DAN-Net: Dual-domain adaptive-scaling non-local network for CT metal artifact reduction

Tao Wang¹, Wenjun Xia¹, Yongqiang Huang¹, Huaiqiang Sun², Yan Liu³, Hu Chen³, Jiliu Zhou¹ and Yi Zhang¹

¹ College of Computer Science, Sichuan University, Chengdu 610065, People’s Republic of China
² Department of Radiology, West China Hospital of Sichuan University, Chengdu 610041, People’s Republic of China
³ College of Electrical Engineering, Sichuan University, Chengdu 610065, People’s Republic of China

E-mail: yzhang@scu.edu.cn

Keywords: computed tomography, image reconstruction, metal artifact reduction, deep learning

Abstract
Metallic implants can heavily attenuate x-rays in computed tomography (CT) scans, leading to severe artifacts in reconstructed images, which significantly jeopardize image quality and negatively impact subsequent diagnoses and treatment planning. With the rapid development of deep learning in the field of medical imaging, several network models have been proposed for metal artifact reduction (MAR) in CT. Despite the encouraging results achieved by these methods, there is still much room to further improve performance. In this paper, a novel dual-domain adaptive-scaling non-local network (DAN-Net) is proposed for MAR. We correct the corrupted sinogram using adaptive scaling first to preserve more tissue and bone details. Then, an end-to-end dual-domain network is adopted to successively process the sinogram and its corresponding reconstructed image is generated by the analytical reconstruction layer. In addition, to better suppress the existing artifacts and restrain the potential secondary artifacts caused by inaccurate results of the sinogram-domain network, a novel residual sinogram learning strategy and non-local module are leveraged in the proposed network model. Experiments demonstrate the performance of the proposed DAN-Net is competitive with several state-of-the-art MAR methods in both qualitative and quantitative aspects.

1. Introduction

Computed tomography (CT) technology has developed rapidly in clinical, industrial, security and other spheres (Cuadros et al. 2019, Han and Baek 2019, Zhao et al. 2020). With the help of CT images, medical diagnosis and treatments can be conducted effectively. However, the effects of noise, photon starvation, beam hardening, scattered radiation and nonlinear partial volume effects are much more severe in the case of metallic implants in scanned regions (Cuadros et al. 2019). Due to these metallic objects, the reconstructed CT images are contaminated by heavy artifacts called ‘metal artifacts’ specifically. These artifacts degrade the imaging quality and severely compromise doctors’ diagnoses. In particular, some artifacts and certain lesions have considerable commonalities, leading to misdiagnosis, and subsequent medical image analysis is difficult (Zhao et al. 2020). Therefore, it is of great significance to reduce metal artifacts in CT images.

During the past several decades, numerous metal artifact reduction (MAR) methods have been proposed. Conventional MAR methods can be grouped into three categories: projection completion methods, iterative reconstruction methods and image postprocessing methods (Mouton et al. 2013). The projection completion methods regard projection data in the metal trace as missing and fill in lost data with estimated values by different interpolation strategies (Zhao et al. 2002, Gu et al. 2006, Mehranian et al. 2013) or image inpainting methods (Duan et al. 2008, Xue et al. 2009, Zhang et al. 2011). Interpolation-based approaches are widely adopted but can hardly guarantee smoothness at the interpolation boundaries (Jeong and Ra 2009). After filtering, discontinuities are amplified at the metal trace boundaries, which introduce new artifacts into the reconstructed CT images. To fully explore the local information in both dimensions of the angle and detector bin, some
diffusion-based image inpainting methods were introduced for projection completion (Duan et al. 2008, Zhang et al. 2011). Although these methods may mitigate the discontinuity to some extent, extra artifacts are still inevitable in the reconstructed images. To smooth the transition region between the metal and nonmetal portions and to suppress secondary artifacts, some prior image-based methods have been proposed (Meyer et al. 2010, Wang et al. 2013, Li et al. 2015), such as the normalized metal artifact reduction (NMAR) method (Meyer et al. 2010). NMAR normalized the projection data with the constraint of prior images obtained by multi-threshold segmentation based on interpolation methods. Corrected CT images can be derived from completed sinograms by filtered back projection (FBP). However, the result of NMAR is limited by the quality of the prior image. In addition, FBP is based on the line integral model, which does not take into account the statistical characteristics of measured data and simply assumes that the measured data are noiseless and all response lines have the same weight, which is not always consistent with the real situation. Iterative reconstruction is an alternative way to tackle these problems, which improves image quality gradually based on constrained optimization, such as the least square method and maximum likelihood. Classical iterative reconstruction MAR methods can be divided into two groups. One uses projection data outside of the metal trace, which can be regarded as clean data (Wang and Snyder 1996, Wang et al. 1999, 2000, Mehranian et al. 2013a, 2013b, Zhang et al. 2018). The other adopts a statistical objective function to decay corrupted projection data (De Man et al. 2000, Van Slambrouck and Neyts 2012). However, iterative methods are usually time-consuming and require manually well-designed regularizers, both of which bring difficulties to clinical application. Image postprocessing methods (Ballhausen et al. 2014, Soltanian-Zadeh et al. 2016) aim to reduce metal artifacts in the image domain without accessing raw projection data. However, since the noise and artifacts in CT images do not obey any specific statistical distribution, postprocessing methods usually cannot suppress the artifacts well and are apt to distort the anatomic structure (Mouton et al. 2013, Yu et al. 2020).

Recently, with the successful applications of deep learning (DL) in many fields (LeCun et al. 2015, Guo et al. 2016, Miotto et al. 2018, Wang et al. 2018), DL-based methods have shown great potential for medical imaging (Chen et al. 2017, Yongqiang et al. 2019). Different network architectures, such as convolutional neural networks (CNNs) and generative adversarial networks, have been utilized to recover the missed data in the metal trace (Long et al. 2015, Ghanii and Karl 2018, Park et al. 2018, Ghanii and Karl 2019, Pimkin et al. 2020). Meanwhile, some studies have been dedicated to using DL methods to reduce metal artifacts in the image domain. Zhang and Yu (2018) proposed a CNN framework (CNNMAR) fusing different MAR methods to improve the performance of artifact reduction. To eliminate metal artifacts from original CT images, Liao et al. (2019) introduced a novel unsupervised artifact disentanglement network (ADN). Gjesteby et al. (2019) took detailed images derived from filtering and base images by NMAR as inputs and mapped them to metal artifact-free images with a dual-stream residual network.

