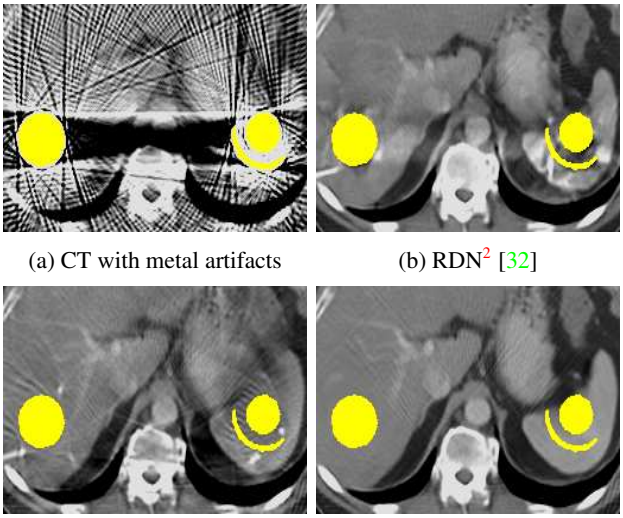
DuDoNet: Dual Domain Network for CT Metal Artifact Reduction

Wei-An Lin*[1] Haofu Liao*[2] Cheng Peng[1] Xiaohang Sun[3] Jingdan Zhang[4]
Jiebo Luo[2] Rama Chellappa[1] Shaohua Kevin Zhou[5,6]
[1]University of Maryland, College Park [2]University of Rochester [3]Princeton University
[4]Z2W Corporation [5]Chinese Academy of Sciences [6]Peng Cheng Laboratory, Shenzhen

Abstract

Computed tomography (CT) is an imaging modality widely used for medical diagnosis and treatment. CT images are often corrupted by undesirable artifacts when metallic implants are carried by patients, which creates the problem of metal artifact reduction (MAR). Existing methods for reducing the artifacts due to metallic implants are inadequate for two main reasons. First, metal artifacts are structured and non-local so that simple image domain enhancement approaches would not suffice. Second, the MAR approaches which attempt to reduce metal artifacts in the X-ray projection (sinogram) domain inevitably lead to severe secondary artifact due to sinogram inconsistency. To overcome these difficulties, we propose an end-to-end trainable Dual Domain Network (DuDoNet) to simultaneously restore sinogram consistency and enhance CT images. The linkage between the sinogram and image domains is a novel Radon inversion layer that allows the gradients to back-propagate from the image domain to the sinogram domain during training. Extensive experiments show that our method achieves significant improvements over other single domain MAR approaches. To the best of our knowledge, it is the first end-to-end dual-domain network for MAR.



(a) CT with metal artifacts (b) RDN[2] [32]

(c) CNNMAR [33] (d) DuDoNet (Ours)

Figure 1: (a) Sample MAR results for a CT image with intense metal artifact. Metal implants are colored in yellow. (b) Artifacts are not fully reduced and a 'white band' is present between the two implants. (c) Organ boundaries on the right are smeared out. (d) DuDoNet effectively reduces metal shadows and recovers fine details.

1. Introduction

Computed tomography (CT) images reconstructed from X-ray projections allow effective medical diagnosis and treatment. However, due to increasingly common metallic implants, CT images are often adversely affected by metal artifacts which not only exhibit undesirable visual effects but also increase the possibility of false diagnosis. This creates the problem of metal artifact reduction (MAR), for which existing solutions are inadequate.

Unlike typical image restoration tasks such as super-resolution [14, 32, 25, 35], compression artifact removal [31, 7], and denoising [3, 16, 15], metal artifacts are often *structured and non-local* (e.g. streaking and shadowing artifacts as in Figure 1a). Modeling such artifacts in image domain is extremely difficult. Therefore, before the emergence of deep learning, most existing works [12, 4, 18, 17] proposed to reduce metal artifact in the X-ray projection (sinogram) domain. The metal-corrupted regions are viewed as missing, and replaced by interpolated values. However, as the projections are taken from a single object under certain geometry, physical constraints should be satisfied by the enhanced sinogram. Otherwise, severe *secondary artifacts* can be introduced in the reconstructed CT images.

* First two authors contributed equally.

[2]The residual dense network (RDN) proposed in [32] without upscaling layers.

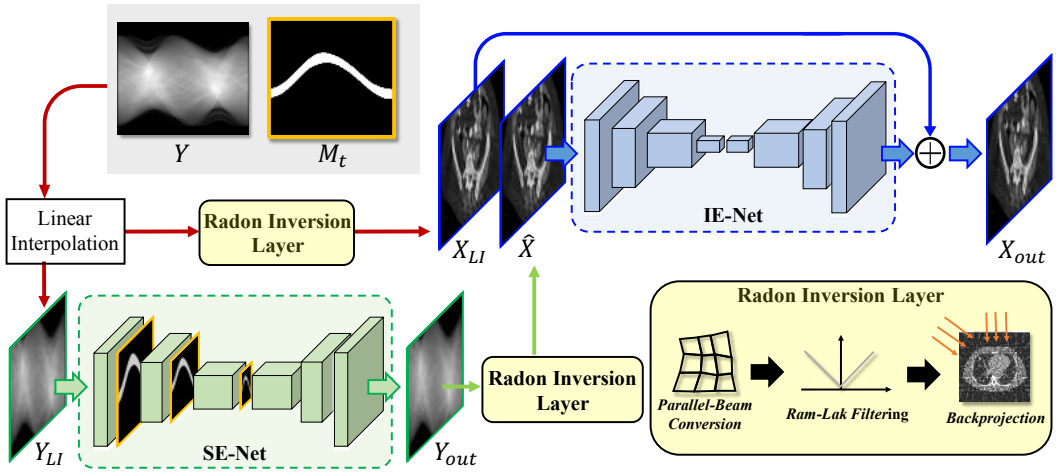


Figure 2: The proposed Dual Domain Network (DuDoNet) for MAR. Given a degraded sinogram Y and a metal trace mask $\mathcal{M}_t$, DuDoNet reduces metal artifacts by simultaneously refining in the sinogram and image domains.

