Dual Domain Diffusion Guidance for 3D CBCT Metal Artifact Reduction

Yongjin Choi, Doeyoung Kwon, Seung Jun Baek*
Korea University, Seoul, South Korea
{dydwls8445,doeyoung,sjbaek}@korea.ac.kr

Abstract

Previous methods to solve the problem of metal artifact reduction (MAR) have mostly focused on 2D MAR, making it challenging to apply to problems with 3-dimensional CT such as CBCT. In this paper, we propose a novel approach for 3D MAR which utilizes two diffusion models to model the metal-free CBCT prior and metal artifact prior. Through dual-domain guidance in the image and projection domains, the 3D connectivity is enhanced in the restored images. Moreover, we propose a memory-efficient technique for an efficient sampling of 3-dimensional data, which reduces the memory usage by orders of magnitude. Experiments show that our method achieves the state-of-the-art performance not only with synthetic data but also with real-world clinical and out-of-distribution data.

1. Introduction

Cone Beam Computed Tomography (CBCT) is widely utilized in dental diagnosis and treatment procedures, including dental implant placement, orthodontic appliances, and orthognathic surgery. However, the presence of metallic inserts, such as dental crowns, implants, and orthodontic devices, can result in non-local streaking and shading artifacts during CBCT imaging. These artifacts can obscure dental and oral structures, complicating the three-dimensional volumetric reconstruction and hindering accurate diagnosis [7, 21]. In particular, the intractable nature of artifact structures complicates mathematical modeling, thereby making their removal a challenging process.

Metal Artifact Reduction (MAR) is a method that involves removing artifacts caused by the scattering of metals in CT images. Numerous hand-crafted model-based approaches have been proposed to address the MAR problem [1, 15, 20]. With the recent advancements in deep learning, there has been a significant increase in attempts to solve the MAR problem using Convolutional Neural Networks (CNN) in the image domain [13, 19, 27], sinogram domain [9, 18, 31], and dual domain [28–30]. Furthermore, there have also been works based on generative models for addressing the MAR problem [17, 18]. Despite the numerous efforts, most of the existing works have primarily addressed the 2D MAR problem. 2D-based MAR methods fail to consider the three-dimensional geometry, which may lead to difficulties in distinguishing CT slices containing metal artifacts from those without metal and potentially leading to unnatural connections among the CT slices. Considering that CBCT produces 3D volumetric data, the applicability of 2D-based approaches in a clinical setting remains limited.

In a fidelity-embedded learning (FEL) Park *et al.* [22] implemented MAR for 3D CBCT by utilizing UNet [24] and wavelet decomposition [10] to remove metal artifacts in each 2D slice individually, followed by updating only the metal-free regions in the 3D cone beam projection based on a simple thresholding and the separable paraboloid surrogate method in an iterative process. While this method represents a unique approach to performing MAR in 3D CBCT, it may not be able to effectively restore images with severe or novel types of artifacts and generalize well to out-of-distribution data due to its inability to sufficiently learn prior related to CBCT and metal artifacts.

In this paper, we proposed to use two distinct diffusion models to separately model the metal-free CBCT prior and the metal artifact prior. By incorporating dual domain diffusion guidance in the image domain and the 3D cone beam projection domain, we removed severe or novel types of metal artifacts, while maintaining a natural continuity along the z-axis. Moreover, we propose a memory-efficient sampling method, which reduces the memory usage by 500× compared to the previous diffusion-based models to solve a similar type of inverse problems to our work. Experiments demonstrate that the proposed model achieves clinically significant performance in the 3D CT volume MAR task. Below we summarize the contributions of our work.

- By using dual domain diffusion guidance in the image domain and 3D cone beam projection domain, we tackled the 3D MAR problem efficiently based on 2D-diffusion models.

* Corresponding Author.

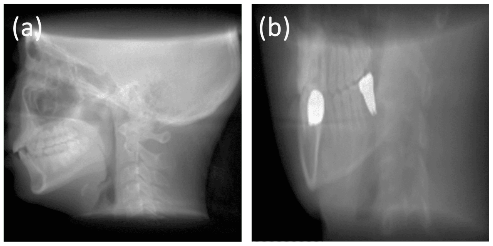

Figure 1. 3D cone beam projection of (a) metal-free CBCT and (b) metal-corrupted CBCT. The bright regions in (b) represent metal inserts.

- We propose a diffusion guidance method which is 500× memory efficient compared to the existing baseline methods using diffusion for 3D tasks.
- Our method not only is effective for synthetic samples, but also generalizes well to a wide range of samples including out-of-distribution and clinical samples.

2. Background

2.1. Cone beam projection & 3D Metal artifacts

The cone beam projection of $X \in \mathbb{R}^{d\times w\times h}$ at detector position p given polychromatic energy E, and energy spectrum $S(E)$ can be calculated as follows [22], see Fig. 1.

$$p = -\gamma \log \int_E S(E) \exp(-A(X_E))dx \tag{1}$$

X_E is the attenuation coefficient of X in energy E, A is a cone beam forward operator, and γ is a projection truncation by detector. The projection p can be reconstructed into a three-dimensional CT volume X using the FDK algorithm [8].

When high attenuation materials, such as metals, are present in CT images, non-local streaking and shadowing artifacts can occur. Furthermore, since CBCT is based on three-dimensional projections rather than individual CT slices, these artifacts exhibit a three-dimensional correlation. In such cases, the metal-corrupted CT can be considered as a combination of metal-free CT and its residual, the metal artifacts.

$$X = X^* + X^m \tag{2}$$

X is a metal-corrupted CT, X^* is a metal-free CT, and X^m is the metal artifacts.

