Adaptive Convolutional Dictionary Network for CT Metal Artifact Reduction

Hong Wang[1,2], **Yuexiang Li**[1,†], **Deyu Meng**[2,3,†], **Yefeng Zheng**[1]
[1]Tencent Jarvis Lab, Shenzhen, China
[2]Xi'an Jiaotong University, Xi'an, China
[3]Peng Cheng Laboratory, Shenzhen, China
{hazelhwang, vicyxli, yefengzheng}@tencent.com, dymeng@mail.xjtu.edu.cn

Abstract

Inspired by the great success of deep neural networks, learning-based methods have gained promising performances for metal artifact reduction (MAR) in computed tomography (CT) images. However, most of the existing approaches put less emphasis on modelling and embedding the intrinsic prior knowledge underlying this specific MAR task into their network designs. Against this issue, we propose an adaptive convolutional dictionary network (ACDNet), which leverages both model-based and learning-based methods. Specifically, we explore the prior structures of metal artifacts, *e.g.*, non-local repetitive streaking patterns, and encode them as an explicit weighted convolutional dictionary model. Then, a simple-yet-effective algorithm is carefully designed to solve the model. By unfolding every iterative substep of the proposed algorithm into a network module, we explicitly embed the prior structure into a deep network, *i.e.,* a clear interpretability for the MAR task. Furthermore, our ACDNet can automatically learn the prior for artifact-free CT images via training data and adaptively adjust the representation kernels for each input CT image based on its content. Hence, our method inherits the clear interpretability of model-based methods and maintains the powerful representation ability of learning-based methods. Comprehensive experiments executed on synthetic and clinical datasets show the superiority of our ACDNet in terms of effectiveness and model generalization. Code and supplementary material are available at https://github.com/hongwang01/ACDNet.

1 Introduction

X-ray computed tomography (CT) has been broadly adopted for clinical diagnosis. Nevertheless, common metallic implants within patients, such as dental fillings and hip prosthesis, would adversely cause the missing of projection data during CT imaging, and thus lead to the obvious streaking artifacts and shadings in the reconstructed CT images. A robust model, which automatically reduces the unsatisfactory metal artifacts and improves the quality of CT images for subsequent clinical treatment, is worthwhile to develop [Wang *et al.*, 2021b; Wang *et al.*, 2021a; Lin *et al.*, 2019].

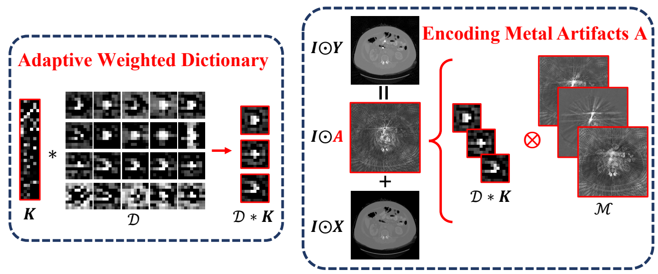


Figure 1: The proposed weighted convolutional dictionary model for encoding metal artifacts $\boldsymbol{A}$ as $(\mathcal{D} * \boldsymbol{K}) \otimes \mathcal{M}$. Here the dictionary $\mathcal{D}$ is sample-invariant and the weighting coefficient $\boldsymbol{K}$ is sample-variant. By updating $\boldsymbol{K}$, the representation kernel for artifacts (*i.e.*, $\mathcal{D} * \boldsymbol{K}$) can be adaptively inferred for every input image $\boldsymbol{Y}$.

Against this metal artifact reduction (MAR) task, traditional model-based methods focused on reconstructing artifact-reduced CT images by filling the metal-affected region in sinogram with different estimation strategies, such as linear interpolation (LI) [Kalender *et al.*, 1987] and normalized MAR [Meyer *et al.*, 2010]. However, these methods always bring secondary artifacts in the restored CT images, since the estimated sinogram cannot finely meet the CT physical imaging constraint. There is another type of conventional MAR methods, which iteratively reconstruct clean CT images from unaffected sinogram [Zhang *et al.*, 2011] or weighted/corrected sinogram [Lemmens *et al.*, 2008]. Although such model-based methods are usually algorithmically interpretable, the pre-designed regularizers cannot flexibly represent the complicated artifacts in different metal-corrupted CT images collected from real applications.

With the rapid development of deep learning (DL), recent years have witnessed the promising progress of DL for the MAR task. Some early works adopted convolutional neural network (CNN) to reconstruct clean sinograms and then utilized the filtered-back-projection transformation to reconstruct the artifact-reduced CT images [Zhang and Yu, 2018; Ghani and Karl, 2019]. Later, researchers adopted different learning strategies, *e.g.,* residual learning [Huang *et al.*, 2018] and adversarial learning [Wang *et al.*, 2018], to directly learn the artifact-reduced CT images. Very recently, there is a new

[†]Corresponding author

research line for the MAR task, which focuses on the mutual learning of sinograms and CT images [Lin *et al.*, 2019; Lyu *et al.*, 2020; Wang *et al.*, 2021b; Wang *et al.*, 2021c].

Attributed to the powerful learning ability of CNN, deep MAR methods generally achieve superior performance over conventional model-based approaches. Yet, there is still some room for further performance improvement. First, most of current deep MAR works pay less attention to exploring the intrinsic prior knowledge of the specific MAR task, such as non-local streaking structures of metal artifacts (see Fig. 1). The explicit embedding of such prior information can assist to regularize the solution space for metal artifact extraction, and further boost the MAR performance of deep networks as well as its generalization ability. Second, the involved image enhancement modules in current deep MAR works [Lin *et al.*, 2019; Lyu *et al.*, 2020] are roughly the variants of U-Net. With such a general network design, the physical interpretability of every network module is unclear.

In this paper, we propose an explicit model to encode the prior observations underlying the MAR task and fully embed it into an adaptive convolutional dictionary network (namely ACDNet). The proposed framework inherits the clear interpretability of model-based methods and maintains the powerful representation ability of learning-based methods. In summary, our main contributions can be concluded as:

1) Prior Formulation. For the MAR task, we explore that artifacts in different metal-corrupted CT images always present **common (*i.e.*, sample-invariant)** patterns, such as non-local repetitive streaking structures, but with the **specific (*i.e.*, sample-invariant)** intensity variation in CT images. Based on such prior observations, we adaptively encode the artifacts in every metal-corrupted CT image as a weighted convolutional dictionary (WCD) model (see Fig. 1).