Despite the encouraging results achieved by the abovementioned sinogram- or image domain-based DL methods, there are still some limitations in single-domain methods (Lin et al. 2019). For sinogram domain-based DL methods, although corrupted projections within metal traces are local, it is difficult to preserve continuity at metal trace boundaries, where secondary artifacts can be introduced easily. In terms of the image domain-based methods, the input CT images reconstructed from corrupted sinograms are full of severe artifacts, which cover most of clinical important details. These images with low quality may lead to misclassification of some structures and artifacts due to their similar patterns. Thus, the goal for MAR becomes twofold. The first goal is to eliminate existing artifacts as much as possible, and the other is to avoid introducing extra artifacts. To this end, combining the merits of both projection and image domain-based methods is meaningful and some end-to-end dual domain networks were proposed very recently. Lin et al. (2019) proposed DuDoNet, which progressively restores sinogram consistency and enhances CT images linked by a differentiable radon inversion layer. Lyu et al. (2020) proposed to improve DuDoNet by specifying the metal mask projection and encoding it into the network. Yu et al. (2020) proposed employing an image-domain network to generate a prior image at first. Then, the sinogram obtained from the prior image was utilized to guide the sinogram-domain network. In Peng et al. (2020), the authors proposed using partial convolution to recover irregular metal trace regions with only valid pixels outside the corrupted areas. Furthermore, an auxiliary inpainting network is introduced to suppress the secondary artifacts in the reconstructed image from the previous step. Both sinograms from the last two steps were fused to generate the final result.

Due to their state-of-the-art performance, dual-domain networks have become the mainstream for MAR. However, current dual domain-based methods still suffer from some critical limitations. Lin et al. (2019), Peng et al. (2020) regarded projection data in the metal trace as missing, which resulted in the loss of details near the metal area in reconstructed CT images. Yu et al. (2020), Lyu et al. (2020) used metal corrupted projection data and corresponding reconstructed CT images as inputs directly. Actually, the data in the metal trace have a much higher amplitude than the data outside the metal trace, and there is a rapid change at the boundary of the metal trace. According to the CT imaging principle (Lyu et al. 2020), due to this amplitude difference, data inside and outside of the metal trace can be regarded as obeying two different data distributions. It is difficult for neural
networks to transform two different data distributions into a uniform distribution. Lin et al (2019) experimentally found that their method did not perform well while taking original sinogram and CT images as inputs. Meanwhile, the change in the boundary will cause weak continuity of the first derivative of projection data in a certain section, which will be further expanded by filtering and will generate extra artifacts (Pan 1999). In addition, artifacts are non-local in the image, which is hard to remove.

To address the problems mentioned above, a novel dual-domain adaptive-scaling non-local network (DAN-Net) for MAR is proposed. The projection data are considered to be composed of two parts: one part comes from the tissues and the other part comes from the metal objects. A rough estimation of tissue-like projection data in the metal trace is obtained by a linear interpolation (LI) operation, and the residual between it and the original projection data is regarded as the contribution of the metal. To weaken the rapid change caused by metal implants and retain the data characteristics, the residual part in the metal trace is adaptively scaled (Kachelrieß et al. 2001, Chen et al. 2002, Watzke and Kalender 2004). The results of this adaptive scaling and corresponding reconstruction by FBP are used as the inputs of our network. In addition, a novel residual sinogram learning strategy is applied in the sinogram-domain network to weaken the rapid change in projection data and improve the smoothness of the projection. To handle the nonlocality of artifacts, a non-local U-Net architecture is employed for image-domain enhancement, which captures long-range dependencies via non-local operations. The whole network is trained in an end-to-end manner so that the image-domain enhancement and sinogram-domain enhancement can benefit each other.

Our main contributions are summarized as follows.

(1) Different from current dual-domain networks, the original sinogram is preprocessed using adaptive scaling and taking the scaled metal projection and its corresponding FBP result as the inputs, which can preliminarily suppress metal artifacts and maintain tissue details.

(2) A novel residual sinogram learning strategy is proposed to avoid transforming two different data distributions into a uniform one and to improve the smoothness of the corrected projection.

(3) A non-local U-Net architecture is designed for image-domain enhancement, which can capture long-range dependencies of metal artifacts and further improve image quality.

The remainder of this paper is organized as follows. The proposed DAN-Net is elaborated in section 2. The experiments and results for the simulated and clinical data are presented in section 3. The results of analytical studies are shown in section 4. Discussion and the conclusion are provided in section 5.

2. Method

2.1. Problem formulation

In our work, we consider the case of a 2D attenuation distribution. If there are metallic objects in the scanner field, the linear attenuation image \( X(E) \) at energy level \( E \) can be expressed as follows:

\[
X(E) = X_{\text{tissue}}(E) \odot (1 - M) + X_{\text{metal}}(E) \odot M, \tag{1}
\]

where \( X_{\text{tissue}}(E) \) and \( X_{\text{metal}}(E) \) represent the attenuation image to be reconstructed and the metal part. \( M \) denotes the metal mask in \( X(E) \) and \( \odot \) is the elementwise multiplication. The projection data \( S_{\text{ma}} \), contaminated by metals, can be calculated as follows (Lin et al. 2019):

\[
S_{\text{ma}} = - \ln \int \eta(E) \exp(-\mathcal{P}(X(E))) \, dE = - \ln \int \eta(E) \exp(-\mathcal{P}(X_{\text{tissue}}(E) \odot (1 - M) + X_{\text{metal}}(E) \odot M)) \, dE = - \ln \int \eta(E) \exp(-\mathcal{P}(X_{\text{tissue}}(E) \odot (1 - M))) \, dE + \int \eta(E) \exp(-\mathcal{P}(X_{\text{metal}}(E) \odot M)) \, dE = S_{\text{tissue}} + S_{\text{metal}}, \tag{2}
\]