2.2. Score-based diffusion models

Diffusion model [12,25] is a generative model that learns data distribution through a forward process, which progressively corrupts original data by adding Gaussian noise, and a reverse process that utilizes a neural network to learn the inverse of the forward process. Among these, the score-based diffusion model proposed by [25] defines the forward process as a Stochastic Diffusion Equation (SDE). In the diffusion process, data x is modeled as $x(t)$, where $t \in [0, 1]$ is a continuous time index. $x(0) \sim p_0$ represents the distribution of the dataset, while $x(T) \sim p_T$ denotes the prior distribution.

$$dx = f(x,t)dt + g(t)dw \tag{3}$$

where w denotes the standard Wiener process (also known as Brownian motion), $f(,t) : \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^d$ represents the diffusion coefficient of $x(t)$, and $g(t) : \mathbb{R} \to \mathbb{R}$ refers to the drift coefficient of $x(t)$. The reverse-time SDE corresponding to Eq. (3) can be expressed as follows.

$$dx = [f(x,t) - g(t)^2 \nabla_{x_t} \log p(x_t)]dt + g(t)d\bar{w} \tag{4}$$

$\bar{w}$ is the standard Wiener process as t moves from $T \to 0$, and dt represents the infinitesimal negative timestep. $\log p(x_t)$ signifies the score function, and if the score can be calculated for each t, p_0 can be sampled through the reverse process. $\log p(x_t)$ can be approximated by s_θ using the DSM objective [26].

$$\min_\theta \mathbb{E}_{t,x_0,x_t\sim p(x_t|x_0)}[\|S_\theta(x_t,t) - \nabla_{x_t} \log p(x_t|x_0)\|_2^2] \tag{5}$$

By defining $f(x,t)$ as 0, $g(t)$ as $\sqrt{d[\sigma^2(t)]/dt}$, and $\sigma(t)$ as a progressively increasing noise scale function, it becomes a Variance Exploding SDE (VE-SDE). The VE-SDE sampling process can be addressed by replacing the score function with a Score Network, a neural network specifically trained for learning the score function.

2.3. Diffusion Posterior Sampling & Blind DPS

Diffusion Posterior Sampling (DPS) [3] addresses inverse problem by employing diffusion model and posterior mean. When we consider the distribution of inverse problem,

$$P(y|x_0) = N(y|A(x_0), \sigma^2 I), y \in \mathbb{R}^m, x \in \mathbb{R}^d \tag{6}$$

y is the measurement, A is the forward projection matrix, and Chung *et al.* [3] wish to recover x_0. Then, the gradient of the log-likelihood that Chung *et al.* [3] aim to solve is as follows.

$$\nabla_{x_t} \log p(x_t|y) = \nabla_{x_t} \log p(y|x_t) + \nabla_{x_t} \log p(x_t) \tag{7}$$

However in this scenario, $\nabla_{x_t} \log p(y|x_t)$ is intractable. To solve this problem, Chung *et al.* [3] introduce posterior mean that approximate X_0. Then Eq. (7) becomes

$$\nabla_{x_t} \log p(x_t|y) = \nabla_{x_t} \log p(y|\hat{x_0}) + \nabla_{x_t} \log p(x_t) \tag{8}$$

$$\hat{x_0} = x_t + \sigma_i^2 S_\theta^x(x_t, t) \tag{9}$$

$S_\theta^x(x_t, t)$ is the score function of x_t and $\hat{x_0}$ is the posterior mean. A blind inverse problem involves finding both x_0 and A when neither is known. Blind DPS [2] extends the method of DPS to solve the blind inverse problem by using two diffusion models that model x_0 and A respectively, and utilizes the posterior mean for each. The solution of Blind DPS is as follows.

$$\nabla_{x_t} \log p(A_t, x_t|y) = \nabla_{x_t} \log p(y|\hat{A}_0, \hat{x_0}) + \nabla_{x_t} \log p(x_t)$$

$$\nabla_{A_t} \log p(A_t, x_t|y) = \nabla_{A_t} \log p(y|\hat{A}_0, \hat{x_0}) + \nabla_{A_t} \log p(A_t) \tag{10}$$

Blind DPS [2] can generate high quality outputs in situations where two or more variables interact by providing guidance to both A_t and x_t from y. However, this approach is time-intensive, as backpropagation for each diffusion model must be carried out at every step. Additionally, it imposes a significant memory burden, especially when reconstructing three-dimensional data.

3. Dual Domain Diffusion Guidance for 3D CBCT MAR

3.1. Projection Guidance for 3D Connectivity

In this work, we aim to separate metal-corrupted CBCT into metal-free CBCT and metal artifacts. where,

$$x_0 = x_0^* + x_0^m \tag{11}$$

x_0 is the metal-corrupted CBCT, x_0^* is the metal-free CBCT, and x_0^m indicates metal artifacts. Leveraging the strength of the diffusion model's prior, we employ distinct diffusion models to separately model the metal-free CBCT and metal artifacts. The posterior distribution at time t is as follows.

$$P(x_t^*, x_t^m|x_0) \propto P(x_0|x_t^*, x_t^m)p(x_t^*, x_t^m) \tag{12}$$

Our work attempts to solve the 3D MAR problem in CBCT. While it is possible to utilize a 3D Diffusion model, this approach is excessively complex, inefficient and requires thousands of 3D volumes for training [23]. Therefore, we use 2D slice-based diffusion models. Additionally, we introduce 3D cone beam projection to address the inconsistency issue along the z-axis. As the cone beam projection provides a lateral view, it can effectively resolve the inconsistency issue along the z-axis.

Key idea. The problem is defined as follows.

$$y = A(x_0) + n \tag{13}$$

Where A is the 3D cone beam projection function, y is the 3D cone beam projection given x_0, and n is Gaussian noise. As mentioned above, because x_0 is created using a 2D slice-based diffusion model, there can be inconsistency problems between the z-axis when we give diffusion guidance only with x_0. If we perform a 3D cone beam projection on x_0, it becomes a projection seen from the side and information about the z-axis is created, see Fig. 1. Therefore, we introduce additional guidance for the 3D cone beam projection y. The posterior distribution given y is defined as follows.

$$P(x_t^*, x_t^m|x_0, y) \propto P(x_0|x_t^*, x_t^m, y)p(x_t^*, x_t^m|y) \tag{14}$$

When generating synthetic y, Gaussian noise is not considered, and since y is determined by x_0, y can be ignored in the conditional distribution specified in Eq. (14). Then, according to the Bayes' rule, it can be expressed as follows.