2) Prior Embedding and Interpretability. To solve the WCD model, we propose an iterative algorithm with only simple operators and easily construct the ACDNet by unfolding every iterative step into the corresponding network module. Similar to learning-based methods, ACDNet automatically learns the priors for artifact-free CT images in a purely data-driven manner, overcoming the disadvantages of hand-crafted priors. Similar to model-based methods, ACDNet is explicitly integrated with the WCD model of artifacts and has clear physical interpretability corresponding to our optimization algorithm. The proposed ACDNet thus integrates the advantages of model-based and learning-based methodologies.

3) Fine Generalization. With the regularization of the explicit WCD model, ACDNet can more accurately extract artifacts complying with prior structures. Comprehensive experiments, including cross-body-site generalization and synthesis-to-clinical generalization, finely substantiate the superiority of our method as well as excellent computation efficiency. Especially, as an image-domain-based method without utilizing sinogram, our ACDNet can be adopted as a plug-in module, which is easily integrated into current computer-aided diagnosis systems to deal with some practical scenarios where the projection data is difficult to acquire.

2 Methodology

In this section, we propose an explicit model to adaptively encode the priors of every metal-affected CT image and derive a simple algorithm for the subsequent network design.

2.1 Weighted Convolutional Dictionary Model

For an observed metal-corrupted CT image $\boldsymbol{Y}$, it is composed of two regions, *i.e.*, metal part and non-metal part. Since metals generally have higher CT values than normal tissues, we ignore the information in the metal region and make efforts to reconstruct the non-metal region of $\boldsymbol{Y}$. Therefore, the decomposition model can be derived as:

$$\boldsymbol{I} \odot \boldsymbol{Y} = \boldsymbol{I} \odot \boldsymbol{X} + \boldsymbol{I} \odot \boldsymbol{A}, \tag{1}$$

where $\boldsymbol{I}$ is a binary non-metal mask; $\boldsymbol{X} \in \mathbb{R}^{H \times W}$ is the to-be-estimated clean CT image; $\boldsymbol{A}$ is the to-be-extracted metal artifacts; $\odot$ is an element-wise multiplication.

From Eq. (1), the estimation of $\boldsymbol{X}$ and $\boldsymbol{A}$ from $\boldsymbol{Y}$ is an ill-posed inverse problem. Against this issue, most of current learning-based methods empirically design complicated network architectures to directly learn $\boldsymbol{X}$ from $\boldsymbol{Y}$, which lack the consideration of **explicit prior knowledge** of MAR (*e.g.,* the patterns of metal artifacts). However, embedding the explicit prior is helpful to finely regularize the solution space of such an ill-posed problem, and thus further improves the performance of deep networks, especially generalization ability. Based on this motivation, we explore the specific prior structures underlying the MAR task, and then propose a strategy to explicitly embed them into deep networks.

Prior Formulation. Specifically, we disclose that for different metal-affected CT images, the metal artifacts are with roughly **common** patterns, *e.g.*, non-local repetitive streaking structures. Besides, due to the mutual influence between normal tissues and metal artifacts, the patterns of metal artifacts in different CT images are not exactly the same and generally have some **specific** characteristics, *e.g.*, pixel intensities of metal artifacts vary between different metal-corrupted CT images. Based on these prior observations, we formulate a weighted convolutional dictionary (WCD) model to encode the metal artifacts $\boldsymbol{A}$ as [Wang *et al.*, 2021d]:

$$\boldsymbol{A} = \sum_{n=1}^{N} (\mathcal{D} * \boldsymbol{K}_n) \otimes \boldsymbol{M}_n = (\mathcal{D} * \boldsymbol{K}) \otimes \mathcal{M}, \tag{2}$$

where $\mathcal{D} \in \mathbb{R}^{p \times p \times d}$ is a sample-invariant dictionary (*i.e.*, a set of convolutional filters) representing **common** local patterns of metal artifacts in different metal-corrupted CT images; $\boldsymbol{K}_n \in \mathbb{R}^{d}$ is a sample-wise weighting coefficient to learn the **specific** convolutional filter for $\boldsymbol{A}$ by computing $\mathcal{D} * \boldsymbol{K}_n$ as $\sum_{i=1}^{d} \mathcal{D}[:,:,i] \odot \boldsymbol{K}_n[i]$ (see Fig. 1); $\boldsymbol{M}_n \in \mathbb{R}^{H \times W}$ is the coefficients representing the locations for the local patterns of metal artifacts; p is the size of convolutional filters; d is the total number of filters in the dictionary $\mathcal{D}$; N is the actual number of filters encoding the artifacts $\boldsymbol{A}$; and the $\otimes$ between $(\mathcal{D} * \boldsymbol{K})$ and $\mathcal{M}$ is a convolutional operation in tensor form. Specifically, $(\mathcal{D} * \boldsymbol{K}_n) \in \mathbb{R}^{p \times p}$, $\boldsymbol{K} \in \mathbb{R}^{d \times N}$, $(\mathcal{D} * \boldsymbol{K}) \in \mathbb{R}^{p \times p \times N}$, $\mathcal{M} \in \mathbb{R}^{H \times W \times N}$, and

$$\begin{aligned} &\boldsymbol{K} = [\boldsymbol{K}_1, \boldsymbol{K}_2, \cdots, \boldsymbol{K}_N], \ \ \mathcal{M} = [\boldsymbol{M}_1, \boldsymbol{M}_2, \cdots, \boldsymbol{M}_N], \\ &\mathcal{D} * \boldsymbol{K} = [\mathcal{D} * \boldsymbol{K}_1, \mathcal{D} * \boldsymbol{K}_2, \cdots, \mathcal{D} * \boldsymbol{K}_N]. \end{aligned} \tag{3}$$

Note that convolutional dictionary model has been verified to be applicable to represent repetitive patterns by existing studies [Wang *et al.*, 2021a]. Compared to these methods, the proposed weighted mechanism in Eq. (2) has two main merits for the MAR task: 1) It delivers not only the sample-invariant/common knowledge, but also the sample-wise characteristics for information embedding, which can adaptively encode the prior structure for every CT image; 2) With such a weighted convolutional dictionary, we can choose an N smaller than d,* which further shrinks the solution space for estimating $\boldsymbol{A}$ and thereby improves the model generalization (see Sec. 4).