Recently, motivated by the success of deep learning in solving ill-posed inverse problems [32, 25, 15, 19, 30, 23], several works have been proposed to overcome the difficulties in MAR. Wang et al. [24] applied the pix2pix model [9] to reduce metal artifact in the CT image domain. Zhang et al. [33] proposed to first estimate a prior image by a convolutional neural network (CNN). Based on the prior image, metal-corrupted regions in the sinogram are filled with surrogate data through several post-processing steps for reduced secondary artifact. Park et al. [20] applied U-Net [22] to directly restore metal-corrupted sinograms. Although metal artifacts can be reduced by these deep learning approaches, we will show that, despite the strong expressive power of deep neural networks, either image domain enhancement or sinogram domain enhancement is limited in being able to restore metal shadows and secondary artifact.

We hereby propose Dual Domain Network (DuDoNet) to address these problems by *learning two CNNs on dual domains to restore sinograms and CT images simultaneously*. Our intuition is that image domain enhancement can be improved by fusing information from the sinogram domain, and inconsistent sinograms can be corrected by the learning signal back-propagated from the image domain to reduce secondary artifacts. Specifically, we propose a novel network (Figure 2) consisting of three parts: *a sinogram enhancement network (SE-Net), a Radon inversion layer (RIL), and an image enhancement network (IE-Net)*. To address the issue that in the sinogram domain, information about small metal implants tends to vanish in higher layers of the network due to down-sampling, we propose a mask pyramid U-Net architecture for SE-Net, which retains metal mask information across multiple scales. The key to our dual-domain learning is RIL that reconstructs CT images using the filtered back-projection (FBP) algorithm and efficiently back-propagates gradients from the image domain to the sinogram domain. Based on RIL, we introduce a Radon consistency loss to penalize secondary artifacts in the image domain. Finally, IE-Net refines CT images via residual learning. Extensive experiments on CT images from hundreds of patients demonstrate that dual domain enhancement generates superior artifact-reduced CT images.

In summary, we make the following contributions:

- We propose an end-to-end trainable dual-domain refinement network for MAR. The network is able to recover details corrupted by metal artifacts.
- We propose a mask pyramid (MP) U-Net to improve sinogram refinement. The MP architecture improves performance especially when small metallic implants are dominated by the non-metal regions.
- We propose a Radon inversion layer (RIL) to enable efficient end-to-end dual domain learning. RIL can benefit the community through its ubiquitous use in various reconstruction algorithms [26, 10, 1, 34].
- We propose a Radon consistency (RC) loss to penalize secondary artifacts in the image domain. Gradients of the loss in the image domain are back-propagated through RIL to the sinogram domain for improved consistency.

2. Backgrounds and Related Works

Tissues inside the human body such as bones and muscles have different X-ray attenuation coefficients μ. If we consider a 2D slice of human body, the distribution of the attenuation coefficients $X = \mu(x, y)$ represents the underlying anatomical structure. The principle of CT imaging is based on the fundamental Fourier Slice Theorem, which guarantees that the 2D function X can be reconstructed

solely from its dense 1D projections. In CT imaging, projections of the anatomical structure X are inferred by the emitted and received X-ray intensities through the Lambert-Beer Law [2]. We consider the following CT model under a polychromatic X-ray source with energy distribution $\eta(E)$:

$$Y = -\log \int \eta(E) \exp\{-\mathcal{P}X(E)\}\, dE, \tag{1}$$

where $\mathcal{P}$ is the projection generation process, and Y represents the projection data (sinogram). The 2D $X(E)$ is the anatomical structure (CT image) we want to recover from the measured projection data Y .

For normal body tissues, $X(E)$ is almost constant with respect to the X-ray energy E. If we let $X = X(E)$, then

$$Y = \mathcal{P}X. \tag{2}$$

Therefore, given measured projection data Y, the CT image $\hat{X}$ can be inferred by using a reconstruction algorithm $\mathcal{P}^{\dagger}$[3]: $\hat{X} = \mathcal{P}^{\dagger}Y$ [11].

However, when metallic implants $I_M(E)$ are present, $X(E) = X + I_M(E)$, where $X(E)$ has large variations with respect to E due to I_M. Eq. (1) becomes

$$Y = \mathcal{P}X - \log \int \eta(E) \exp\{-\mathcal{P}I_M(E)\} dE, \tag{3}$$

where the region of $\mathcal{P}I_M$ in Y will be referred to as *metal trace* in the rest of the paper. When the reconstruction algorithm $\mathcal{P}^{\dagger}$ is applied,

$$\mathcal{P}^{\dagger}Y = \hat{X} - \mathcal{P}^{\dagger} \log \int \eta(E) \exp\{-\mathcal{P}I_M(E)\} dE. \tag{4}$$

The term after $\hat{X}$ in (4) is the metal artifact. It is clear that perfect MAR can be achieved only if the last term in Eq. (4) is suppressed while the term $\hat{X}$ is unaffected. However, it is generally an ill-posed problem since both terms contribute to the region of metal trace.

2.1. Inpainting-based Methods

One commonly adopted strategy in MAR is to formulate sinogram completion as an image inpainting task. Data within the metal trace are viewed as missing and filled through interpolation. Linear interpolation (LI) [12] is a widely used method in MAR due to its simplicity. Meyer et al. [18] proposed the NMAR algorithm, where sinograms are normalized by tissue priors before performing LI. NMAR requires proper tissue segmentation in the image domain, which is unreliable when severe metal artifacts are present. Mehranian et al. [17] restored sinograms by enforcing sparsity constraints in the wavelet domain. In general, inpainting-based approaches fail to replace the data of $\mathcal{P}X$ in (3) within metal trace by consistent values. *It is this introduced inconsistency in sinogram data that leads to noticeable secondary artifacts after reconstruction.*

[3]We use $\mathcal{P}^{\dagger}$ to denote the linear operation for reconstruction.