$$P(x_t^*, x_t^m|x_0, y) \propto P(x_0|x_t^*, x_t^m)P(y|x_t^*, x_t^m)P(x_t^*)P(x_t^m) \tag{15}$$

We focus more on MAR in the image domain and for efficient sampling, we incorporate 3D cone beam projection guidance only once per every k step of diffusion sampling. This is sufficient to solve the 3D connectivity problem. Therefore, the sampling process varies depending on whether the diffusion sampling step is a multiple of k or not. The gradient of the log-likelihood term with respect to k is as follows.

$$\begin{aligned}\nabla_{x_t^*} \log p(x_t^*, x_t^m|x_0, y) \simeq \alpha_1 \nabla_{x_t^*} \log p(x_0|x_t^*, x_t^m)\\ + \beta_1 \nabla_{x_t^*} \log p(y|x_t^*, x_t^m) + \nabla_{x_t^*} \log p(x_t^*)\end{aligned}$$

$$\begin{aligned}\nabla_{x_t^m} \log p(x_t^*, x_t^m|x_0, y) \simeq \alpha_2 \nabla_{x_t^m} \log p(x_0|x_t^*, x_t^m)\\ + \beta_2 \nabla_{x_t^m} \log p(y|x_t^*, x_t^m) + \nabla_{x_t^m} \log p(x_t^m)\end{aligned} \tag{16}$$

α_1, α_2, β_1, and β_2 are diffusion guidance hyperparameters. When the sampling step is a multiple of k, we set β_1 and β_2 as zero.

3.2. Posterior mean Guidance for efficient sampling

In Eq. (16), the terms $\log p(x_0|x_t^*, x_t^m)$ and $\log p(y|x_t^*, x_t^m)$ are typically intractable. According to Blind DPS [2], if x_t^* and x_t^m can be sampled from independent diffusion models, the terms $\log p(x_0|x_t^*, x_t^m)$ and $\log p(y|x_t^*, x_t^m)$ can be approximated as $\log p(x_0|\hat{x_0^*}, \hat{x_0^m})$ and $\log p(y|\hat{x_0^*}, \hat{x_0^m})$, a tractable distribution, where $\hat{x_0^*}$ and $\hat{x_0^m}$ are the posterior means of x_t^* and x_t^m. Since $x_0^* := x_t^* + \sigma_t^2 S_\theta^*(x_t^*, t)$, a diffusion model must backpropagate whole parameters in every reverse diffusion step to calculate $\nabla_{x_t^*} \log p(x_0|\hat{x_0^*}, \hat{x_0^m})$. The same applies to the metal artifact diffusion model. For the 3D cone beam projection diffusion guidance, when there are M slices in CBCT, it necessitates the loading of $2M$ diffusion models

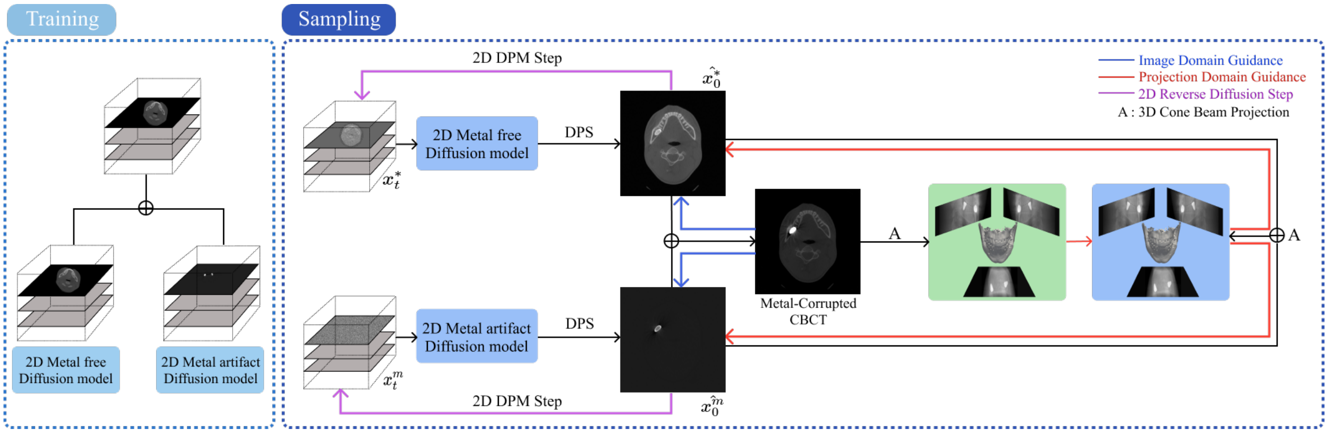
Training
Sampling
2D Metal free Diffusion model
2D Metal artifact Diffusion model
2D DPM Step
2D Metal free Diffusion model
DPS
x_t^*
$\hat{x_0^*}$
2D Metal artifact Diffusion model
DPS
x_t^m
$\hat{x_0^m}$
2D DPM Step
Metal-Corrupted CBCT
A
A
Image Domain Guidance
Projection Domain Guidance
2D Reverse Diffusion Step
A : 3D Cone Beam Projection

Figure 2. Overview of our method. (left) In training process, The 2D metal-free diffusion model and the 2D metal artifact diffusion model learn the metal-free prior and metal artifact prior, respectively. (right) In sampling process, the noisy outputs of the metal-free diffusion model and the metal artifact diffusion model (x_t^*, x_t^m) predict $\hat{x_0^*}$ and $\hat{x_0^m}$ through diffusion posterior sampling (DPS). The sum of the two is used to calculate the residual from the metal-corrupted CBCT, and the $\hat{x_0^*}$ and $\hat{x_0^m}$ are updated through guidance in the image domain (blue line). Then, the updated $\hat{x_0^*}$ and $\hat{x_0^m}$ are added to perform a 3D cone beam projection, and the residual is calculated from the 3D cone beam projection of the metal-corrupted CBCT to update $\hat{x_0^*}$ and $\hat{x_0^m}$ through projection domain guidance (red line). Finally, reverse sampling to predict x_{t-1}^*, x_{t-1}^m is performed (purple line).