where \( \eta(E) \) denotes the intensity distribution with spectral energy at \( E \) and \( \mathcal{P} \) is the forward projection operation. As shown in equation (2), \( S_{\text{ma}} \) can be regarded as containing two parts: one is contributed by the attenuation of tissues, denoted as \( S_{\text{tissue}} \) and the other is produced by metal objects, denoted as \( S_{\text{metal}} \). For MAR, if we simply discard the projection data in the metal trace, projections contributed from both tissue and metal will be lost, and the reconstructed CT image has to take the risk of losing tissue details around the metallic implants. On the other hand, for the LI based method, the projection data in the metal trace are usually estimated by performing LI, referred to as \( S_{\text{LI}} \). The residual between \( S_{\text{ma}} \) and \( S_{\text{LI}} \), notated as \( S_{\text{sub}} = S_{\text{ma}} - S_{\text{LI}} \), is regarded as the metals’ contribution. Ideally, it is expected that \( S_{\text{LI}} \approx S_{\text{tissue}} \) and \( S_{\text{sub}} \approx S_{\text{metal}} \). However, \( S_{\text{LI}} \) is just
a coarse estimation of $S_{\text{true}}$, and some useful information is still reserved in $S_{\text{sub}}$. Based on these considerations, our method attempts to retrieve the rest valuable information from $S_{\text{sub}}$.

2.2. The proposed DAN-Net
To simultaneously leverage the advantages of both sinogram- and image-domain information, we adopt a dual-domain joint learning strategy for CT MAR, and back-propagation of gradients is conducted by the analytical reconstruction layer. Figure 1 depicts the overview of our proposed DAN-Net. More details are presented in subsequent sections.

2.2.1. Adaptive scaling
When x-rays pass through a metal material with high attenuation coefficients, the intensity of low energy will be significantly reduced. At this time, the beam-hardening effect will be more pronounced, leading to an abrupt change in projection data at metal trace boundaries. As we mentioned before, this change will raise more artifacts after filtering. As equation (2) shows, projection data penetrating through metals usually contains two parts. One is contributed by the tissues, and another is contributed by metals. For interpolation-based methods, this part of projection data is treated as contaminated and erased from the original projection. By doing this, the metal artifacts can be largely suppressed in the reconstructed image. However, the removed projection data contains all the information of the metals as well as some information of bones around the metals. This leads to
the loss of metals and blurred bones in the reconstructed image after correction. In DL-based MAR, Lin et al (2019) used LI corrected projection data and corresponding reconstructed CT images as the network inputs, which reduced most artifacts but resulted in the loss of details around the metals. In Lyu et al (2020), Yu et al (2020), the authors directly took metal corrupted projection data and metal-contaminated CT images as inputs, which preserved more details in the final corrected CT images but some artifacts are also preserved. To simultaneously take advantage of both strategies, eliminate the rapid shift in projection data caused by the metal and maintain more useful information, in this paper, adaptive scaling (Chen et al 2002) is adopted, which can be written as the following formula for simplicity:

\[
\begin{align*}
S_{sub} &= S_{ma} - S_{LI} \\
S_{ret} &= \lambda \times S_{sub} \\
S_{pre} &= S_{LI} + S_{ret}
\end{align*}
\]

where \( \lambda \in [0, 1] \) is the scaling parameter to control the trade-off between artifact reduction and detail preservation around the metallic implant in the final reconstructed CT images. As shown in equations (3)–(5), instead of simply removing all the projection data corrupted by the metals, the metal projection is adaptively scaled to compensate for the inaccurate LI-based correction and the value of the metal projection is lowered down by multiplying it by the scaling parameter. This lowered projection is devoted to diminishing the impact of some errors within the metal projections when back-projecting them to the positions of other tissue in the reconstructed image. \( S_{ret} \) and \( S_{pre} \) represent the scaled metal projection and the corrected projection after adaptive scaling, respectively. As a result, a greater \( \lambda \) will keep more tissue details but lead to more artifacts as well, while a smaller \( \lambda \) will generate fewer artifacts but lose more tissue details. Typically, the value of \( \lambda \) is set between 0.3 and 0.5 according to Chen et al (2002) and we chose \( \lambda = 0.4 \) experimentally in this paper. The corresponding adaptively scaled CT images is obtained as \( X_{pre} = \mathcal{P}^{-1}(S_{pre}) \), where \( \mathcal{P}^{-1} \) denotes FBP operator.

Figure 2 shows the results of one example after adaptive scaling. All images were reconstructed from corresponding sinogram by FBP using R-L filter. 2D parallel-beam geometry is adopted for simulation. 367 detector bins and 361 sampling views uniformly distributed from 0° to 180° are assumed. Due to the mathematical property of LI, which discards the original corrupted projection data within metal trace region and uses projection data outside the metal traces to obtain a rough estimation, the result of LI cannot preserve all the details well and seems blurry. It can be seen that (D2) has less artifacts than (B2) and more bone and tissue details, especially nearing to the metal, are preserved than (C2). To further demonstrate the ability of adaptive scaling for structure preservation, the profile, which are indicated by a yellow line in figure 2(A2), is plotted in figure 3. It can be noticed that in the region marked by a green box, the result after adaptive scaling has a more consistent shape to the ground truth and LI smooths the peak belonging to a bone.

2.2.2. Sinogram domain network

To complete the sinogram, we train a neural network \( G_{sino} \) to process the projection data. If we take only the LI corrected sinogram \( S_{LI} \) as the input of \( G_{sino} \) the CT image reconstructed from the output of \( G_{sino} \) will be oversmoothed, and some tissue details will be lost (Lyu et al 2020). On the other hand, it is challenging to restore information directly from the original corrupted sinogram because the projection data inside and outside the metal trace follow two different distributions. To remedy these drawbacks, instead of taking the original
sinogram or L1 refined result as the input of $G_{sino}$, we propose a residual sinogram learning strategy for $G_{sino}$, e.g. taking $S_{res}$ as input to enhance the smoothness of projection data, retrieving useful information from the metal mask region $M_t$ and improving the continuity at the metal trace boundary. Meanwhile, current networks contain down-sampling operations, which will cause the information loss of metal projection (Ghani and Karl 2019). In this work, U-Net is utilized as the backbone of $G_{sino}$ and the details of $G_{sino}$ are shown in figure 4.