2.2. MAR by Iterative Reconstruction

In iterative reconstruction, MAR can be formulated as the following optimization problem:

$$\hat{X} = \min_{X} \|(1 - \mathcal{M}_t) \odot (\mathcal{P}X - Y)\|^2 + \lambda R(X), \tag{5}$$

where $\mathcal{M}_t$ is the metal trace mask. $\mathcal{M}_t = 1$ on the metal trace and $\mathcal{M}_t = 0$ otherwise. R is some regularization function, e.g. total variation (TV) [8] and sparsity constraints in the wavelet domain [29]. Eq. (5) is often solved through iterative approaches such as the split Bregman algorithm. Iterative reconstruction usually suffers from long processing time as they require multiplying and inverting huge matrices in each iteration. More importantly, handcrafted regularization $R(X)$ does not capture the structure of metal artifacts and would result in an over-smoothed reconstruction. Recently, Zhang et al. [29] proposed a re-weighted JSR method which combines NMAR into (5) and jointly solves for X and interpolated sinogram. Similar to NMAR, the weighting strategy in re-weighted JSR requires tissue segmentation. In phantom study, better performance against NMAR is achieved by re-weighted JSR. However, the improvements remain limited for non-phantom CT images.

2.3. Deep Learning based Methods for MAR

Convolutional neural networks have the ability to model complex structures within data. Motivated by the success of DNNs in solving inverse problems, Gjesteby et al. [6] and Park et al. [20] proposed to refine sinograms using a CNN for improved consistency. Zhang et al. [33] proposed a CNNMAR model to first estimate a prior image by a CNN and then correct sinogram similar to NMAR. However, even with the strong expressive power of CNNs, these approaches still suffer from secondary artifacts due to inconsistent sinograms.

Gjesteby et al. [5], Xu et al. [27] and Wang et al. [24] proposed to reduce metal artifact directly in the CT image domain. The metal artifacts considered in these works are mild and thus can be effectively reduced by a CNN. We will show in our experiments that image domain enhancement is not sufficient for mitigating intense metal shadows.

3. Proposed Method

As shown in Figure 2, our proposed model consists of three parts: (a) a sinogram enhancement network (SE-Net), (b) a Radon inversion layer (RIL), and (c) an image enhancement network (IE-Net). Inputs to the model include a degraded sinogram $Y \in \mathbb{R}^{H_s \times W_s}$ and the corresponding metal trace mask $\mathcal{M}_t \in \{0, 1\}^{H_s \times W_s}$. Notice that we use H_s to represent the detector size and W_s to represent the number of projection views. The region where $\mathcal{M}_t = 1$

is the metal trace. Given the inputs, we first apply LI [12] to generate an initial estimate for the sinogram data within metal trace. The resulting interpolated sinogram is denoted by Y_{LI}. SE-Net then restores Y_{LI} within the metal trace through a mask pyramid U-Net architecture. To maintain sinogram consistency, we introduce a Radon consistency (RC) loss. A sinogram will be penalized by the RC loss if it leads to secondary artifacts in the image domain after passing through RIL. Finally, the reconstructed CT image $\hat{X} \in \mathbb{R}^{H_c \times W_c}$ is refined by IE-Net via residual learning.

3.1. Sinogram Enhancement Network

Sinogram enhancement is extremely challenging since geometric consistency should be retained to prevent secondary artifact in the reconstructed CT image, so prior works only replace data within the metal trace. Similarly, given a metal-corrupted sinogram Y and metal trace mask $\mathcal{M}_t$, SE-Net $\mathcal{G}_s$ learns to restore the region of Y_{LI} in $\mathcal{M}_t = 1$. In sinogram domain enhancement, when the metal size is small, or equivalently, the metal trace is small, information about metal trace is dominated by non-metal regions in higher layers of network due to down-sampling. To retain the mask information, we propose to fuse $\mathcal{M}_t$ through a mask pyramid U-Net architecture. The output of SE-Net is written as

$$Y_{out} = \mathcal{M}_t \odot \mathcal{G}_s(Y_{LI}, \mathcal{M}_t) + (1 - \mathcal{M}_t) \odot Y_{LI}. \quad (6)$$

We use an L_1 loss to train SE-Net:

$$\mathcal{L}_{\mathcal{G}_s} = \|Y_{out} - Y_{gt}\|_1, \quad (7)$$

where Y_{gt} is the ground truth sinogram without metal artifact.

3.2. Radon Inversion Layer

Although sinogram inconsistency is reduced by SE-Net, there is no existing mechanism to penalize secondary artifacts in the image domain. The missing key element is an efficient and *differentiable* reconstruction layer. Therefore, we propose a novel RIL f_R to reconstruct CT images from sinograms and at the same time allow back-propagation of gradients. We hightlight that trivially inverting $\mathcal{P}$ in existing deep learning frameworks would require a time and space complexity of $\mathcal{O}(H_s W_s H_c W_c)$, which is prohibitive due to limited GPU memory.

In this work, we consider the projection process $\mathcal{P}$ as the Radon transform under fan-beam geometry with arc detectors [11]. The distance between an X-ray source and its rotation center is D. The resulting fan-beam sinograms Y_{fan} are represented in coordinates (γ, β). To reconstruct CT images from $Y_{fan}(\gamma, \beta)$, we adopt the fan-beam filtered back-projection (FBP) algorithm as the forward operation of RIL.

Our RIL consists of three modules: (a) a parallel-beam conversion module, (b) a filtering module, and (c) a back-projection module. The parallel-beam conversion module transforms $Y_{fan}(\gamma, \beta)$ to its parallel-beam counterpart $Y_{para}(t, \theta)$ through a change of variables. The FBP algorithm in coordinate (t, θ) becomes more effective and memory-efficient than in (γ, β). Parallel-beam FBP is then realized by the subsequent filtering and back-projection modules.