into memory simultaneously. This approach is highly memory inefficient, see Tab. 3. To circumvent this problem, we hijack the path to x_t^*, x_t^m and instead provide guidance for $\hat{x_0^*}, \hat{x_0^t}$, rather than for x_t^*, x_t^m.

$$\nabla_{x_t^*} \log p(x_t^*, x_t^m | x_0, y) \simeq \alpha_1 \nabla_{\hat{x_0^*}} \log p(x_0 | \hat{x_0^*}, \hat{x_0^m}) + \beta_1 \nabla_{\hat{x_0^*}} \log p(y | \hat{x_0^*}, \hat{x_0^m}) + \nabla_{x_t^*} \log p(x_t^*) \quad (17)$$

$$\nabla_{x_t^m} \log p(x_t^*, x_t^m | x_0, y) \simeq \alpha_2 \nabla_{\hat{x_0^m}} \log p(x_0 | \hat{x_0^*}, \hat{x_0^m}) + \beta_2 \nabla_{\hat{x_0^m}} \log p(y | \hat{x_0^*}, \hat{x_0^m}) + \nabla_{x_t^m} \log p(x_t^m) \quad (18)$$

In Eq. (17), Eq. (18) we expect that, as the diffusion steps progress, $(\hat{x_0^*} + \hat{x_0^m})$ should gradually approach x_0. Intuitively, before taking the reverse diffusion step for x_t^* and x_t^m, we can update the posterior mean via guidance to make it closer to x_0, and then take the reverse diffusion step. Our training and sampling processes are shown in Fig. 2 and our sampling algorithm is given in Algorithm 1. In Algorithm 1, $[j]$ refers to the j-th slice of the CT scan.

Algorithm 1 Dual Domain Diffusion Guidance

```
Require: : $N, M, y, \alpha_1, \alpha_2, \beta_1, \beta_2, K, \{\sigma_i\}_{i=1}^N$
$x_N^*, x_N^m \in \mathbb{R}^{d_z, d_x, d_y}, Z \sim \mathcal{N}(0, I)$
for $i \leftarrow N - 1$ to $0$ do
    for $j \leftarrow 1$ to $M$ do
        $\hat{S}^*[j] \leftarrow \hat{S}_\theta^*(x_{i+1}^*[j], \sigma_{i+1})$
        $\hat{S}^m[j] \leftarrow \hat{S}_\theta^m(x_{i+1}^m[j], \sigma_{i+1})$
        $\hat{x}_0^*[j] \leftarrow x_{i+1}^*[j] + \sigma_{i+1}^2 \hat{S}^*$
        $\hat{x}_0^m[j] \leftarrow x_{i+1}^m[j] + \sigma_{i+1}^2 \hat{S}^m$
        $\hat{x}_0^{*'}[j] \leftarrow \hat{x}_0^*[j] - \alpha_1 \nabla_{\hat{x}_0^*[j]} \|x_0[j] - (\hat{x}_0^*[j] + \hat{x}_0^m[j])\|_2$
        $\hat{x}_0^{m'}[j] \leftarrow \hat{x}_0^m[j] - \alpha_2 \nabla_{\hat{x}_0^m[j]} \|x_0[j] - (\hat{x}_0^*[j] + \hat{x}_0^m[j])\|_2$
    end for
    if (i%K == 0) then
        $\hat{x}_0^{*''} \leftarrow \hat{x}_0^{*'} - \beta_1 \nabla_{\hat{x}_0^*} \|y - A(\hat{x}_0^* + \hat{x}_0^m)\|_2$
        $\hat{x}_0^{m''} \leftarrow \hat{x}_0^{m'} - \beta_2 \nabla_{\hat{x}_0^m} \|y - A(\hat{x}_0^* + \hat{x}_0^m)\|_2$
    end if
    $x_i^{*'} \leftarrow \hat{x}_0^{*''} - \sigma_i^2 \hat{S}^* + \sqrt{\sigma_{i+1}^2 - \sigma_i^2} Z$
    $x_i^{m'} \leftarrow \hat{x}_0^{m''} - \sigma_i^2 \hat{S}^m + \sqrt{\sigma_{i+1}^2 - \sigma_i^2} Z$
end for
return $x_0^*, x_0^m$
```

4. Experiment

4.1. Dataset

Synthesized Dataset. We obtained 48 dental CBCT scans from a commercial CBCT scanner (i-Cat 17-19, Imaging Sciences International) with a tube voltage of 120kVp and tube current of 5mA. The voxel size was 768×768×576 with 0.3mm real scale along each axis. These scans were from patients without metal inserts. Data collection and experiment were conducted with the approval of the Institutional Review Board (IRB number: 2020AN0410) of our organization.

Method	PSNR ↑	SSIM ↑		
		Axial	Coronal	Sagittal
InDuDoNet+ [29]	24.99	<u>0.9669</u>	<u>0.9674</u>	<u>0.9678</u>
ACDNet [32]	24.92	0.9661	0.9670	0.9675
FEL [22]	33.12	0.9621	0.9528	0.9539
Blind DPS [2]	<u>35.48</u>	0.9667	0.9413	0.9419
Ours	**38.21**	**0.9887**	**0.9875**	**0.9873**

Table 1. Synthetic Metal-corrupted CBCT MAR evaluation results (PSNR, SSIM). **Bold**: best, <u>under</u>: second best.