To retain sufficient information of metal projection, a mask pyramid network (Liao et al 2019) is introduced to explicitly feed the metal projection information into each layer (Lyu et al 2020). Thus, we have

$$M_p = \mathcal{P}(M)$$

(6)

$$M_t = \delta(M_p > 0),$$

(7)

where $\delta(\cdot)$ is a binary indicator function. Since our main goal is to retrieve information in the metal trace, we only refine the sinogram in the metal trace. The corrected sinogram can be written as

$$F_{sino} = G_{sino}(S_{res}, M_p, M_t)$$

(8)

$$S_{sino} = F_{sino} + S_{gt}.$$ 

(9)

To preserve the details around the metals and avoid over smoothing, sinogram loss $\mathcal{L}_{sino}$, which is implemented by L1 norm, is adopted to measure the differences between $S_{sino}$ and the ground truth $S_{gt}$ according to Lin et al (2019), Lyu et al (2020), Yu et al (2020) and Peng et al (2020) as:

$$\mathcal{L}_{sino} = \| (S_{sino} - S_{gt}) \odot M_t \|_1,$$ 

(10)

where $S_{gt}$ is the sinogram which is not been contaminated by metal and is obtained by performing forward projection operation on clean CT image following the same procedure as Zhang and Yu (2018), Lin et al (2019), Lyu et al (2020) and Yu et al (2020). $\|\cdot\|_1$ represents the L1 norm. Then, $X_{sino} = \mathcal{P}^{-1}(S_{sino})$ can be obtained using an analytical reconstruction layer, which is differentiable and easily injected into neural networks. To alleviate the secondary artifacts in the reconstructed CT image, the reconstruction loss $\mathcal{L}_{\text{FBP}}$ between $X_{sino}$ and the ground truth image $X_{gt}$ is utilized with L1 norm as:

$$\mathcal{L}_{\text{FBP}} = \| (X_{sino} - X_{gt}) \odot (1 - M) \|_1.$$ 

(11)

2.2.3. Image domain net

To suppress the secondary artifacts introduced by the errors of projection data completion in $G_{sino}$, we also utilize U-Net as the backbone to enhance the reconstructed CT images. For computational efficiency, we halve the channel numbers. It is well known that convolution is a local operator whose receptive field is limited by the size of filters. Once the network is insufficiently deep, it is difficult to capture the latent features in long-range dependencies. Since metal artifacts are non-local, convolution-based postprocessing methods may fail to remove the artifacts well. To tackle this problem, a non-local network (NLN) (Wang et al 2018), which can capture long-range dependencies via non-local operations, is introduced into our proposed image domain.
Figure 5. An illustration of the NLN. $X \in \mathbb{R}^{T \times C \times H \times W}$ and $Z \in \mathbb{R}^{T \times C \times H \times W}$ are the input and output feature maps, respectively. $T$, $C$, $H$ and $W$ represent the batch size, channel numbers, height, and width of the input feature maps, respectively.

Figure 6. An illustration of the $G_{\text{im}}$. K: kernel, S: stride, P: padding sizes and D: dilation.

Table 1. Summary of the symbols used in this paper.

| Categories        | Symbols | Meaning                           |
|-------------------|---------|-----------------------------------|
| Adaptive scaling  | $S_{\text{metal}}$ | Metal-contaminated projection    |
|                   | $S_{L1}$ | Linear interpolation corrected sinogram |
|                   | $S_{\text{sub}}$ | Residual of $S_{\text{metal}}$ and $S_{L1}$ |
|                   | $S_{\text{res}}$ | Scaled metal projection          |
|                   | $X_{\text{pre}}$ | Adaptively scaled sinogram       |
|                   | $X_{\text{sino}}$ | Adaptively scaled CT image       |
| Sinogram domain   | $G_{\text{sino}}$ | Sinogram Domain Network          |
|                   | $S_{\text{sino},c}$ | Corrected Sinogram after $G_{\text{sino}}$ |
|                   | $X_{\text{sino}}$ | Reconstructed CT image from $S_{\text{sino}}$ |
|                   | $M_T$     | Metal mask projection            |
|                   | $M_t$     | Metal trace                      |
| Image domain      | $G_{\text{im}}$ | Image Domain Network             |
|                   | $X_{\text{im}}$ | Reconstructed CT image from $S_{\text{im}}$ |
|                   | $X_{\text{out}}$ | Outputs of $G_{\text{im}}$      |
|                   | $M$       | Metal mask                       |
| Ground truth      | $X_{\text{gt}}$ | Ground truth CT image            |
|                   | $S_{\text{gt}}$ | Ground truth CT sinogram         |
| Loss              | $\mathcal{L}_{\text{sino}}$ | Sinogram loss                    |
|                   | $\mathcal{L}_{\text{FBP}}$ | Reconstruction loss              |
|                   | $\mathcal{L}_{\text{im}}$ | Image domain network loss        |
| Others            | $\mathcal{P}$  | Forward projection               |
|                   | $\mathcal{P}^{-1}$ | FBP                              |
network $G_{im}$, NLNs originate from the non-local means denoising method (Buades et al 2005). Different from non-local means, which performs weighted summation with similar pixels, NLN captures feature maps globally. A generic non-local operation is defined as

$$y_j = \frac{1}{c(x)} \sum_{j \in S} f(x_i, x_j) g(x_j),$$

where $x_i$ represents the $i$th element to be replaced, and $y_j$ is the result. $S$ represents a search window. The pairwise function $f$ computes the similarity between $x_i$ and $x_j$, which is expressed as follows:

$$f(x_i, x_j) = \exp(\theta_i(x_i)^T) \exp(\theta_j(x_j)),$$

where $\theta_i(x_i) = W_1 x_i$ and $\theta_j(x_j) = W_2 x_j$ are two embeddings of feature maps, and $W_1$ and $W_2$ are the learnable weight matrices. The function $g$ serves to compute a representation of the input signal at the position of $j$. According to Wang et al (2018), linear embedding is selected as $g$ here: $g(x_i) = W_g x_i$, where $W_g$ is a learned weighting matrix. $c(x)$ represents the normalization factor, which is defined as