Parallel-beam Conversion Module. We utilize the property that a fan beam sinogram $Y_{fan}(\gamma, \beta)$ can be converted to its parallel beam counterpart $Y_{para}(t, \theta)$ through the following change of variables [11]:

$$\begin{cases} t = D \sin \gamma, \\ \theta = \beta + \gamma. \end{cases} \quad (8)$$

The change of variable in (8) is implemented by grid sampling in (t, θ), which allows back-propagation of gradients. With Y_{para}, CT images can be reconstructed through the following Ram-Lak filtering and back-projection modules.

Ram-Lak Filtering Module. We apply the Ram-Lak filtering to Y_{para} in the Fourier domain.

$$Q(t, \theta) = \mathcal{F}_t^{-1} \left\{ |\omega| \cdot \mathcal{F}_t \left\{ Y_{para}(t, \theta) \right\} \right\}, \quad (9)$$

where $\mathcal{F}_t$ and $\mathcal{F}_t^{-1}$ are the Discrete Fourier Transform (DFT) and inverse Discrete Fourier Transform (iDFT) with respect to the detector dimension.

Backprojection Module. The filtered parallel-beam sinogram Q is back-projected to the image domain for every projection angle θ by the following formula:

$$X(u, v) = \int_0^{\pi} Q(u \cos \theta + v \sin \theta, \theta) d\theta. \quad (10)$$

It is clear from (10) that the computation is highly parallel. We make a remark here regarding the property of RIL f_R. Due to the back-projection nature of f_R, the derivative with respect to the input Y_{out} is actually the projection operation $\mathcal{P}$. That is, any loss in the image domain will be aggregated and projected to the sinogram domain. This desirable property enables joint learning in sinogram and image domains.

Radon Consistency Loss. With the differentiable RIL, we introduce the following Radon consistency (RC) loss to penalize secondary artifacts in $\hat{X} = f_R(Y_{out})$ after reconstruction.

$$\mathcal{L}_{RC} = \|f_R(Y_{out}) - X_{gt}\|_1, \quad (11)$$

where X_{gt} is the ground truth CT image without metal artifact.

Difference from DL-based Reconstruction. Our RIL is designed to combine the image formation process (CT reconstruction) with deep neural networks and achieve improved MAR by dual-domain consistency learning. Methods in [26, 10, 1, 34] target *image formation via deep learning*, which is not the main focus of this work.

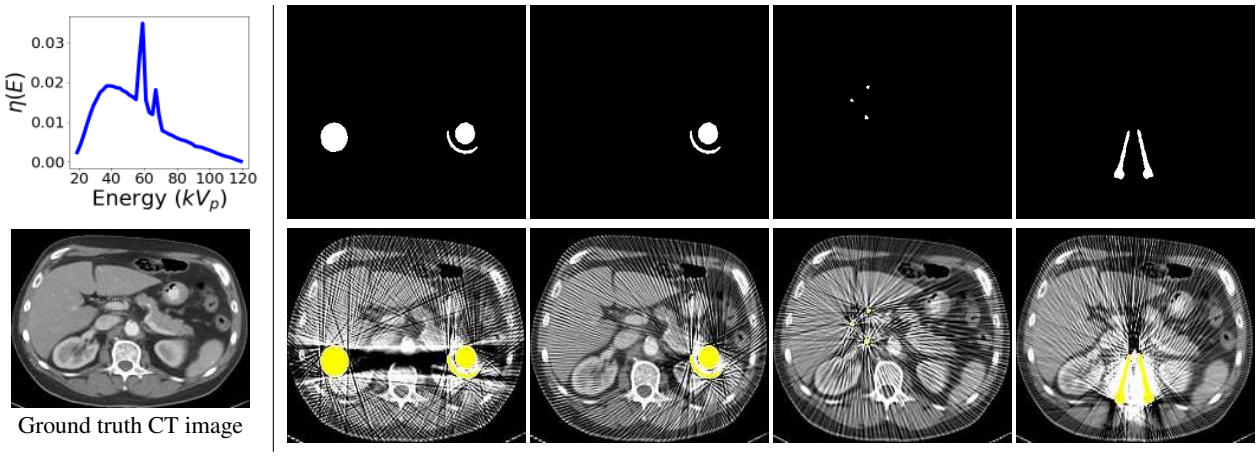


Figure 3: Sample simulated metal artifact on patient CT. The X-ray spectrum is shown in the upper-left corner. Metallic implants are colored in yellow for better visualization.

3.3. Image Enhancement Network

Since our ultimate goal is to reduce visually undesirable artifacts in image domain, we further apply a U-Net $\mathcal{G}_i$ to enhance $\hat{X}$ by residual learning:

$$X_{out} = X_{LI} + \mathcal{G}_i(\hat{X}, X_{LI}), \tag{12}$$

where $X_{LI} = f_R(Y_{LI})$ is reconstructed from Y_{LI}, the linearly interpolated sinogram. $\mathcal{G}_i$ is also optimized by L_1 loss.

$$\mathcal{L}_{\mathcal{G}_i} = \|X_{out} - X_{gt}\|_1. \tag{13}$$

The full objective function of our model is:

$$\mathcal{L} = \mathcal{L}_{\mathcal{G}_s} + \mathcal{L}_{RC} + \mathcal{L}_{\mathcal{G}_i}. \tag{14}$$

One could tune and balance each term in (14) for better performance. However, we found that the default setting works sufficiently well.

4. Experimental Results

Following the de facto practice in the literature [33], our evaluations consider simulated metal artifacts on real patient CTs. Various effects are considered including polychromatic X-ray, partial volume effect, and Poisson noise. The simulated artifacts exhibit complicated structures and cannot be easily modelled by a very deep CNN. All the compared approaches are evaluated on the same dataset, and superior performance is achieved by our method. Evaluations on clinical data is presented in the supplementary material.