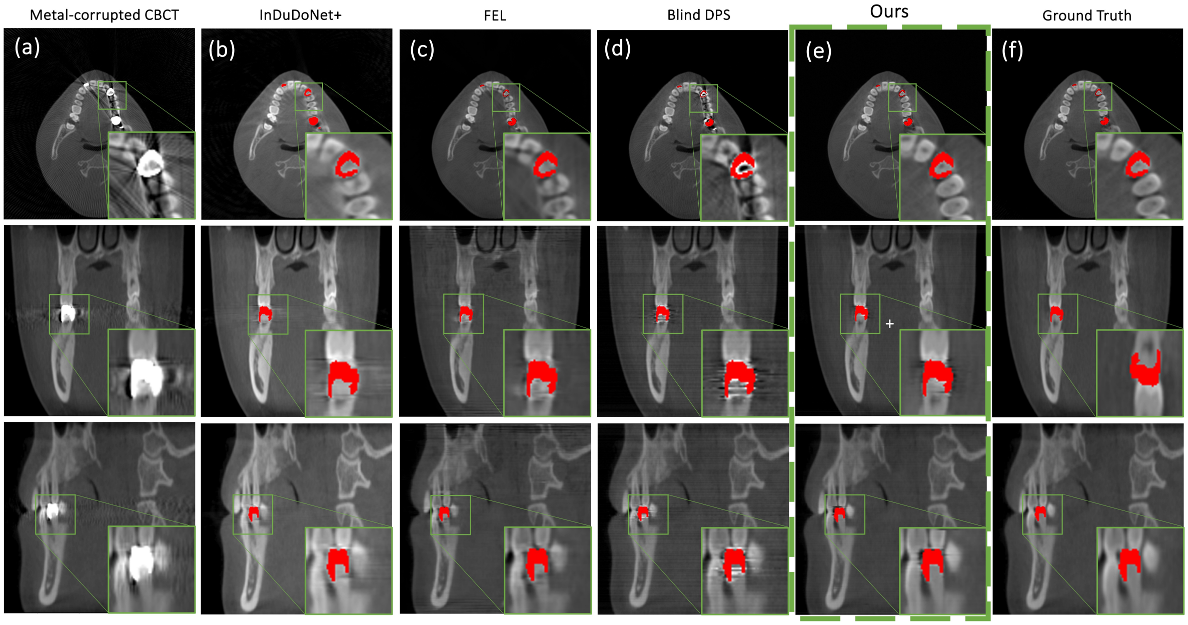

Figure 3. Synthetic metal-corrupted CBCT MAR qualitative results. (First row: axial, Second row: coronal, Third row: sagittal) Red: metal inserts. (a) Metal-corrupted CBCT, (b) InDuDoNet+ (c) FEL (d) Blind DPS (e) Ours, (f) GT. HU values were clipped between -1000, 2500.

To train the diffusion model, synthetic metal artifacts were created in the CBCT volumes. To simulate synthetic implants or crowns, whole tooth or enamel was segmented from the CBCT scans, and metal inserts were generated by randomly selecting 1 to 8 teeth. Metal inserts were assumed to be Au or Zr. 3D cone beam projections were conducted using 120 kVp tube voltage and 5 mA tube current, and the polychromatic model described in Eq. (1). The attenuation coefficient values were adopted from previous research [14]. Poisson and electronic noise were added to the generated metal projections. The electronic noise was modeled as Gaussian noise, similar to the method by [16, 22].

The metal projections were added to CBCT projections, and the FDK algorithm was employed to reconstruct metal-affected 3D CBCT volumes. For the projections and FDK algorithm, We adopted the Tomosipo library [11]. Metal artifacts were created by subtracting the CBCT volumes without metal presence from the metal-affected CBCT volumes of the same patient.

For the training process, we utilized 23 out of 48 artifact-free CBCT scans to create the diffusion model for generating CBCT and the diffusion model for generating metal artifacts. The remaining 25 scans were used as a test set. Each CBCT slice was resized to 384×384 for computational efficiency like [4], and only the slices from 250 to 506, where teeth were consistently present among the total 576 slices, were used for training and evaluation.

Clinical Dataset. The CBCT volumes of 3 patients with actual metal artifacts were obtained using the same equipment as the synthesized dataset. These CBCT volumes were used as a test set to verify the clinical validity. The number and types of metal inserts in the 3 CBCT volumes were diverse, including implants, crowns. The number of inserts ranged from a single instance up to nearly the entire tooth in severe cases.

Out of Distribution Dataset. Cui *et al*. collected 4938

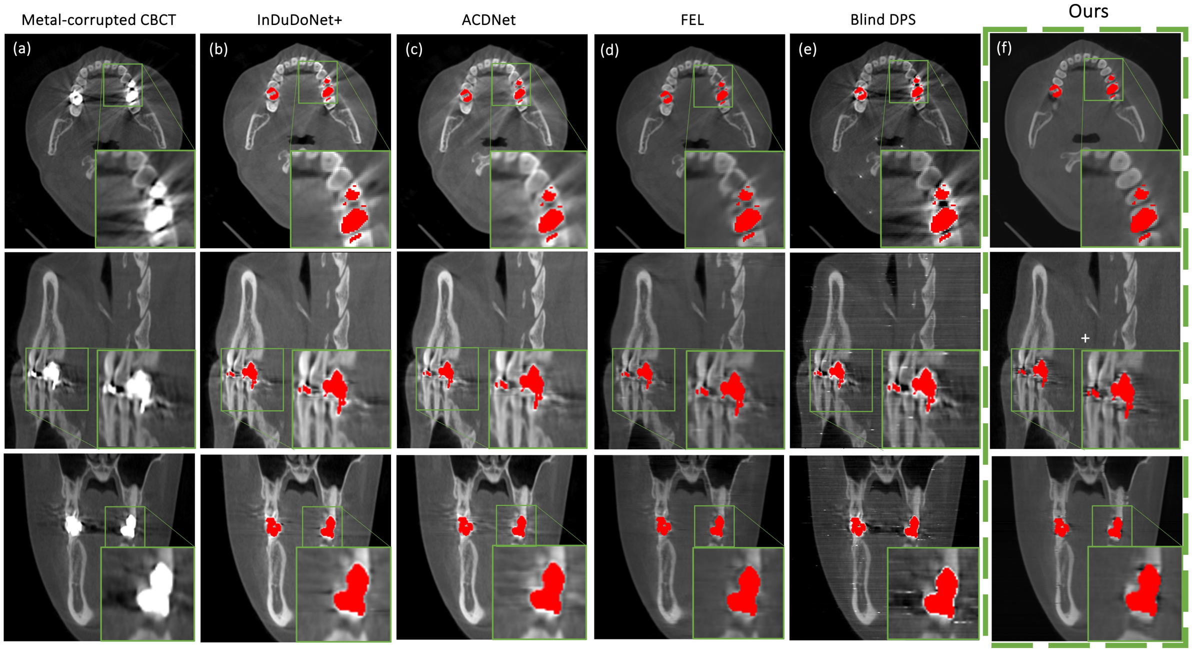

Figure 4. Clinical metal-corrupted CBCT MAR qualitative results. (First row: axial, Second row: coronal, Third row: sagittal) Red: Metal inserts. (a) Metal-corrupted CBCT, (b) InDuDoNet+ (c) ACDNet (d) FEL (e) Blind DPS (f) Ours, HU values were clipped between -1000, 2500.