By substituting Eq. (2) into Eq. (1), we can derive:

$$\boldsymbol{I} \odot \boldsymbol{Y} = \boldsymbol{I} \odot \boldsymbol{X} + \boldsymbol{I} \odot ((\mathcal{D} * \boldsymbol{K}) \otimes \mathcal{M}) . \tag{4}$$

As shown, our goal is to estimate the sample-wise $\boldsymbol{K}$, $\mathcal{M}$, and $\boldsymbol{X}$ from $\boldsymbol{Y}$. Note that the non-metal mask $\boldsymbol{I}$ is pre-known (see Sec. 4) and the common dictionary $\mathcal{D}$ can be automatically learnt from training data (see Sec. 3). With the maximum-a-posterior framework, the optimization problem is:

$$\begin{aligned} \min_{\boldsymbol{K},\mathcal{M},\boldsymbol{X}} & \left\| \boldsymbol{I} \odot (\boldsymbol{Y} - \boldsymbol{X} - (\mathcal{D} * \boldsymbol{K}) \otimes \mathcal{M}) \right\|_F^2 \\ & + \alpha f_1(\boldsymbol{K}) + \beta f_2(\mathcal{M}) + \gamma f_3(\boldsymbol{X}) \\ \text{subject to} \quad & \|\boldsymbol{K}_n\|_2 = 1, n = 1, 2, \ldots, N, \end{aligned} \tag{5}$$

where $f_1(\cdot)$, $f_2(\cdot)$, and $f_3(\cdot)$ are regularization functions, delivering the prior knowledge of $\boldsymbol{K}$, $\mathcal{M}$, and $\boldsymbol{X}$, respectively; α, β, and γ are regularization weights.

2.2 Optimization Algorithm

Traditional solvers for the problem (5) often involve complex Fourier and inverse Fourier transforms, which are difficult to integrate into deep networks. In this regard, we design a new optimization algorithm with only simple operators. Concretely, the proximal gradient technique [Beck and Teboulle, 2009] is adopted to alternately update $\boldsymbol{K}$, $\mathcal{M}$, and $\boldsymbol{X}$:

Solving $\boldsymbol{K}$. From the problem (5), at the $(t+1)$-th iteration, $\boldsymbol{K}$ is solved as follows:

$$\begin{aligned} & \boldsymbol{K}^{(t+1)} = \operatorname*{argmin}_{\boldsymbol{K}} \left\| \boldsymbol{I} \odot \left(\boldsymbol{Y} - \boldsymbol{X}^{(t)} - (\mathcal{D} * \boldsymbol{K}) \otimes \mathcal{M}^{(t)} \right) \right\|_F^2 + \alpha f_1(\boldsymbol{K}), \\ & \text{subject to} \quad \|\boldsymbol{K}_n\|_2 = 1, n = 1, 2, \ldots, N, \end{aligned} \tag{6}$$

and the quadratic approximation [Beck and Teboulle, 2009] of the objective function in the problem (6) is:

$$\begin{aligned} \boldsymbol{K}^{(t+1)} = \operatorname*{argmin}_{\boldsymbol{K} \in \Omega} \quad & g_1\left(\boldsymbol{K}^{(t)}\right) + \frac{1}{2\eta_1} \left\| \boldsymbol{K} - \boldsymbol{K}^{(t)} \right\|_F^2 \\ & + \left\langle \boldsymbol{K} - \boldsymbol{K}^{(t)}, \nabla g_1\left(\boldsymbol{K}^{(t)}\right) \right\rangle + \alpha f_1(\boldsymbol{K}), \end{aligned} \tag{7}$$

where $\Omega = \{\boldsymbol{K} | \|\boldsymbol{K}_n\|_2 = 1, n = 1, 2, \cdots, N\}$; $g_1\left(\boldsymbol{K}^{(t)}\right) = \left\| \boldsymbol{I} \odot \left(\boldsymbol{Y} - \boldsymbol{X}^{(t)} - \left(\mathcal{D} * \boldsymbol{K}^{(t)}\right) \otimes \mathcal{M}^{(t)} \right) \right\|_F^2$; η_1 is the stepsize. Hence, we can get the following equation:

$$\begin{aligned} \frac{\partial g_1\left(\boldsymbol{K}^{(t)}\right)}{\partial \boldsymbol{K}_n} = & \left(U_3^f \left(\mathcal{D} \otimes^d \boldsymbol{M}_n^{(t)} \right) \right) \\ & \operatorname{vec}\left(\boldsymbol{I} \odot \left(\left(\mathcal{D} * \boldsymbol{K}^{(t)} \right) \otimes \mathcal{M}^{(t)} + \boldsymbol{X}^{(t)} - \boldsymbol{Y} \right) \right), \end{aligned} \tag{8}$$

where $\otimes^d$ is the depth-wise convolutional computation; $U_3^f(\cdot)$ denotes the unfolding operation at the 3^{rd} mode; and $\operatorname{vec}(\cdot)$ denotes the vectorization.

Clearly, Eq. (7) can be equivalently written as:

$$\boldsymbol{K}^{(t+1)} = \operatorname*{argmin}_{\boldsymbol{K} \in \Omega} \frac{1}{2} \left\| \boldsymbol{K} - \left(\boldsymbol{K}^{(t)} - \eta_1 \nabla g_1\left(\boldsymbol{K}^{(t)}\right) \right) \right\|_F^2 + \alpha \eta_1 f_1(\boldsymbol{K}) . \tag{9}$$

For general prior $f_1(\cdot)$ [Donoho, 1995], Eq. (9) is derived as:

$$\boldsymbol{K}^{(t+1)} = \operatorname{prox}_{\alpha\eta_1}\left(\boldsymbol{K}^{(t+0.5)} \right), \tag{10}$$

where $\boldsymbol{K}^{(t+0.5)} = \boldsymbol{K}^{(t)} - \eta_1 \nabla g_1\left(\boldsymbol{K}^{(t)}\right)$; $\nabla g_1\left(\boldsymbol{K}^{(t)}\right) = \left[\frac{\partial g_1\left(\boldsymbol{K}^{(t)}\right)}{\partial \boldsymbol{K}_1}, \frac{\partial g_1\left(\boldsymbol{K}^{(t)}\right)}{\partial \boldsymbol{K}_2}, \ldots, \frac{\partial g_1\left(\boldsymbol{K}^{(t)}\right)}{\partial \boldsymbol{K}_N} \right]$; $\operatorname{prox}_{\alpha\eta_1}(\cdot)$ is a proximal operator which is related to $f_1(\cdot)$. The constraint space Ω can be achieved by introducing a normalization operation into $\operatorname{prox}_{\alpha\eta_1}(\cdot)$ (see Sec. 3).