Metal Artifact Dataset. Recently, Yan et al. [28] released a large-scale CT dataset DeepLesion for lesion detection. Due to its high diversity and quality, we use a subset of images from DeepLesion to synthesize metal artifact. 4,000 images from 320 patients are used in the training set and 200 images from 12 patients are used in the test set. All images are resized to 416×416. We collect a total of 100 metal shapes. 90 metal shapes are paired with the 4,000 images, yielding 360,000 combinations in the training set. 10 metal shapes are paired with the 200 images, yielding 2,000 combinations in the test set. In the training set, the sizes of the metal implants range from 16 to 4967 pixels. In the test set, the sizes of the metal implants range from 32 to 2054 pixels.

We adopt similar procedures as in [33] to synthesize metal-corrupted sinograms and CT images. We assume a polychromatic X-ray source with spectrum $\eta(E)$ in Figure 3. To simulate Poisson noise in the sinogram, we assume the incident X-ray has 2×10^7 photons. Metal partial volume effect is also considered. The distance from the X-ray source to the rotation center is set to 39.7cm, and 320 projection views are uniformly spaced between 0-360 degrees. The resulting sinograms have size 321×320. Figure 3 shows some sample images with simulated metal artifacts.

Evaluation Metrics. We choose peak signal-to-noise ratio (PSNR) and structured similarity index (SSIM) for quantitative evaluations. In DeepLesion, each CT image is provided with a dynamic range, within which the tissues are clearly discernible. We use the dynamic range as the peak signal strength when calculating PSNR.

Implementation Details. We implement our model using the PyTorch [21] framework. All the sinograms have size 321×320, and all the CT images have size 416×416. To train the model, we use the Adam [13] optimizer with $(\beta_1, \beta_2) = (0.5, 0.999)$, and a batch size of 8. The learning rate starts from 2×10^{-4}, and is halved for every 30 epochs. The model is trained on two Nvidia 1080Ti for 380 epochs.

PSNR(dB)/SSIM	Large Metal ⟶ Small Metal					Average
A) SE-Net$_0$	22.88/0.7850	24.52/0.8159	27.38/0.8438	28.61/0.8549	28.93/0.8581	26.46/0.8315
B) SE-Net	23.06/0.7868	24.71/0.8178	27.66/0.8463	28.91/0.8575	29.19/0.8604	26.71/0.8337
C) IE-Net	27.54/0.8840	29.49/0.9153	31.96/0.9368	34.38/0.9498	33.90/0.9489	31.45/0.9269
D) SE-Net$_0$+IE-Net	28.46/0.8938	30.67/0.9232	33.71/0.9458	36.17/0.9576	35.74/0.9571	32.95/0.9355
E) SE-Net+IE-Net	28.28/0.8921	30.49/0.9221	33.76/0.9456	36.26/0.9576	36.01/0.9574	32.96/0.9350
F) SE-Net$_0$+IE-Net+RCL	28.97/0.8970	31.14/0.9254	34.21/0.9476	36.58/0.9590	36.15/0.9586	33.41/0.9375
G) SE-Net+IE-Net+RCL	29.02/0.8972	31.12/0.9256	34.32/0.9481	36.72/0.9595	36.36/0.9592	33.51/0.9379

Table 1: Quantitative evaluations for different components in DuDoNet.

4.1. Ablation Study

In this section, we evaluate the effectiveness of different components in the proposed approach. Performance is evaluated on the artifact-reduced CT images. When evaluating SE-Nets without image domain refinement, we use the reconstructed CT images $\hat{X}$. We experiment on the following configurations:

A) SE-Net$_0$: The sinogram enhancement network without mask pyramid network.

B) SE-Net: The full sinogram enhancement module.

C) IE-Net: Image enhancement module. IE-Net is applied to enhance X_{LI} without $\hat{X}$.

D) SE-Net$_0$+IE-Net: Dual domain learning with SE-Net$_0$ and IE-Net.

E) SE-Net+IE-Net: Dual domain learning with SE-Net and IE-Net.

F) SE-Net$_0$+IE-Net+RCL: Dual domain learning with Radon consistency loss.

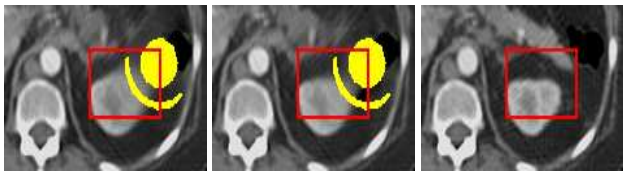
G) SE-Net+IE-Net+RCL: Our full network.

Notice that the configurations including SE-Net$_0$, SE-Net and IE-Net are single domain enhancement approaches.

Table 1 summarizes the performance of different models. Since there are totally 10 metal implants in the test set, for conciseness, we group the results according to the size of metal implants. The sizes of the 10 metal implants are: [2054, 879, 878, 448, 242, 115, 115, 111, 53, 32] in pixels. We simply put every two masks into one group.

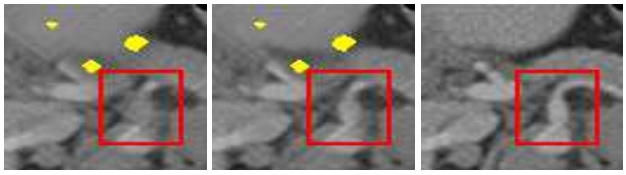
From E and G, it is clear that the use of the RC loss improves the performance over all metal sizes for at least 0.3 dB. In Figure 4, the model trained with RC loss better recovers the shape of the organ.

From F and G, we observe an interesting trend that the proposed mask pyramid architecture results in ∼0.2 dB gain when the metal size is small, and the performance is nearly identical when the metal is large. The reason is that the mask pyramid retains metal information across multiple scales. Figure 5 demonstrates that in the proximity of small metal implants, the model with mask pyramid recovers the fine details.