Method	**PSNR↑**	**SSIM ↑**		
		Axial	Coronal	Sagittal
InDuDoNet+ [29]	21.50	0.9314	0.9252	0.9265
ACDNet [32]	22.60	<u>0.9397</u>	<u>0.9385</u>	<u>0.9404</u>
FEL [22]	20.77	0.7499	0.7212	0.7309
Blind DPS [2]	<u>26.40</u>	0.8770	0.7754	0.7794
Ours	**31.83**	**0.9565**	**0.9533**	**0.9557**

Table 2. Synthetic out-of-distribution metal-corrupted CBCT MAR evaluation results (PSNR, SSIM). **Bold**: best, <u>under</u>: second best.

Method	**Blind DPS [2]**	**Ours**
Memory (GB)	5299.7	**9.3**

Table 3. Memory usage of Blind DPS and ours when a CBCT is 256×384×384 size, and dual domain guidance is considered.

CBCT volumes from 15 different hospitals for CBCT segmentation work [5,6] and provided 50 volumes publicly for research purposes. Among them, we selected 6 clean CBCT volumes without any metal artifacts to create the synthetic dataset. The imaging equipment, projection geometry, and resolution were all different from ours. The size of a single CBCT slice was 400×400, with the number of slices varying between 240 and 280. Out of these, we cropped and resized the volumes to the dimensions of 200×256×256, and then generated synthetic metal-corrupted CBCT images using the method described in Eq. (13).

4.2. Implementation Details

For training, metal-free CBCT images, metal-affected CBCT images, and metal artifact residuals were clipped from -1000 to 2500 HU, from -1000 to 13,000 HU, and from -3500 to 14,000 HU, followed by normalization.

Both diffusion models employed VE-SDE and ncsnpp models [25]. With a batch size of 4, two A100 GPUs were utilized to train each model, through 320K and 190K training iterations respectively. Sampling was discretized with N=1000 and performed using ancestral sampling [12]. For sampling, the parameters α_1 and α_2 were set to 0.01 and 0.003, while β_1 and β_2 were both set to 0.1 experimentally. K was set to 10.

4.3. Comparison

Comparison methods. In our study, we compare the 3D CBCT MAR with various state-of-the-art (SOTA) methods.

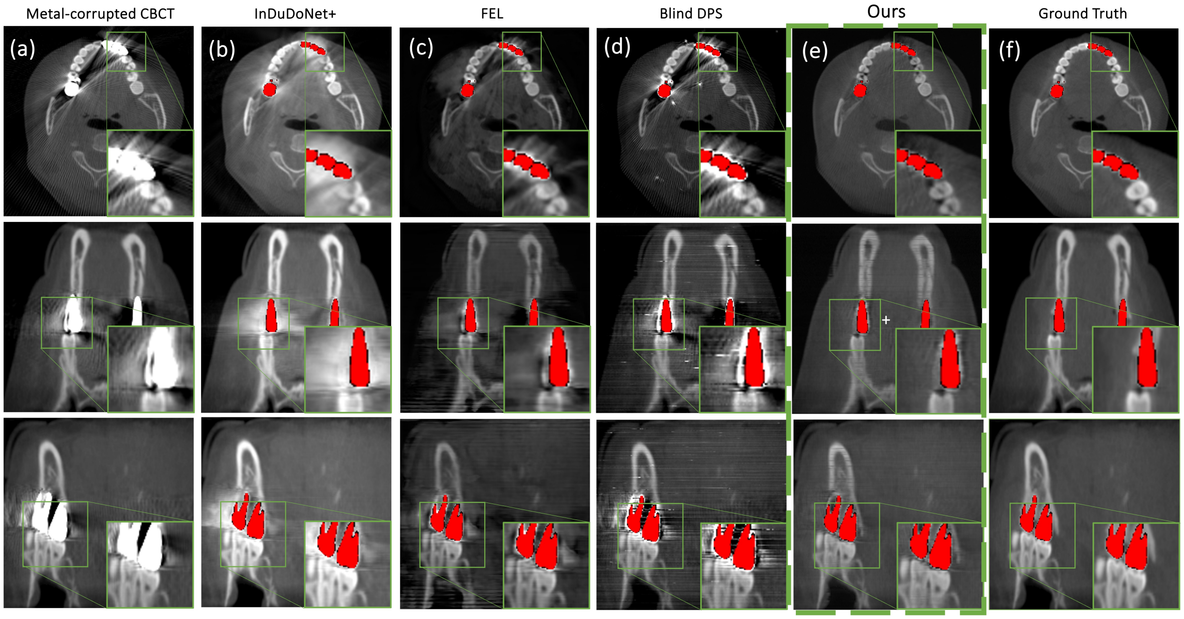

Figure 5. Synthetic out of distribution metal-corrupted CBCT MAR qualitative results. (First row: axial, Second row: coronal, Third row: sagittal) Red: Metal inserts. (a) Metal-corrupted CBCT, (b) InDuDoNet+ (c) FEL (d) Blind DPS (e) Ours, (f) GT, HU values were clipped between -1000, 2500.

We evaluated our method against 2D CT MAR approaches, which include InDuDoNet+ [29] that performs dual domain optimization and ACDNet [32] that employs metal kernel modeling. As for 3D CBCT MAR method, we compare our work with FEL [22], which utilizes 2D CNN and 3D cone beam projection domain optimization. Finally, we compare our approach against Blind DPS [2], which models kernel prior and image prior using a diffusion model to solve the blind inverse problem.

