artifact reduction,” in Int. Conf. Med. Image Comput. Comput. Assist. Interv. Springer, 2021, pp. 107–118.
[23] G. A. Baxes, Digital image processing: principles and applications. John Wiley & Sons, Inc., 1994.
[24] Y. Yang and S. Soatto, “FDA: Fourier domain adaptation for semantic segmentation,” in Proc. IEEE/CVF Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2020, pp. 4085–4095.
[25] J.-H. Lee, M. Heo, K.-R. Kim, and C.-S. Kim, “Single-image depth estimation based on Fourier domain analysis,” in Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2018, pp. 330–339.

[26] X. Mao, Y. Liu, W. Shen, Q. Li, and Y. Wang, “Deep residual Fourier transformation for single image deblurring,” arXiv preprint arXiv:2111.11745, 2021.

[27] I. Tolstikhin et al., “MLP-Mixer: An all-MLP architecture for vision,” Adv. Neural Inf. Process. Syst., vol. 34, 2021.

[28] Beer, “Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten,” Annalen der Physik, vol. 162, no. 5, pp. 78–88, 2010.

[29] M. Ronchetti, “TorchRadon: Fast differentiable routines for computed tomography,” arXiv preprint arXiv:2009.14788, 2020.

[30] R. Girshick, “Fast R-CNN,” in Proc. IEEE Int. Conf. Comput. Vis., 2015, pp. 1440–1448.

[31] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in Eur. Conf. Comput. Vis. Springer, 2016, pp. 694–711.

[32] Z. Huang, J. Zhang, Y. Zhang, and H. Shan, “DU-GAN: Generative adversarial networks with dual-domain U-Net based discriminators for low-dose CT denoising,” IEEE Trans. Instrum. Meas., vol. 71, p. 4500512, 2022.

[33] H. Shan et al., “Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction,” Nat. Mach. Intell., vol. 1, no. 6, pp. 269–276, 2019.

[34] C. Peng et al., “A cross-domain metal trace restoring network for reducing x-ray CT metal artifacts,” IEEE Trans. Med. Imaging, vol. 39, no. 12, pp. 3831–3842, 2020.

[35] K. Yan, X. Wang, L. Lu, and R. M. Summers, “DeepLesion: Automated deep mining, categorization and detection of significant radiology image findings using large-scale clinical lesion annotations,” arXiv preprint arXiv:1710.01766, 2017.

[36] W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity,” IEEE Trans. Image Process., vol. 13, no. 4, 2004.

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