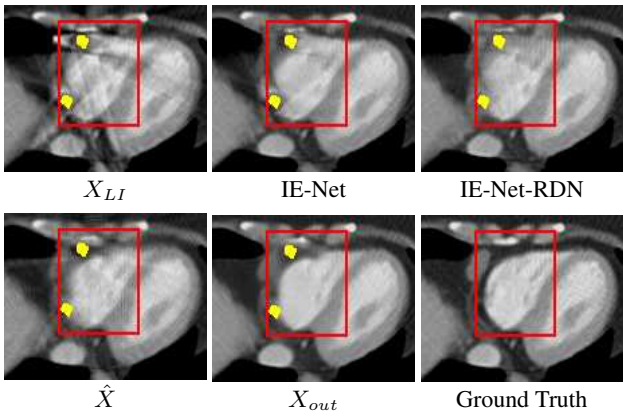
Without RC loss | With RC loss | Ground Truth

Figure 4: Visual comparisons between models without RC loss (E in Table 1) and our full model (G in Table 1).



Without MP | With MP | Ground Truth

Figure 5: Visual comparisons between models without MP (F in Table 1) and our full model (G in Table 1).



X_{LI} | IE-Net | IE-Net-RDN

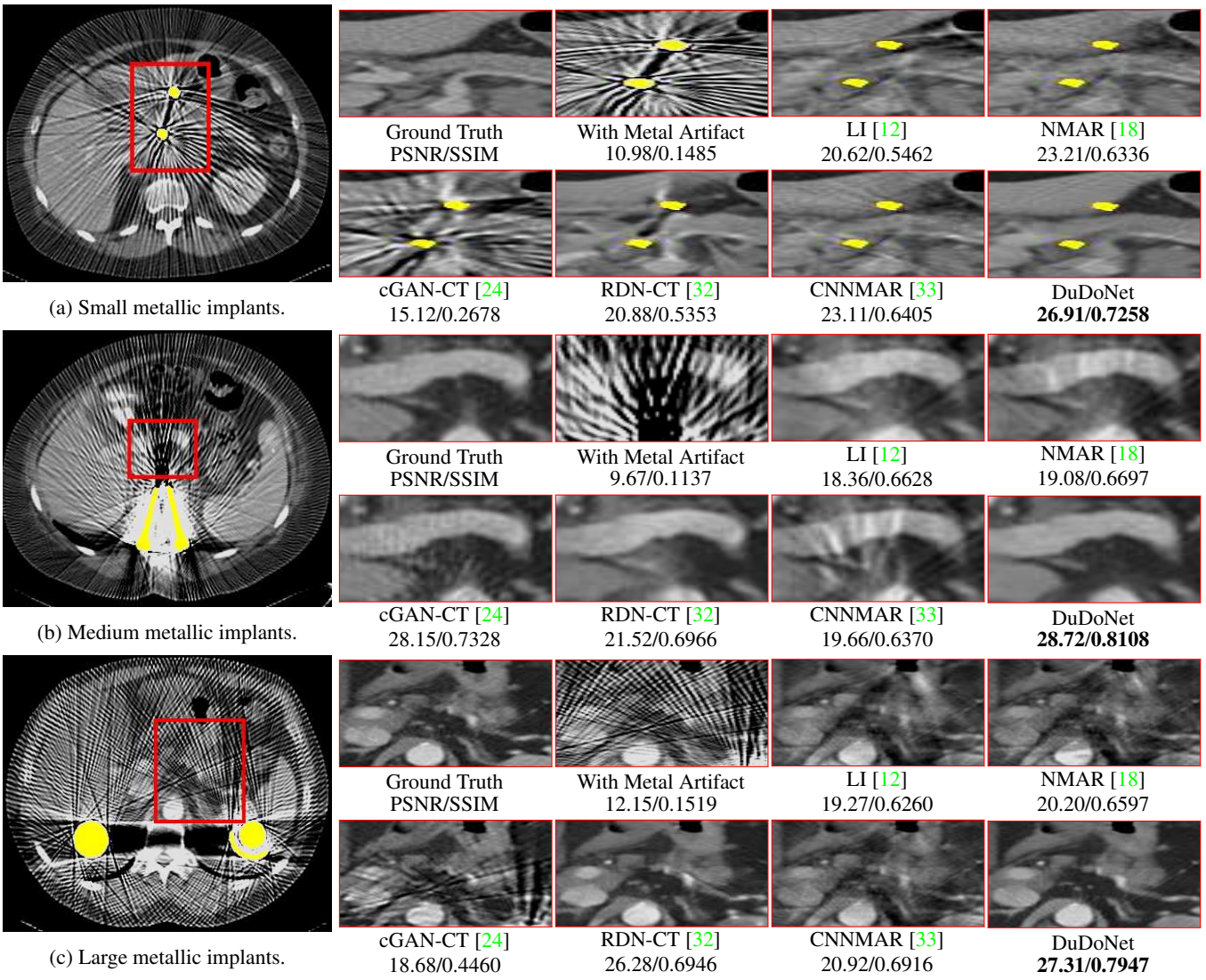
$\hat{X}$ | X_{out} | Ground Truth

Figure 6: Visual comparisons between models without SE-Net (top row IE-Net and IE-Net-RDN) and our full model (bottom row $\hat{X}$ and X_{out}).