Synthetic dataset Results. Tab. 1 presents the quantitative results for synthetic data. Both PSNR and SSIM show significant improvements over existing methods. 2D MAR approaches, such as InDuDoNet+ [29] and ACDNet [32], exhibit a considerable drop in PSNR, which is due to the change of the HU value distribution, as illustrated in Fig. 3. Our method outperforms FEL [22] and Blind DPS [2] in terms of both PSNR and SSIM. As demonstrated in Fig. 3 (e), our method effectively removes metal artifacts across axial, coronal, and sagittal planes. Traditional CNN-based methods, such as InDuDoNet+ (Fig. 3 (b)), and FEL (Fig. 3 (c)), tend to blur CT during the MAR process. Meanwhile, Blind DPS (Fig. 3 (d)) creates weird artifacts and exhibits reduced connectivity in the coronal and sagittal planes.

Clinical dataset Results. Fig. 4 shows the results of a clinical case. InDuDoNet+ [29], ACDNet [32], and FEL [22] effectively remove metal artifacts in coronal and sagittal planes; however, they produce some smoothing or residual scattering around the metal insert and exhibit poor performance in removing artifacts around metal inserts in the axial plane, see Fig. 4 (b), (c), (d). Blind DPS [2] introduces strange artifacts and poor connectivity in coronal and sagittal planes, (Fig. 4 (e)). In contrast, our method effectively removes metal artifacts in all three planes (Fig. 4 (f)).

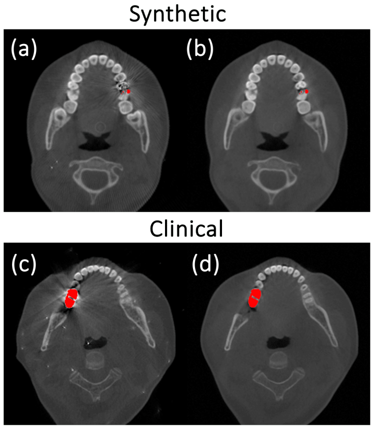

Figure 6. Synthetic and clinical MAR results of Blind DPS and ours w/o projection domain guidance. (a) is the synthetic output of Blind DPS, (b) is the synthetic output of ours w/o projection guidance. (c) shows the clinical output of Blind DPS and (d) shows the clinical output of ours w/o projection domain guidance. HU values were clipped between -1000, 2500.

Method	PSNR	SSIM		
		Axial	Coronal	Sagittal
w/o image domain guidance	25.31	0.8543	0.8103	0.8154
w/o projection domain guidance	37.08	0.9841	0.9594	0.9595
Ours	**38.21**	**0.9887**	**0.9878**	**0.9873**

Table 4. Ablation results of synthetic metal-corrupted CBCT MAR. (PSNR, SSIM)

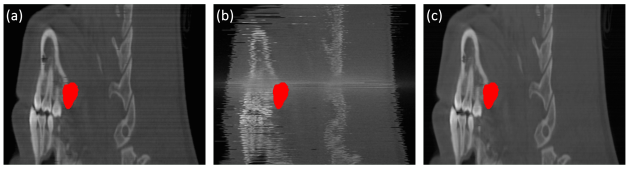

Figure 7. Ablation results of Synthetic metal-Corrupted CBCT MAR sagittal view. Red: Metal inserts. (a) w/o projection domain guidance (b) w/o image domain guidance (c) Ours, HU values were clipped between -1000, 2500.

Synthetic out-of-distribution dataset Results. Tab. 2 presents the quantitative results for out-of-distribution synthetic data from [5]. Overall performance is slightly lower due to the smaller margin in the CBCT volume compared to our synthetic dataset, and the use of different imaging equipment. InDuDoNet+ [29] and ACDNet [32] exhibit similar trends as our synthetic dataset. As shown in Fig. 5, they face challenges in handling changing HU value distributions and accurately removing artifacts around metal inserts. FEL [22] experiences a significant performance drop in the out-of-distribution dataset, suggesting a poor generalization capability. Blind DPS [2] produces strange artifacts in the CT image, and the SSIM performance sharply drops in the coronal and sagittal planes compared to the axial plane. Conversely, our method effectively removes metal artifacts in out-of-distribution datasets, demonstrating consistent results across axial, coronal, and sagittal planes.

Posterior mean Guidance for efficient sampling. As exhibited in Tab. 3, our method shows significantly reduced memory usage compared to Blind DPS. When there are M slices of CBCT, it uses more than $2M\times$ less memory. However, as observed in Fig. 6, our method presents superior performance. The posterior means $\hat{x_0^*}$ and $\hat{x_0^m}$ gradually approximate the desired x_0^* and x_0^m. By providing guidance to these posterior means that approach the final values of interest, rather than to x_t^* and x_t^m, we can remove metal artifacts more efficiently and effectively.

4.4. Ablation

To demonstrate the necessity of dual guidance, we conducted an ablation study. In w/o projection domain guidance, the parameters $\alpha_1, \alpha_2, \beta_1, \beta_2$ were set to 0.01, 0.003, 0, 0, and in w/o image domain guidance, the parameters were set to 0, 0, 100, 35. To compensate the absence of image domain guidance, the scale of β_1,β_2 was increased, and k was set to 1. As shown in Tab. 4, when both guidance is added, the SSIM of the axial, coronal, and sagittal planes increase to similar levels. When one of the guidances is excluded, the SSIM values for the coronal and sagittal planes drop. The reason why the SSIM for the coronal and sagittal planes drop when projection domain guidance is missing is that while connectivity along the z-axis increases, SSIM of the axial plane increases more significantly because the learned view of the diffusion model is an axial plane. Without projection domain guidance, there exists a problem with connectivity along the z-axis which results in lower SSIM for coronal and sagittal planes compared to axial. As shown in the Fig. 7, when there is no projection domain guidance, the 3D connectivity drops. When image domain guidance was not used, 3D connectivity was present but strange artifacts occurred because the guidance of the diffusion, which was directly learned from the image domain, was absent. In contrast, when dual domain guidance was present, the 3D connectivity was improved, and flickering did not occur.

5. Conclusion

In this work, we proposed a dual domain guided diffusion model for 3D metal artifact reduction. The diffusion model that models the metal-free CBCT prior and the metal artifact prior were used to model each distribution. Using image domain guidance, we remove the metal artifacts on each slice, and the cone beam projection domain guidance ensures 3D continuity. This approach provides excellent results in 3D MAR not only for dataset with the same distribution as the training data but also for out-of-distribution dataset and clinical dataset, thereby achieving state-of-the-art performance.