Solving $\mathcal{M}$. Similarly, at the $(t+1)$-th iteration, $\mathcal{M}$ can be updated by solving the quadratic approximation of the subproblem with respect to $\mathcal{M}$ as:

$$\begin{aligned} \mathcal{M}^{(t+1)} = \operatorname*{argmin}_{\mathcal{M}} & \frac{1}{2} \left\| \mathcal{M} - \left(\mathcal{M}^{(t)} - \eta_2 \nabla g_2\left(\mathcal{M}^{(t)}\right) \right) \right\|_F^2 \\ & + \beta \eta_2 f_2(\mathcal{M}), \end{aligned} \tag{11}$$

where η_2 is the stepsize and $g_2\left(\mathcal{M}^{(t)}\right) = \left\| \boldsymbol{I} \odot \left(\boldsymbol{Y} - \boldsymbol{X}^{(t)} - \left(\mathcal{D} * \boldsymbol{K}^{(t+1)} \right) \otimes \mathcal{M}^{(t)} \right) \right\|_F^2$. Similarly, the solution of problem (11) is deduced as:

$$\mathcal{M}^{(t+1)} = \operatorname{prox}_{\beta\eta_2}\left(\mathcal{M}^{(t+0.5)} \right), \tag{12}$$

where $\operatorname{prox}_{\beta\eta_2}(\cdot)$ is a proximal operator related to the regularization function $f_2(\cdot)$ about $\mathcal{M}$; $\mathcal{M}^{(t+0.5)} = \mathcal{M}^{(t)} - \eta_2 \nabla g_2\left(\mathcal{M}^{(t)}\right)$; $\nabla g_2\left(\mathcal{M}^{(t)}\right) = \left(\mathcal{D} * \boldsymbol{K}^{(t+1)} \right) \otimes^T \left(\boldsymbol{I} \odot \left(\left(\mathcal{D} * \boldsymbol{K}^{(t+1)} \right) \otimes \mathcal{M}^{(t)} + \boldsymbol{X}^{(t)} - \boldsymbol{Y} \right) \right)$; and $\otimes^T$ is the transposed convolution operation.

Solving $\boldsymbol{X}$. Given $\boldsymbol{K}^{(t+1)}$ and $\mathcal{M}^{(t+1)}$, the quadratic approximation of the subproblem about $\boldsymbol{X}$ is derived as:

$$\boldsymbol{X}^{(t+1)} = \operatorname*{argmin}_{\boldsymbol{X}} \frac{1}{2} \left\| \boldsymbol{X} - \left(\boldsymbol{X}^{(t)} - \eta_3 \nabla g_3\left(\boldsymbol{X}^{(t)}\right) \right) \right\|_F^2 + \gamma \eta_3 f_3(\boldsymbol{X}), \tag{13}$$

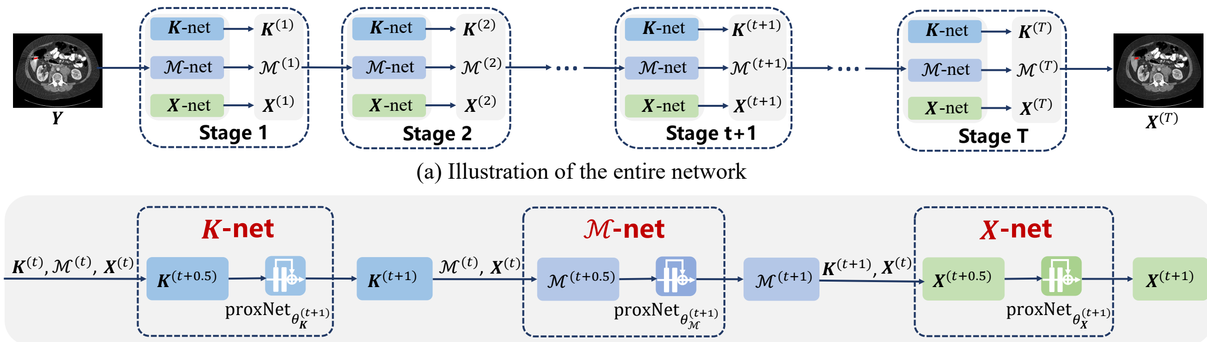
where $g_3\left(\boldsymbol{X}^{(t)}\right) = \left\| \boldsymbol{I} \odot \left(\boldsymbol{Y} - \boldsymbol{X}^{(t)} - \left(\mathcal{D} * \boldsymbol{K}^{(t+1)} \right) \otimes \mathcal{M}^{(t+1)} \right) \right\|_F^2$. Then, the updating rule of $\boldsymbol{X}$ is written as:

$$\boldsymbol{X}^{(t+1)} = \operatorname{prox}_{\gamma\eta_3}\left(\boldsymbol{X}^{(t+0.5)} \right), \tag{14}$$

where $\boldsymbol{X}^{(t+0.5)} = (\boldsymbol{1} - \eta_3 \boldsymbol{I}) \odot \boldsymbol{X}^{(t)} + \eta_3 \boldsymbol{I} \odot \left(\boldsymbol{Y} - \left(\mathcal{D} * \boldsymbol{K}^{(t+1)} \right) \otimes \mathcal{M}^{(t+1)} \right)$; $\operatorname{prox}_{\gamma\eta_3}(\cdot)$ is dependent on the prior function $f_3(\cdot)$ about $\boldsymbol{X}$.

Eqs. (10), (12), and (14) constitute the entire iterative process for solving the problem (5). As seen, the proposed algorithm only contains simple operators, which makes it easier to accordingly build the deep network framework upon the algorithm. Note that $\operatorname{prox}_{\alpha\eta_1}(\cdot)$, $\operatorname{prox}_{\beta\eta_2}(\cdot)$, and $\operatorname{prox}_{\gamma\eta_3}(\cdot)$ are implicit operators, which are automatically learnt from training data by virtue of the powerful prior fitting capability of

*In all our experiments below, $N = 6$ and $d = 32$.