$$c(x) = \sum_{j \in S} f(x_i, x_j).$$

To insert non-local operations into the neural network, a residual connection is adopted:

$$z_i = W_3 y_i + x_i,$$

where $x_i$ denotes input data. According to Wang et al (2018), at the deeper layer of the network, the feature maps’ spatial size is small and non-local module is insufficient to provide precise spatial information. If non-local module is used at the first layer of the network, expensive computational cost needs to be paid. In addition, using multiple non-local modules together can bring better results. Based on these considerations, we embed non-local modules after the second and third down-sampling layers, as depicted in figure 1. In NLN, $W_1, W_2, W_g$ and $W_3$ are obtained by $1 \times 1$ convolution. Figure 5 shows the non-local module. To focus on the artifact-impacted regions, $X_{sino}$ and $X_{pre}$ are concatenated as the inputs of $G_{im}$. A residual learning strategy is also adopted, and the output of $G_{im}$ denoted as $X_{im}$ is written as:

$$X_{im} = G_{im}(X_{sino}, X_{pre}).$$

The details of $G_{im}$ are shown in figure 6. $G_{im}$ is also optimized with L1 loss in the image domain:

$$L_{im} = \| (X_{im} - X_{gt}) \odot (1 - M) \|.$$  

In summary, the total objective function is:

$$\mathcal{L} = \alpha_1 \times L_{sino} + \alpha_2 \times L_{FBP} + \alpha_3 \times L_{im},$$

where $\alpha_1$, $\alpha_2$ and $\alpha_3$ are the weighting parameters of different components. In our experiments, we empirically set $\alpha_1 = \alpha_2 = \alpha_3 = 1$.

The symbols used in this paper are summarized in table 1.

### 3. Experiments

In this section, the data generation, details of neural networks, training strategies and experimental results will be shown in detail.

#### 3.1. Dataset

For data simulation, we followed the procedure of Pan (1999) and used the DeepLesion dataset [55], which has high diversity and good quality. For metal mask simulation, the shape, size and positions of masks should be delicately designed to cover real clinical scenes. In this work, we employed the masks generated from Lin et al (2019), containing 100 manually segmented metal implants with all kinds of metal implants, such as dental fillings, spine fixed crews, hip prostheses, coiling and wires. Specifically, we randomly selected 1000 CT images from the DeepLesion dataset and 90 metal masks to synthesize 90 000 combinations in the training set. The remaining 200 CT images and 10 masks were adopted for evaluation. The original CT images were resized to $256 \times 256$ for computational efficiency. To simulate Poisson noise, a polychromatic x-ray source was employed, and the incident beam x-ray was set to $2 \times 10^7$ photons [56]. The partial volume effects and scatter were also taken into consideration. Without loss of generality, our experiments were restricted to 2D parallel-beam geometry, i.e. the sinograms of CT images were obtained by the radon function with MATLAB R2017b. For the sampling condition, 367 detector bins and 361 sampling views uniformly distributed from 0° to 180° were assumed. Therefore, the sinogram had a size of $367 \times 361$. Unlike Pan (1999), we truncated the CT values to [0, 4095], which better conforms to the real situation.
Figure 7. Visual comparison using different methods on the simulated dataset with different metal sizes. (A1)–(A3): reference images; (B1)–(B3): metal corrupted images; (C1)–(C3): corresponding results of LI; (D1)–(D3): corresponding results of NMAR; (E1)–(E3): corresponding results of CNNMAR; (F1)–(F3): corresponding results of DuDoNet; (G1)–(G3): corresponding results of ADN; (H1)–(H3): corresponding results of DAN-Net. The display window is [−375, 560] HU.

Table 2. Quantitative comparison of different methods on the simulated dataset.

| Methods | Uncorrected | LI | NMAR | CNNMAR | DuDoNet | ADN | DAN-Net |
|---------|-------------|----|------|--------|---------|-----|---------|
| PSNR    | 15.33       | 30.74 | 30.83 | 32.15  | 36.82   | 33.60 | 40.61   |
| SSIM    | 0.6673      | 0.9224 | 0.9270 | 0.9508 | 0.9777  | 0.9275 | 0.9872  |
| RMSE (HU)| 139.88     | 52.57 | 49.53 | 35.04  | 25.42   | 59.29 | 17.83   |
3.2. Implementation details

We trained our network in an end-to-end manner, and the model was implemented with the PyTorch framework [57]. The back-projection was implemented by the numba library in Python, which can improve the computational efficiency, aided by CUDA. The network was optimized by the Adam optimizer with the parameters \( (\beta_1, \beta_2) = (0.5, 0.999) \). The learning rate was initialized to 0.0002 and halved every 20 epochs. The network was trained with 200 epochs on an NVIDIA 1080Ti GPU with 11 GB memory, and the batch size was 4.

3.3. Comparison with state-of-the-art methods

The proposed DAN-Net was compared with several state-of-the-art MAR methods: LI (Gjesteby et al. 2016), NMAR (Meyer et al. 2010), CNNMAR (Zhang and Yu 2018), DuDoNet (Lin et al. 2019) and ADN (Liao et al. 2019).

![Figure 8. Intermediate sinogram visual comparison of case 3 in figure 7 using different sinogram enhancement methods on the simulated dataset. (A) Reference sinogram; (B) metal corrupted sinogram; (C) corresponding results for LI; (D) corresponding results for NMAR; (E) corresponding results for CNNMAR; (F) corresponding results for DuDoNet; and (G) corresponding results for DAN-Net.](image)

![Table 3. Quantitative comparison of different methods over the ROIs indicated in figure 7. The best scores are highlighted in bold.](table)
LI and NMAR are classic methods widely used in MAR. CNNMAR is a well-known application of DL in MAR that comprehensively demonstrates the effectiveness and potential of CNN-based methods. DuDoNet is a supervised dual-domain framework in MAR that incorporates an extra sinogram enhancement network to ease the learning of the image domain. ADN is a state-of-the-art unsupervised framework in MAR that disentangles CT images corrupted by metal artifacts into an artifact-free domain and a pure artifact domain; and then decodes disentangled representations of artifact-free domains to artifact-suppressed images. For the LI4 and ADN5 methods, we used publicly released codes. Because there are no public implementations of the NMAR, CNNMAR and DuDoNet method, we reimplemented it following the original paper.