Effect of Dual Domain Learning. In the proposed framework, IE-Net enhances X_{LI} by fusing information

PSNR(dB)/SSIM	Large Metal ⟶ Small Metal					Average
LI [12]	20.20/0.8236	22.35/0.8686	26.76/0.9098	28.50/0.9252	29.53/0.9312	25.47/0.8917
NMAR [18]	21.95/0.8333	24.43/0.8813	28.63/0.9174	30.84/0.9281	31.69/0.9402	27.51/0.9001
cGAN-CT [24]	26.71/0.8265	24.71/0.8507	29.80/0.8911	31.47/0.9104	27.65/0.8876	28.07/0.8733
RDN-CT [32]	28.61/0.8668	28.78/0.9027	32.40/0.9264	34.95/0.9446	34.00/0.9376	31.74/0.9156
CNNMAR [33]	23.82/0.8690	26.78/0.9097	30.92/0.9394	32.97/0.9513	33.11/0.9520	29.52/0.9243
DuDoNet (Ours)	**29.02/0.8972**	**31.12/0.9256**	**34.32/0.9481**	**36.72/0.9595**	**36.36/0.9592**	**33.51/0.9379**

Table 2: Quantitative evaluation of MAR approaches in terms of PSNR and SSIM.



(a) Small metallic implants.

(b) Medium metallic implants.

(c) Large metallic implants.

Figure 7: Visual comparisons on MAR for different types of metallic implants.

from SE-Net. We study the effect of dual domain learning by visually comparing our full pipeline (G in Table 1) with single domain enhancement IE-Net (C in Table 1). In addition to the U-Net architecture, we also consider IE-Net with RDN architecture, which is denoted as IE-Net-RDN. Visual comparisons are shown in Figure 6. We observe that single domain models IE-Net and IE-Net-RDN fail to recover corrupted organ boundaries in X_{LI}. In our dual domain refinement network, SE-Net first recovers inconsistent sinograms and reduces secondary artifacts as in $\hat{X}$. IE-Net then refines $\hat{X}$ to recover the fine details.

Effect of LI sinogram. The inputs to our network are

the linear interpolated sinogram Y_{LI} and its reconstructed CT X_{LI}. One possible alternative is to directly input the metal corrupted sinogram and CT, and let the network learn to restore the intense artifacts. However, we experimentally found out this alternative approach does not perform well. Metal shadows and streaking artifacts are not fully suppressed.

4.2. Comparison with State-of-the-Art Methods

In this section, we compare our model with the following methods: LI [12], NMAR [18], cGAN-CT [24], RDN-CT [32] and CNNMAR [33]. We use cGAN-CT to refer the approach by Wang et al. [24] which applies cGAN for image domain MAR. RDN [32] was originally proposed for image super-resolution (SR). The fundamental building unit of RDN is the residual dense block (RDB). Recently, it has been shown that by stacking multiple RDBs or its variant, the residual in residual dense blocks (RRDBs) [25], local details in natural images can be effectively recovered. We build a very deep architecture with 10 RDBs ($\sim$80 conv layers) for direct image domain enhancement, which is denoted by RDN-CT. Specifically, we select $D = 10, C = 8, G = 64$, following the notations in [32]. Inputs to RDN-CT are 128×128 patches.

Quantitative Comparisons. Table 2 shows quantitative comparisons. We observe that the state-of-the-art sinogram inpainting approach CNNMAR achieves higher SSIM than image enhancement approaches (e.g. RDN and cGAN-CT) especially when the size of metal is small. The reason is that sinogram inpainting only modifies data within the metal trace and recovers the statistics reasonably well. In most of the cases, CNNMAR also outperforms cGAN-CT in terms of PSNR. However, when CNN is sufficiently deep (e.g. RDN-CT), image enhancement approaches generally achieve higher PSNR. Our dual domain learning approach jointly restores sinograms and CT images, which attains the best performance in terms of both PSNR and SSIM *consistently in all categories*.

Visual Comparisons. Figure 7 shows visual comparisons. Figure 7a considers metal artifacts resulted from two small metallic implants. From the zoomed figure (with metal artifact), we can perceive severe streaking artifacts and intense metal shadows between the two implants. We observe that sinogram inpainting approaches such as LI, NMAR and CNNMAR effectively reduce metal shadows. However, fine details are either corrupted by secondary artifacts as in LI or blurred as in NMAR and CNNMAR. Image domain approaches such as cGAN-CT and RDN-CT produce sharper CT images but fail to suppress metal shadows. Our method effectively reduces metal shadows and at the same time retains fine details. Figure 7b shows a degraded CT image with long metal implants. We observe similar trend that sinogram inpainting approaches do not perform well in regions with intense streaking artifact. In this example, image domain methods reduce most of the artifacts. It is possibly due to that fact that the pattern of the artifact in Figure 7b is monotonous compared to Figures 7a and 7c. However, noticeable speckle noise is present in the result by cGAN-CT, and RDN-CT does not fully recover details in the middle. Figure 7c considers metal artifacts result from two large metallic implants. Likewise, sinogram inpainting methods and direct image domain enhancement have limited capability of suppressing metal artifacts. More visual comparisons are presented in the supplemental material.

4.3. Running Time Comparisons

On an Nvidia 1080Ti GPU, it takes 0.24 ms for RIL to reconstruct a sinogram of size 321×320 to a CT image of size 416×416, and 11.40 ms for back-propagation of gradients. RIL requires 16 MB of memory for forward pass and 25 MB for back-propagation. In Table 3 we compare the running time of different MAR approaches. With the running time of LI included, DuDoNet runs almost $4\times$ faster than the very deep architecture RDN while achieving superior performance.

LI [12]	NMAR [18]	cGAN-CT [24]	RDN-CT [32]	CNNMAR [33]	DuDoNet (Ours)
0.0832	0.4180	0.0365	0.5150	0.6043	0.1335

Table 3: Comparison of running time measured in seconds.

5. Conclusion

In this paper, we present the Dual Domain Network for metal artifact reduction. In particular, we propose to jointly improve sinogram consistency and refine CT images through a novel Radon inversion layer and a Radon consistency loss, along with a mask pyramid U-Net. Experimental evaluations demonstrate that while state-of-the-art MAR methods suffer from secondary artifacts and very-deep neural networks have limited capability of directly reducing metal artifacts in image domain, our dual-domain model can effectively suppress metal shadows and recover details for CT images. At the same time, our network is computationally more efficient. Future work includes investigating the potential of the dual-domain learning framework for other signal recovery tasks, such as super-resolution, noise reduction, and CT reconstruction from sparse X-ray projections.