(a) Illustration of the entire network

(b) Network structure at the $(t+1)$-th stage

Figure 2: (a) The proposed ACDNet consists of T stages. (b) The detailed structure at any stage where $\boldsymbol{K}^{(t+1)}$, $\mathcal{M}^{(t+1)}$, and $\boldsymbol{X}^{(t+1)}$ are successively updated by $\boldsymbol{K}$-net, $\mathcal{M}$-net, and $\boldsymbol{X}$-net, respectively, based on Eqs. (10), (12), and (14).

deep networks. This manner has been fully used and widely validated to be effective and helpful to explore interpretable knowledge by recent studies (*e.g.*. [Wang *et al.*, 2021d; Xie *et al.*, 2019]). The details are in Sec. 3.

3 Network Design and Analysis

Due to the specific design of joint model-driven (*i.e.*, prior knowledge) and data-driven (*i.e.*, deep networks) frameworks, deep unfolding techniques have achieved great success in computer vision tasks, such as single image rain removal [Wang *et al.*, 2021d] and low-light image enhancement [Liu *et al.*, 2022]. Inspired by this, we specifically design an adaptive convolutional dictionary network (ACDNet) for the MAR task by unfolding the iterative algorithm in Sec. 2.2 into the corresponding network structure.

Fig. 2 (a) displays the entire network structure with T stages, which correspond to T iterations of the optimization algorithm in Sec. 2.2. For every stage, our network consists of three sub-nets, *i.e.*, $\boldsymbol{K}$-net, $\mathcal{M}$-net, and $\boldsymbol{X}$-net corresponding to addressing the three subproblems, *i.e.*,, solving $\boldsymbol{K}$, $\mathcal{M}$, and $\boldsymbol{X}$, respectively. Fig. 2 (b) shows the detailed network connections at every stage, which are constructed by sequentially unfolding the iterative rules, *i.e.*, Eq. (10), Eq. (12), and Eq. (14), respectively. With the unfolding operations, every network module corresponds to the specific iterative step of the proposed optimization algorithm. Thus, the entire network framework has a clear physical interpretability. Next we present the information of sub-nets in detail:

$\boldsymbol{K}$-net. At the $(t+1)$-th stage, $\boldsymbol{K}^{(t+0.5)}$ is computed and fed to $\text{proxNet}_{\theta_{\boldsymbol{K}}^{(t+1)}}(\cdot)$ to execute the operator $\text{prox}_{\alpha\eta_1}(\cdot)$. Then, the updated weighting coefficient is: $\boldsymbol{K}^{(t+1)} = \text{proxNet}_{\theta_{\boldsymbol{K}}^{(t+1)}}\left(\boldsymbol{K}^{(t+0.5)}\right)$, where $\text{proxNet}_{\theta_{\boldsymbol{K}}^{(t+1)}}(\cdot)$ is a residual structure, *i.e.*, [*Linear+ReLU+Linear+Skip Connection+Normalization at the dimension* d].

$\mathcal{M}$-net. Given $\boldsymbol{K}^{(t+1)}$, $\mathcal{M}^{(t)}$, and $\boldsymbol{X}^{(t)}$, we can compute $\mathcal{M}^{(t+1)} = \text{proxNet}_{\theta_{\mathcal{M}}^{(t+1)}}\left(\mathcal{M}^{(t+0.5)}\right)$, where $\text{proxNet}_{\theta_{\mathcal{M}}^{(t+1)}}(\cdot)$ is the unfolding form of $\text{prox}_{\beta\eta_2}(\cdot)$—ResNet with three [*Conv+BN+ReLU+Conv+BN+Skip Connection*] residual blocks.

$\boldsymbol{X}$-net. Similarly, given $\boldsymbol{K}^{(t+1)}$, $\mathcal{M}^{(t+1)}$, and $\boldsymbol{X}^{(t)}$, the artifact-reduced CT image can be updated by $\boldsymbol{X}^{(t+1)} = \text{proxNet}_{\theta_{\boldsymbol{X}}^{(t+1)}}\left(\boldsymbol{X}^{(t+0.5)}\right)$, where $\text{proxNet}_{\theta_{\boldsymbol{X}}^{(t+1)}}(\cdot)$ has the same network structure to $\text{proxNet}_{\theta_{\mathcal{M}}^{(t+1)}}(\cdot)$.

With T-stage optimization, we can get the final CT image $\boldsymbol{X}^{(T)}$. All the involved parameters are $\{\theta_{\boldsymbol{K}}^{(t)}, \theta_{\mathcal{M}}^{(t)}, \theta_{\boldsymbol{X}}^{(t)}\}_{t=1}^{T}$, $\{\eta_i\}_{i=1}^{3}$, and the common dictionary $\mathcal{D}$. They can be automatically learnt from training samples in an end-to-end manner. Note that $\mathcal{D}$ is achieved by a common convolutional layer. More details, including the initialization ($\boldsymbol{K}^{(0)}$, $\mathcal{M}^{(0)}$, $\boldsymbol{X}^{(0)}$), are given in *Supplemental Material (SM)*.

ProxNet. Following recent great deep unfolding-based works [Wang *et al.*, 2021b; Xie *et al.*, 2019], we also set proximal operator with ResNet. Although it is very hard to inversely derive the form of regularizer by integrating ResNet function due to its complicated form, but as has comprehensively substantiated by previous research, it does be helpful to explore insightful structural prior. Besides, adopting such learning-based manner to automatically learn implicit regularizers is more flexible, avoiding hand-crafted prior design.

Interpretability. Unlike most of current deep MAR networks, which are heuristically built based on the off-the-shelf network blocks, ACDNet is naturally constructed under the guidance of the optimization algorithm with careful data fidelity term design. In this regard, the entire network integrates the interpretability of model-based methods. Besides, such interpretability is visually validated by Fig. 3 of *SM*.