Structural similarity (SSIM), peak signal-to-noise ratio (PSNR) and root mean squared error (RMSE) are adopted as quantitative metrics. Table 2 lists the quantitative results obtained by calculating the mean values of both metrics on all of the test images using different methods. It is observed that the traditional MAR methods LI and NMAR significantly improve SSIM, PSNR value and degrade RMSE value compared with uncorrected CT images. NMAR outperforms LI since it takes advantage of both prior images and the LI method. CNNMAR fuses...
the merits of different MAR methods based on DL technology and outperforms conventional methods. ADN is an advanced unsupervised DL-based method that achieves similar performance to CNNMAR without the need for paired training data. DuDoNet and our method attain remarkable improvements on both SSIM and PSNR since they simultaneously leverage the advantages of the sinogram domain and image domain. Compared with DuDoNet, DAN-Net gets the better scores of each metric, which demonstrates the performance of our proposed method quantitatively.

For qualitative comparisons, the visual results are shown in figure 7, presenting three representative metallic implants with different sizes. In figure 7, metal-free images, metal-corrupted images and the results using different MAR methods are included. For better visualization, the simulated metal masks are colored in red. It can be seen that in the case of small metallic implants, the traditional methods, LI and NMAR, still contain some radial artifacts, while DL-based methods perform better. When metal objects get larger, LI and NMAR perform even worse. LI and NMAR introduce obvious new artifacts in figures 7(C1)–(C3) and (D1)–(D3). CNNMAR distorted structures and missing tissue details can be observed in figure 7(E3).

Another point that needs to be mentioned is that, in the third case, other methods fail to preserve the details around metallic implants, while DAN-Net maintains these structural details more completely. Figure 8 shows...
the corresponding intermediate sinogram enhancement results. Considering that ADN is an image postprocessing method, its sinogram enhancement is not presented. In the regions indicated by the blue arrows in figure 8(C), there are obvious artificial boundaries, whereas in the results of other methods, these boundaries are inconspicuous. In figures 8(D)–(G), as indicated by the green arrows, NMAR, CNNMAR and DuDoNet generate visible differences from the reference sinogram (figure 8(A)), but DAN-Net achieves the most visibly consistent sinogram with the reference. Two typical ROIs near and far away from the metals for each case, totally six ROIs which are indicated by the blue boxes (ROI (a)–(f)) in figures 7(B1)–(B3), are chosen to evaluate the quantitative performance in local regions. The results are listed in table 3. It is easy to notice that the proposed DAN-Net outperforms all the other methods in terms of all metrics except only three case. HU values along the yellow line in figures 7(B1)–(B3) of the ground truth versus the images reconstructed using different methods are plotted in figures 9–11, respectively and the proposed DAN-Net achieves the most consistent results to the ground truth.
3.4. Clinical study

To verify the performance of proposed DAN-Net in a clinical scenario, three clinical CT images were tested. In this experiment, the metal artifacts were empirically segmented using 2000 HU as the threshold. The test images were normalized to the same range as the training data. Figures 12–14 present the MAR results using different methods. It is observed that DAN-Net suppresses most of the metal artifacts and preserves the fine-grained anatomical structures around the metals, which supplies coherent results to the simulated data and demonstrates the potential for real clinical application. Meanwhile, the performance of most MAR methods is

Table 4. Quantitative comparison of different variants of our method on the simulated dataset.

| Methods      | Sino-Net | Res-Sino-Net | IM-Net | Non-local-IM-Net | Ma-Dual-Net | DAN-Net |
|--------------|----------|--------------|--------|------------------|-------------|---------|
| PSNR         | 31.43    | 31.71        | 33.79  | 34.75            | 34.15       | **40.61** |
| SSIM         | 0.9232   | 0.9494       | 0.9520 | 0.9720           | 0.9597      | **0.9872** |
| RMSE(HU)     | 47.60    | 39.44        | 30.14  | 24.01            | 31.77       | **17.83** |

Figure 15. Sinograms and corresponding reconstructions with sinogram-domain enhancement methods. The simulated metal masks are colored red. (A1) and (A2): ground truth. (B1) and (B2): Sino-Net. (C1) and (C2): Res-Sino-Net. The display window is [−375, 560] HU.

Figure 16. Reconstructions using image-domain and dual-domain enhancement methods. The simulated metal masks are colored in red. The reference image is figure 15 (A1). (A): IM-Net, (B): Non-local-IM-Net, (C): Ma-Dual-Net and (D): DAN-Net. The display window is [−375, 560] HU.
dependent on the previous results of segmentation, and our method will also benefit from a more accurate segmentation algorithm.

4. Ablation study

In this section, we investigate the effectiveness of different modules of the proposed DAN-Net. The ablation study configurations are listed as follows:

1. Sino-Net: the sinogram-domain network without residual learning.
2. Res-Sino-Net: the sinogram-domain network with residual learning.
3. IM-Net: the image-domain network without a non-local module.
4. Non-local-IM-Net: the image-domain network with the non-local module.
5. Ma-Dual-Net: a dual-domain network with sinogram-domain residual learning and image-domain non-local modules, but without an adaptively scaled sinogram; and
6. DAN-Net: full module taking $S_{res}$ and $X_{pre}$ as inputs.

The quantitative results of the ablation study are given in table 4 and the visual results are shown in figures 15 and 16.

4.1. Effect of sinogram-domain residual learning

To evaluate the performance of our residual sinogram learning strategy, Sino-Net and Res-Sino-Net were trained using $S_{ma}$ and $S_{res}$ as input respectively to complete the projection data within the metal trace. In table 4, it can be seen that the residual sinogram learning strategy improves the scores of all metrics. In figure 15(B1), evident dark artifacts appear. On the contrary, in figure 15(C1), residual learning recovers more details, which indicates that this residual strategy can ease network learning better. It can be seen that figure 15(B2) is obviously brighter than figures 15(A2) and (C2), which suggests possible data variation occurred in Sino-Net. In contrast, figures 15(A2) and (C2) look more consistent.