Acknowledgements: This research was supported by MURI from the Army Research Office under the Grant No. W911NF-17-1-0304, NSF award #17228477, and the Morris K. Udall Center of Excellence in Parkinson's Disease Research by NIH.

References

[1] J. Adler and O. Öktem. Learned primal-dual reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, June 2018. 2, 4

[2] Beer. Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten. *Annalen der Physik und Chemie*, 162(5):78–88, 1852. 3

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007. 1

[4] X. Duan, L. Zhang, Y. Xiao, J. Cheng, Z. Chen, and Y. Xing. Metal artifact reduction in ct images by sinogram tv inpainting. In *Nuclear Science Symposium Conference Record, 2008. NSS'08. IEEE*, pages 4175–4177. IEEE, 2008. 1

[5] L. Gjesteby, Q. Yang, Y. Xi, B. Claus, Y. Jin, B. De Man, and G. Wang. Reducing metal streak artifacts in ct images via deep learning: Pilot results. In *The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, pages 611–614, 2017. 3

[6] L. Gjesteby, Q. Yang, Y. Xi, Y. Zhou, J. Zhang, and G. Wang. Deep learning methods to guide ct image reconstruction and reduce metal artifacts. In *SPIE Medical Imaging*, 2017. 3

[7] J. Guo and H. Chao. Building dual-domain representations for compression artifacts reduction. In *European Conference on Computer Vision*, pages 628–644. Springer, 2016. 1

[8] H. Gupta, K. H. Jin, H. Q. Nguyen, M. T. McCann, and M. Unser. Iterative metal artifact reduction for x-ray computed tomography using unmatched projector/backprojector pairs. *IEEE Transactions on Medical Imaging*, 43(6):3019–3033, 2016. 3

[9] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1125–1134, 2017. 2

[10] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, Sept 2017. 2, 4

[11] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. Society for Industrial and Applied Mathematics, 2001. 3, 4

[12] W. A. Kalender, R. Hebel, and J. Ebersberger. Reduction of ct artifacts caused by metallic implants. *Radiology*, 164(2):576–577, 1987. 1, 3, 4, 7, 8

[13] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 5

[14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, page 4, 2017. 1

[15] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila. Noise2Noise: Learning image restoration without clean data. In *International Conference on Machine Learning (ICML)*, volume 80, pages 2965–2974, 2018. 1, 2

[16] M. Makitalo and A. Foi. Optimal inversion of the anscombe transformation in low-count poisson image denoising. *IEEE transactions on Image Processing*, 20(1):99–109, 2011. 1

[17] A. Mehranian, M. R. Ay, A. Rahmim, and H. Zaidi. X-ray ct metal artifact reduction using wavelet domain l0 sparse regularization. *IEEE Transactions on Medical Imaging*, 32:1707–1722, 2013. 1, 3

[18] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß. Normalized metal artifact reduction (nmar) in computed tomography. *Medical physics*, 37(10):5482–5493, 2010. 1, 3, 7, 8

[19] J. Pan, W. Ren, Z. Hu, and M. Yang. Learning to deblur images with exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018. 2

[20] H. S. Park, Y. E. Chung, S. M. Lee, H. P. Kim, and J. K. Seo. Sinogram-consistency learning in ct for metal artifact reduction. *arXiv preprint arXiv:1708.00607*, 2017. 2, 3

[21] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. Automatic differentiation in pytorch. In *NIPS-W*, 2017. 5

[22] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 234–241. Springer, 2015. 2

[23] D. Ulyanov, A. Vedaldi, and V. Lempitsky. Deep image prior. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 2

[24] J. Wang, Y. Zhao, J. H. Noble, and B. M. Dawant. Conditional generative adversarial networks for metal artifact reduction in ct images of the ear. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, 2018. 2, 3, 7, 8

[25] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018. 1, 2, 8

[26] T. Würfl, F. C. Ghesu, V. Christlein, and A. Maier. Deep learning computed tomography. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 432–440. Springer International Publishing, 2016. 2, 4

[27] S. Xu and H. Dang. Deep residual learning enabled metal artifact reduction in ct. In *Medical Imaging 2018: Physics of Medical Imaging*, volume 10573, page 105733O. International Society for Optics and Photonics, 2018. 3

[28] K. Yan, X. Wang, L. Lu, L. Zhang, A. P. Harrison, M. Bagheri, and R. M. Summers. Deep lesion graphs in the wild: Relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018. 5

[29] H. Zhang, B. Dong, and B. Liu. A reweighted joint spatial-radon domain ct image reconstruction model for metal artifact reduction. *SIAM J. Imaging Sciences*, 11:707–733, 2018. 3

[30] H. Zhang and V. M. Patel. Densely connected pyramid dehazing network. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[31] X. Zhang, W. Yang, Y. Hu, and J. Liu. Dmcnn: Dual-domain multi-scale convolutional neural network for compression artifacts removal. In *IEEE International Conference on Image Processing (ICIP)*, 2018. 1

[32] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1, 2, 7, 8

[33] Y. Zhang and H. Yu. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE Transactions on Medical Imaging*, 2018. 1, 2, 3, 5, 7, 8

[34] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao. A sparse-view ct reconstruction method based on combination of densenet and deconvolution. *IEEE Transactions on Medical Imaging*, 37(6):1407–1417, June 2018. 2, 4

[35] Z. Zhong, T. Shen, Y. Yang, Z. Lin, and C. Zhang. Joint sub-bands learning with clique structures for wavelet domain super-resolution. *arXiv preprint arXiv:1809.04508*, 2018. 1