Remark. Deep unfolding technique is a general tool for network design. To apply it, the key challenge is how to elaborately design model and optimization algorithm to make it work for specific applications. For the MAR task, we have specifically made some substantial ameliorations: 1) The adaptive prior of artifacts is incorporated into data fidelity term which can finely guide the network learning and boost the generalization ability; 2) Unlike other solvers with complicated computations (*e.g.*, Fourier transformation and matrix inversion), the proposed optimization algorithm is elaborately designed which only contains simple operators and makes the unfolding process easy and proper; 3) Compared to current popular dual-domain MAR methods, ACDNet only

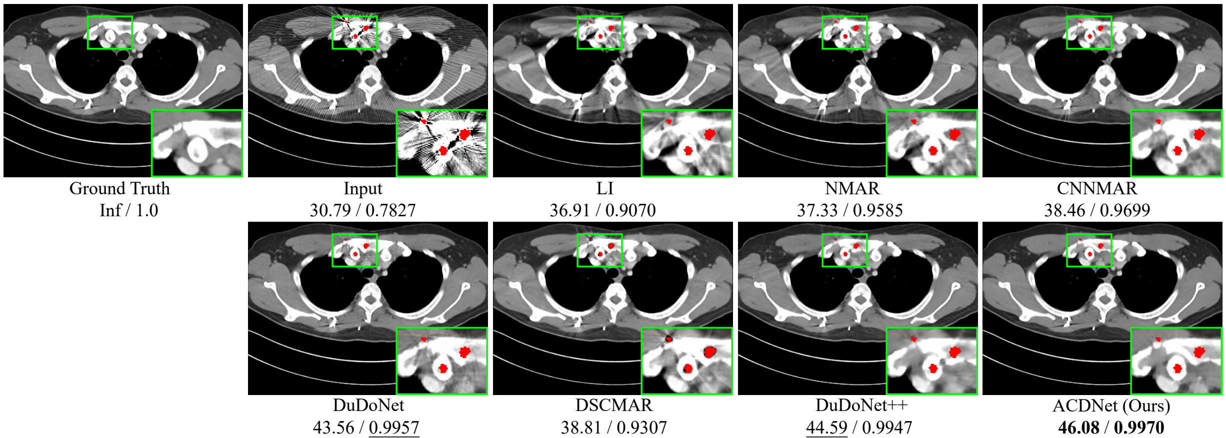


Figure 3: Performance comparison of different MAR approaches on a metal-corrupted CT image selected from the synthesized DeepLesion data. PSNR (dB)/SSIM below is for reference. The red pixels stand for metallic implants.

Methods	Large Metal		⟶		Small Metal	Average
Input	24.12/0.6761	26.13/0.7471	27.75/0.7659	28.53/0.7964	28.78/0.8076	27.06/0.7586
LI [Kalender *et al.*, 1987]	27.21/0.8920	28.31/0.9185	29.86/0.9464	30.40/0.9555	30.57/0.9608	29.27/0.9347
NMAR [Meyer *et al.*, 2010]	27.66/0.9114	28.81/0.9373	29.69/0.9465	30.44/0.9591	30.79/0.9669	29.48/0.9442
CNNMAR [Zhang and Yu, 2018]	28.92/0.9433	29.89/0.9588	30.84/0.9706	31.11/0.9743	31.14/0.9752	30.38/0.9644
DuDoNet [Lin *et al.*, 2019]	29.87/0.9723	30.60/0.9786	31.46/0.9839	31.85/0.9858	31.91/0.9862	31.14/0.9814
DSCMAR [Yu *et al.*, 2020]	34.04/0.9343	33.10/0.9362	33.37/0.9384	32.75/0.9393	32.77/0.9395	33.21/0.9375
DuDoNet++ [Lyu *et al.*, 2020]	36.17/0.9784	38.34/0.9891	40.32/0.9913	41.56/0.9919	42.08/0.9921	39.69/0.9886
ACDNet (Ours)	**37.91/0.9872**	**39.30/0.9920**	**41.14/0.9949**	**42.43/0.9961**	**42.64/0.9965**	**40.68/0.9933**

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

uses CT image domain data, which is more friendly to practical applications where the projection data is difficult to acquire. Besides, ACDNet can be easily integrated into current computer-aided diagnosis systems as a plug-in module.

Loss Function. With supervision on the extracted artifact $\boldsymbol{A}^{(t)}$ and CT image $\boldsymbol{X}^{(t)}$ at every stage, the training loss is:

$$\mathcal{L}=\sum_{t=0}^{T}\mu_t\boldsymbol{I}\odot\left\|\boldsymbol{X}-\boldsymbol{X}^{(t)}\right\|_F^2+\omega_1\left(\sum_{t=0}^{T}\mu_t\boldsymbol{I}\odot\left\|\boldsymbol{X}-\boldsymbol{X}^{(t)}\right\|_1\right) + \omega_2\left(\sum_{t=0}^{T}\mu_t\boldsymbol{I}\odot\left\|\boldsymbol{Y}-\boldsymbol{X}-\boldsymbol{A}^{(t)}\right\|_1\right), \quad (15)$$

where $\boldsymbol{X}$ is the clean (*i.e.*, ground truth) CT image. In all experiments, μ_T is set to 1; μ_t $(t=0,1,\cdots,T-1)$ is 0.1; ω_1 and ω_2 are empirically set to 5×10^{-4}.

Implementation Details. ACDNet is optimized through Adam optimizer based on PyTorch. The framework is trained on an NVIDIA Tesla V100-SMX2 GPU with a batch size of 32. The initial learning rate is 2×10^{-4} and divided by 2 at epochs [50, 100, 150, 200]. The total number of epochs is 300. The size of input image patch is 64×64 pixels and it is randomly flipped horizontally and vertically. More explanations are included in *SM*.

4 Experiments

In this section, we conduct extensive experiments to validate the effectiveness of our method.†

†More results, including interpretability verification, ablation study, convergence process, and P-values analysis are given in *SM*.

4.1 Dataset & Experimental Setting

Synthesized Data. Following the simulation procedure in [Yu *et al.*, 2020], we randomly choose 1,200 clean CT images from the public DeepLesion dataset [Yan *et al.*, 2018] and collect 100 metal masks from [Zhang and Yu, 2018] to synthesize the paired clean/metal-corrupted CT images. Specifically, 90 metals together with 1000 clean CT images for training and 10 ones together with the remaining 200 clean CT images for testing. Similar to [Lin *et al.*, 2019; Lyu *et al.*, 2020], we sequentially take every two testing metal masks as one group for performance evaluation. In addition, to evaluate the cross-body-site generalization performance, clean dental CT images [Yu *et al.*, 2020] are adopted and the corresponding metal-corrupted images are generated under the same simulation protocol on DeepLesion.