4.2. Effect of image-domain non-local module

To further suppress artifacts in the image domain, a non-local U-Net architecture is adopted. To validate the effectiveness of this modification, IM-Net and Non-local-IM-Net were trained without and with the non-local module in image domain, respectively. Both networks take the concatenation of $X_{ma}$ and $X_{pre}$ as the inputs. In table 4, the non-local-IM-Net has better quantitative scores than IM-Net. For the qualitative comparison, it is observed that artifacts are better suppressed in the results of Non-local-IM-Net than IM-Net in figures 16(A) and (B).

4.3. Effect of adaptive scaling

In this section, the impact of adaptive scaling is sensed. Ma-Dual-Net took $S_{ma}$ and $X_{ma}$ as inputs and DAN-Net took $S_{res}$ and $X_{pre}$ as inputs. In table 4, our approach outperforms Ma-Dual-Net in quantitative aspects. The visual comparison is also presented in figures 16(C) and (D); DAN-Net retrieves many more structural details around the metallic implants.

5. Discussions and conclusion

Due to the insertion of metals, the imaging quality of CT images will significantly degrade. Over the past few decades, many MAR methods have been proposed to alleviate the effects of metal artifacts in CT images. In conventional methods, projection data in the metal trace are regarded as missing and some interpolation methods are often applied to fill the missed projection data. Nonetheless, since most interpolation methods cannot guarantee continuity near the interpolation boundary, there are apparent borderlines in the corrected sinogram, and secondary artifacts can be introduced. Furthermore, since the projection data in the metal trace are simply abandoned and replaced with the value estimated by the data outside the metal trace, the information within the metal trace is lost, leading to the loss of tissue details around the metal in the reconstructed CT image. Therefore, not only secondary artifacts are introduced but also details are lost in interpolation-based methods. In practice, it is difficult for single-domain methods to address these problems (Lin et al. 2019). However,
interpolation-based methods can generate a proper initial estimation for DL-based methods. In our work, we also introduce this technique.

In this work, we combine the advantages of conventional MAR approaches and DL-based methods to further improve the performance. To restrain artifacts and maintain tissue details more efficiently, adaptive scaling on the original projection data in the metal trace is applied. Then, the adaptively scaled sinogram and corresponding reconstructed CT images are utilized as the inputs of our network. Because a metal has a much higher attenuation coefficient, the projection data inside and outside of the metal trace can be regarded as obeying two different data distributions. It is difficult to convert two different data distributions to a unified distribution for normal networks. To tackle this problem, a residual learning strategy that only modifies the metal trace region values of the scaled sinogram is used. To alleviate the new artifacts introduced in image domain enhancement, we propose a non-local U-Net architecture that can capture long-range dependencies to suppress metal artifacts further.

However, there are some limitations to our work, and we will dedicate ourselves to solving them in the future. In an end-to-end training manner, it is preferable to obtain the adaptive parameter by learning instead of through a manual setting. Fortunately, the subsequent filtering can reduce the influence of inaccurate parameter selection according to Chen et al (2002). In the future, we will investigate how to integrate this parameter learning into the model to minimize human interference.

We trained and evaluated our networks on simulated datasets, and few clinical CT images were used to validate the effectiveness of our model. In the future, we will collect large-scale clinical CT images to evaluate the performance of our method more comprehensively and systematically.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61871277, 61902264 and in part by the Sichuan Science and Technology Program under Grant 2021JDJQ0024, 2019YFS0125.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical statements

The public DeepLesion dataset used in this study was originally collected from The National Institutes of Health’s (NIH) Clinical Center under their IRB approval. The clinical dataset used in this paper were collected in West China Hospital of Sichuan University, and approved by the IRB.