Clinical Data. A public clinical dataset, *i.e.*, CLINIC-metal [Liu *et al.*, 2021], which contains 14 metal-corrupted volumes with pixel-wise annotations of multiple bone structures (*i.e.,* sacrum, left hip, right hip, and lumbar spine), is used for evaluation. Following [Yu *et al.*, 2020], the clinical metal masks are segmented with thresholding of 2500 HU.

Baselines. Representative MAR methods are used, including traditional LI [Kalender *et al.*, 1987] and NMAR [Meyer *et al.*, 2010], learning-based CNNMAR [Zhang and Yu, 2018], DuDoNet [Lin *et al.*, 2019], DSCMAR [Yu *et al.*, 2020], and DuDoNet++ [Lyu *et al.*, 2020].

Evaluation Metric. We adopt the PSNR/SSIM for quantitative comparison on synthesized data and only visual comparison on clinical data due to the lack of clean CT images.

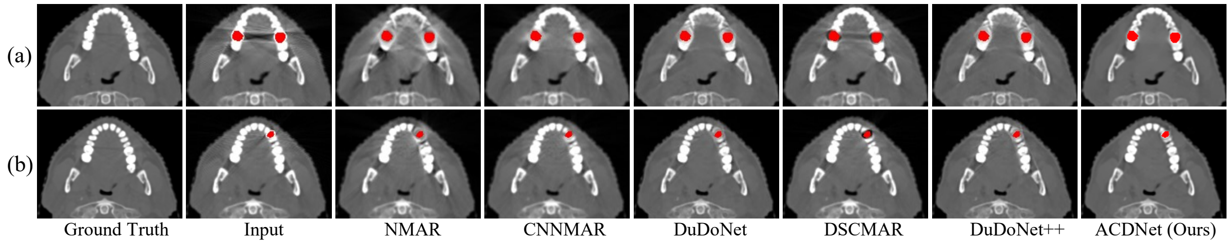


Figure 4: Generalization results. Artifact-reduced results on the synthesized dental CT images with different metallic implants.

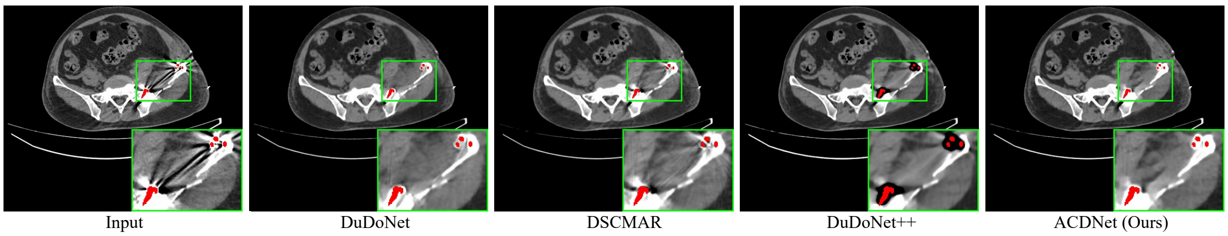


Figure 5: Generalization results. Performance comparison on a real clinical metal-affected CT image from CLINIC-metal.

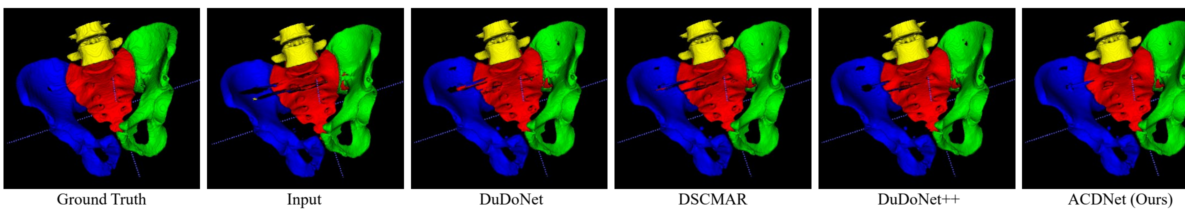


Figure 6: Downstream segmentation results of an artifact-reduced CLINIC-metal volume reconstructed by different MAR approaches.

4.2 Performance Evaluation

DeepLesion Data. Fig. 3 presents the visual comparison on an image from synthesized metal-corrupted DeepLesion. As shown, our approach significantly removes the artifacts and finely recovers evident details. The average PSNR/SSIM scores on the entire DeepLesion dataset are listed in Table 1. It can be observed that ACDNet consistently achieves the best scores with varying sizes of metallic implants.

Dental Data. Fig. 4 shows the MAR visual results on the synthesized dental CT images with different metallic implants, where deep MAR methods are trained on the synthesized DeepLesion (focusing on abdomen and thorax) and directly tested on the dental CT images. Due to the regularization with the explicit WCD model, ACDNet can more accurately identify the artifacts and accomplish the better reconstruction of artifact-reduced CT images. The results show the excellent cross-body-site generalizability of our method.

Clinical Data. Fig. 5 presents the MAR comparison of generalization results on a clinical metal-corrupted CT image. Our ACDNet finely preserves more bone structures and removes more artifacts. Fig. 6 displays the downstream segmentation results of an artifact-reduced CLINIC-metal volume, which are reconstructed by different MAR approaches. The proposed ACDNet is significantly superior to other approaches, which reveals its excellent potential for clinical applications. Due to the limited space, we only provide the results of the latest MAR methods.

Computation Efficiency. Table 2 reports that our method has fewer network parameters (*i.e.,* storage cost) and computation consumption, compared to other methods.

Methods	#Parameters	Testing Time
DuDoNet [Lin *et al.*, 2019]	25,834,251	0.4225
DSCMAR [Yu *et al.*, 2020]	25,834,251	0.3638
DuDoNet++ [Lyu *et al.*, 2020]	25,983,627	0.8062
ACDNet (Ours)	**1,602,809**	**0.3138**

Table 2: The number of network parameters and average testing time (seconds) computed on 2000 images with size 416×416 pixels on an NVIDIA Tesla V100-SMX2 GPU.

5 Conclusion and Future Work

In this paper, for the MAR task, we have proposed a weighted convolutional dictionary model to explicitly and adaptively encode the structural prior of metal artifacts for every input image. Then, by unfolding a simple-yet-effective optimization algorithm, we have easily built the deep network, which integrated the advantages of both model-based methods and learning-based methods. Experiments executed on three public datasets verified the excellent generalizability and excellent computational efficiency of our method.

Following current state-of-the-art (SOTA) MAR methods, we adopt the thresholding manner to simply segment the metal for clinical data. An unsatisfactory threshold possibly makes tissues be wrongly regarded as metals and most MAR methods would fail to recover image details. Although our method consistently achieve good performances for different sizes of metals and show the robustness, as a direction to further boost the performance of our framework, the joint optimization of automated metal localization and MAR is worthwhile to further explore. Besides, like SOTA MAR methods, we process CT images slice by slice for fairness. It would be very meaningful to do 3D prior modeling for the MAR task.

Acknowledgments

This work was founded by the China NSFC project under contract U21A6005, the Macao Science and Technology Development Fund under Grant 061/2020/A2, the major key project of PCL under contract PCL2021A12, the Key-Area Research and Development Program of Guangdong Province, China (No. 2018B010111001), National Key R&D Program of China (2018YFC2000702), the Scientific and Technical Innovation 2030-"New Generation Artificial Intelligence" Project (No. 2020AAA0104100).

References

[Beck and Teboulle, 2009] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[Donoho, 1995] David L Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, 1995.

[Ghani and Karl, 2019] Muhammad Usman Ghani and W Clem Karl. Fast enhanced CT metal artifact reduction using data domain deep learning. *IEEE Transactions on Computational Imaging*, 6:181–193, 2019.

[Huang *et al.*, 2018] Xia Huang, Jian Wang, Fan Tang, Tao Zhong, and Yu Zhang. Metal artifact reduction on cervical CT images by deep residual learning. *Biomedical Engineering Online*, 17(1):1–15, 2018.

[Kalender *et al.*, 1987] Willi A Kalender, Robert Hebel, and Johannes Ebersberger. Reduction of CT artifacts caused by metallic implants. *Radiology*, 164(2):576–577, 1987.

[Lemmens *et al.*, 2008] Catherine Lemmens, David Faul, and Johan Nuyts. Suppression of metal artifacts in CT using a reconstruction procedure that combines MAP and projection completion. *IEEE Transactions on Medical Imaging*, 28(2):250–260, 2008.

[Lin *et al.*, 2019] Wei-An Lin, Haofu Liao, Cheng Peng, Xiaohang Sun, Jingdan Zhang, Jiebo Luo, Rama Chellappa, and Shaohua Kevin Zhou. DuDoNet: Dual domain network for CT metal artifact reduction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10512–10521, 2019.

[Liu *et al.*, 2021] Pengbo Liu, Hu Han, Yuanqi Du, Heqin Zhu, Yinhao Li, Feng Gu, Honghu Xiao, Jun Li, Chunpeng Zhao, Li Xiao, et al. Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. *International Journal of Computer Assisted Radiology and Surgery*, 16(5):749–756, 2021.

[Liu *et al.*, 2022] Xinyi Liu, Qi Xie, Qian Zhao, Hong Wang, and Deyu Meng. Low-light image enhancement by retinex based algorithm unrolling and adjustment. *arXiv preprint arXiv:2202.05972*, 2022.

[Lyu *et al.*, 2020] Yuanyuan Lyu, Wei-An Lin, Haofu Liao, Jingjing Lu, and S Kevin Zhou. Encoding metal mask projection for metal artifact reduction in computed tomography. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 147–157, 2020.

[Meyer *et al.*, 2010] Esther Meyer, Rainer Raupach, Michael Lell, Bernhard Schmidt, and Marc Kachelrieß. Normalized metal artifact reduction (NMAR) in computed tomography. *Medical Physics*, 37(10):5482–5493, 2010.

[Wang *et al.*, 2018] Jianing Wang, Yiyuan Zhao, Jack H Noble, and Benoit M Dawant. Conditional generative adversarial networks for metal artifact reduction in CT images of the ear. In *International Conference on Medical Image Computing and Computer Assisted Intervention*, pages 3–11, 2018.

[Wang *et al.*, 2021a] Hong Wang, Yuexiang Li, Nanjun He, Kai Ma, Deyu Meng, and Yefeng Zheng. DICDNet: Deep interpretable convolutional dictionary network for metal artifact reduction in CT images. *IEEE Transactions on Medical Imaging*, 2021.

[Wang *et al.*, 2021b] Hong Wang, Yuexiang Li, Haimiao Zhang, Jiawei Chen, Kai Ma, Deyu Meng, and Yefeng Zheng. InDuDoNet: An interpretable dual domain network for CT metal artifact reduction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 107–118, 2021.

[Wang *et al.*, 2021c] Hong Wang, Yuexiang Li, Haimiao Zhang, Deyu Meng, and Yefeng Zheng. InDuDoNet+: A model-driven interpretable dual domain network for metal artifact reduction in CT images. *arXiv preprint arXiv:2112.12660*, 2021.

[Wang *et al.*, 2021d] Hong Wang, Qi Xie, Qian Zhao, Yong Liang, and Deyu Meng. RCDNet: An interpretable rain convolutional dictionary network for single image deraining. *arXiv preprint arXiv:2107.06808*, 2021.

[Xie *et al.*, 2019] Qi Xie, Minghao Zhou, Qian Zhao, Deyu Meng, Wangmeng Zuo, and Zongben Xu. Multispectral and hyperspectral image fusion by MS/HS fusion net. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1585–1594, 2019.

[Yan *et al.*, 2018] Ke Yan, Xiaosong Wang, Le Lu, Ling Zhang, Adam P Harrison, Mohammadhadi Bagheri, and Ronald M Summers. Deep lesion graphs in the wild: Relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9261–9270, 2018.

[Yu *et al.*, 2020] Lequan Yu, Zhicheng Zhang, Xiaomeng Li, and Lei Xing. Deep sinogram completion with image prior for metal artifact reduction in CT images. *IEEE Transactions on Medical Imaging*, 40(1):228–238, 2020.

[Zhang and Yu, 2018] Yanbo Zhang and Hengyong Yu. Convolutional neural network based metal artifact reduction in X-ray computed tomography. *IEEE Transactions on Medical Imaging*, 37(6):1370–1381, 2018.

[Zhang *et al.*, 2011] Xiaomeng Zhang, Jing Wang, and Lei Xing. Metal artifact reduction in X-ray computed tomography (CT) by constrained optimization. *Medical Physics*, 38(2):701–711, 2011.