ORCID iDs

Yi Zhang https://orcid.org/0000-0001-7201-2092

References

Ballhausen H, Reiner M, Ganswindt U, Belka C and Sohn M 2014 Post-processing sets of tilted CT volumes as a method for metal artifact reduction, Radiat. Oncol. 9 1–10
Buades A, Coll B and Morel J 2005 A non-local algorithm for image denoising Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (IEEE) pp 60–5
Chen H, Zhang Y, Zhang W, Liao P and Wang G 2017 Low-dose CT via convolutional neural network Biomed. Opt. Express 8 679–94
Chen L M, Liang Y, Sandison G A and Rydberg J 2002 Novel method for reducing high-attenuation object artifacts in CT reconstructions, in medical imaging 2002: Image Processing Int. Soc. Opt. Photonics 4684 841–50
Cuadros A, Ma X and Arce G R 2019 Compressive spectral x-ray tomography based on spatial and spectral coded illumination Opt. Express 27 10745–64
De Man B, Nyuits J, Dupont P, Marchal G and Suetens P 2000 Reduction of metal streak artifacts in x-ray computed tomography using a transmission maximum a posteriori algorithm, IEEE Trans. Nucl. Sci. 47 977–81
Duan X, Zhang L, Xiao Y, Cheng J, Chen Z and Xing Y 2008 Metal artifact reduction in CT images by sinogram TV inpainting 2008 IEEE Nuclear Science Symp. Conf. Record (IEEE) pp 4175–7
Ghani M U and Karl W C 2018 Deep learning based sinogram correction for metal artifact reduction Electron. Imaging 2018 471–2
Ghani M U and Karl W C 2019 Fast enhanced CT metal artifact reduction using data domain deep learning IEEE Trans. Comput. Imaging 6 181–93
Gjesteby L, De Man B, Jin Y, Paganetti H, Verbarg J, Giantoudi D and Wang G 2016 Metal artifact reduction in CT: where are we after four decades? IEEE Access 4 5826–49
Gjesteby L A, Shan H, Yang Q, Xi Y and Wang G 2019 A dual-stream deep convolutional network for reducing metal streak artifacts in CT images Phys. Med. Biol. 64 235003
Gu J, Zhang L, Chen Z, Xing Y and Huang Z 2006 A method based on interpolation for metal artifacts reduction in CT images J. X-Ray Sci. Technol. 14 11–9
Guo Y, Liu Y, Oerlemans A, Lao S, Wu S and Lew M S 2016 Deep learning for visual understanding: a review  Neurocomputing 187 27–48
Han C and Baek J 2019 Multi-pass approach to reduce cone-beam artifacts in a circular orbit cone-beam CT system Opt. Express 27 10108–26
Jeong K Y and Ra J B 2009 Metal artifact reduction based on sinogram correction in CT presented at the 2009 IEEE Nuclear Science Symp. Conf. Record (NSS/MIC) (IEEE) pp 3480–3
Kachelrieß M, Watzke O and Kalender W A 2001 Generalized multi-dimensional adaptive filtering for conventional and spiral single-slice, multi-slice, and cone-beam CT Med. Phys. 28 475–90
LeCun Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 436–44
Li M, Zhang J, Zhang T, Guan Y, Xu P and Sun M 2015 A prior-based metal artifact reduction algorithm for x-ray CT J. X-Ray Sci. Technol. 23 229–41
Liao H, Lin W, Huo Z, Vogelsang L, Sehnert W J, Zhou S K and Luo J 2019 Generative mask pyramid network for ct/cbct metal artifact reduction with joint projection-sinogram correction Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (Springer) pp 77–85
Liao H, Lin W, Zhou S K and Luo J 2019 ADN: artifact disentanglement network for unsupervised metal artifact reduction IEEE Trans. Med. Imaging 39 634–43
Lin W, Liao H, Peng C, Sun X, Zhang J, Luo J, Chellappa R and Zhou S K 2019 DadoNet: Dual domain network for ct metal artifact reduction Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (IEEE) pp 10512–21
Long J, Shelhamer E and Darrell T 2015 Fully convolutional networks for semantic segmentation IEEE Trans. Pattern Anal. Mach. Intell. 39 640–51
Lyu Y, Lin W, Liao H, Lu J and Zhou S K 2020 Encoding metal mask projection for ct metal artifact reduction in computed tomography Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (Springer) pp 147–57
Mehranian A, Ay M R, Rahimim A and Zaidi H 2013b 3D prior image constrained projection completion for x-ray CT metal artifact reduction with joint projection-sinogram correction IEEE Trans. Med. Imaging 32 1707–22
Mehranian A, Ay M R, Rahimim A and Zaidi H 2013b 3D prior image constrained projection completion for x-ray CT metal artifact reduction IEEE Trans. Med. Imaging 32 1707–22
Meyer E, Raupach R, Lell M, Schmidt B and Kachelrieß M 2010 Normalized metal artifact reduction IEEE Trans. Med. Imaging 37 5492–93
Mirotto R, Wang F, Wang S, Jiang X and Dudley J T 2018 Deep learning for healthcare: review, opportunities and challenges Brief Bioinform. 19 1236–46
Mouton A, Megherbi N, Van Slambrouck K, Nuyts J and Breckon T P 2013 An experimental survey of metal artifact reduction in computed tomography J. X-Ray Sci. Technol. 21 193–226
Pan X 1999 Optimal noise control in and fast reconstruction of fan-beam computed tomography image Med. Phys. 26 689–97
Park H S, Lee S M, Kim H P, Seo J K and Chung Y E 2018 CT sinogram-consistency learning for metal-induced beam hardening correction Med. Phys. 45 5376–84
Peng C, Li B, Wang H, Zhao Z, Qiu B and Chen D Z 2020 An irregular metal trace inpainting network for x-ray CT metal artifact reduction Med. Phys. 47 1087–100
Pimkin A, Samoylenko A, Antipina N, Ovechkina A, Golanov A, Dalechina A and Belyaev M 2020 Multidomain CT metal artifacts reduction Proc. of IEEE Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (Springer) pp 77–85
Wang G, Ye J C, Mueller K and Fessler J A 2018 Image reconstruction is a new frontier of machine learning IEEE Trans. Med. Imaging 37 1289–96
Wang J, Wang S, Chen Y, Wu J, Costatries J L and Luo L 2013 Metal artifact reduction in CT using fusion based prior image Med. Phys. 40 81903
Wang X, Girshick R, Gupta A and He K 2018 Non-local Neural Networks Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (IEEE) pp 7794–803
Watzke O and Kalender W 2004 A pragmatic approach to metal artifact reduction in CT: merging of metal artifact reduced images Eur. Radiol. 14 849–56
Xue H, Zhang L, Xiao Y, Chen Z and Xing Y 2009 Metal artifact reduction in dual energy CT by sinogram segmentation based on active contour model and TV inpainting 2009 IEEE Nuclear Science Symp. Conf. Record (NSS/MIC) (IEEE) pp 904–8
Yazdi M and Beaulieu L 2006 A novel approach for reducing metal artifacts due to metallic dental implants Nuclear Science Symp. Conf. Record (IEEE) pp 3260–3
Yongqiang H, Zexin L, Zhimin S, Maosong R and Jiliu Z 2019 Simultaneous denoising and super-resolution of optical coherence tomography images based on generative adversarial network Opt. Express 27 12289–307
Yu L, Zhang Z, Li X and Xing L 2020 Deep sinogram completion with image prior for metal artifact reduction in CT images IEEE Trans. Med. Imaging 40 228–38
Zhang H, Dong B and Liu B 2018 A reweighted joint spatial-rational domain ct image reconstruction model for metal artifact reduction Siam J. Imaging Sci. 11 707–33
Zhang Y, Yu T, Hu J, Liu Y and Zhou J 2011 A new CT metal artifacts reduction algorithm based on fractional-order sinogram inpainting J. X-Ray Sci. Technol. 19 373–84
Zhang Y and Yu H 2018 Convolutional neural network based metal artifact reduction in x-ray computed tomography *IEEE Trans. Med. Imaging* 37 1370–81

Zhao S, Bae K T, Whiting B and Wang G 2002 A wavelet method for metal artifact reduction with multiple metallic objects in the field of view *J. X-Ray Sci. Technol.* 10 67–76

Zhao X, Chen P, Wei J and Qu Z 2020 Spectral CT imaging method based on blind separation of polychromatic projections with Poisson prior *Opt. Express* 28 12780–94