Acknowledgements. This work was supported by ICT Creative Consilience program (IITP-2020-0-01819) supervised by the Institute for Information & communications Technology Planning & Evaluation, National Research Foundation of Korea (NRF) grant (No. 2021R1A2C1007215 and No.2022R1A5A1027646, and the Korea Medical Device Development Fund grant (Project Number: 1711195279, RS-2021-KD000009) by MSIT of the Korean Government.

References

[1] Zhiqian Chang, Dong Hye Ye, Somesh Srivastava, Jean-Baptiste Thibault, Ken Sauer, and Charles Bouman. Prior-guided metal artifact reduction for iterative x-ray computed tomography. *IEEE transactions on medical imaging*, 38(6):1532–1542, 2018. 1

[2] Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6059–6069, 2023. 3, 5, 6, 7, 8

[3] Hyungjin Chung, Jeongsol Kim, Michael T Mccann, Marc L Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. *arXiv preprint arXiv:2209.14687*, 2022. 2

[4] Hyungjin Chung, Dohoon Ryu, Michael T McCann, Marc L Klasky, and Jong Chul Ye. Solving 3d inverse problems using pre-trained 2d diffusion models. 2023. 5

[5] Zhiming Cui, Yu Fang, Lanzhuju Mei, Bojun Zhang, Bo Yu, Jiameng Liu, Caiwen Jiang, Yuhang Sun, Lei Ma, Jiawei Huang, et al. A fully automatic ai system for tooth and alveolar bone segmentation from cone-beam ct images. *Nature communications*, 13(1):2096, 2022. 6, 8

[6] Zhiming Cui, Bojun Zhang, Chunfeng Lian, Changjian Li, Lei Yang, Wenping Wang, Min Zhu, and Dinggang Shen. Hierarchical morphology-guided tooth instance segmentation from cbct images. In *International Conference on Information Processing in Medical Imaging*, pages 150–162. Springer, 2021. 6

[7] Bruno De Man, Johan Nuyts, Patrick Dupont, Guy Marchal, and Paul Suetens. Metal streak artifacts in x-ray computed tomography: a simulation study. In *1998 IEEE Nuclear Science Symposium Conference Record. 1998 IEEE Nuclear Science Symposium and Medical Imaging Conference (Cat. No. 98CH36255)*, volume 3, pages 1860–1865. IEEE, 1998. 1

[8] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *Josa a*, 1(6):612–619, 1984. 2

[9] 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. 1

[10] Amara Graps. An introduction to wavelets. *IEEE computational science and engineering*, 2(2):50–61, 1995. 1

[11] Allard A Hendriksen, Dirk Schut, Willem Jan Palenstijn, Nicola Viganó, Jisoo Kim, Daniël M Pelt, Tristan Van Leeuwen, and K Joost Batenburg. Tomosipo: fast, flexible, and convenient 3d tomography for complex scanning geometries in python. *Optics Express*, 29(24):40494–40513, 2021. 5

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 2, 6

[13] 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*, 17:1–15, 2018. 1

[14] John H Hubbell and Stephen M Seltzer. Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients 1 kev to 20 mev for elements z= 1 to 92 and 48 additional substances of dosimetric interest. Technical report, National Inst. of Standards and Technology-PL, Gaithersburg, MD (United . . . , 1995. 5

[15] Willi A Kalender, Robert Hebel, and Johannes Ebersberger. Reduction of ct artifacts caused by metallic implants. *Radiology*, 164(2):576–577, 1987. 1

[16] Patrick J La Riviere and David Matthew Billmire. Reduction of noise-induced streak artifacts in x-ray computed tomography through spline-based penalized-likelihood sinogram smoothing. *IEEE transactions on medical imaging*, 24(1):105–111, 2005. 5

[17] Junghyun Lee, Jawook Gu, and Jong Chul Ye. Unsupervised ct metal artifact learning using attention-guided β-cyclegan. *IEEE Transactions on Medical Imaging*, 40(12):3932–3944, 2021. 1

[18] Haofu Liao, Wei-An Lin, Zhimin Huo, Levon Vogelsang, William J Sehnert, S Kevin Zhou, and Jiebo Luo. Generative mask pyramid network for ct/cbct metal artifact reduction with joint projection-sinogram correction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, pages 77–85. Springer, 2019. 1

[19] Haofu Liao, Wei-An Lin, S Kevin Zhou, and Jiebo Luo. Adn: artifact disentanglement network for unsupervised metal artifact reduction. *IEEE Transactions on Medical Imaging*, 39(3):634–643, 2019. 1

[20] 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. 1

[21] Hyoung Suk Park, Sung Min Lee, Hwa Pyung Kim, Jin Keun Seo, and Yong Eun Chung. Ct sinogram-consistency learning for metal-induced beam hardening correction. *Medical physics*, 45(12):5376–5384, 2018. 1

[22] Hyoung Suk Park, Jin Keun Seo, Chang Min Hyun, Sung Min Lee, and Kiwan Jeon. A fidelity-embedded learning for metal artifact reduction in dental cbct. *Medical physics*, 49(8):5195–5205, 2022. 1, 2, 5, 6, 7, 8

[23] Walter HL Pinaya, Petru-Daniel Tudosiu, Jessica Dafflon, Pedro F Da Costa, Virginia Fernandez, Parashkev Nachev, Sebastien Ourselin, and M Jorge Cardoso. Brain imaging generation with latent diffusion models. In *MICCAI Workshop on Deep Generative Models*, pages 117–126. Springer, 2022. 3

[24] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015. 1

[25] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020. 2, 6

[26] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011. 2

[27] 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*, 41(4):869–880, 2021. 1

[28] 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 *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pages 107–118. Springer, 2021. 1

[29] Hong Wang, Yuexiang Li, Haimiao Zhang, Deyu Meng, and Yefeng Zheng. Indudonet+: A deep unfolding dual domain network for metal artifact reduction in ct images. *Medical Image Analysis*, 85:102729, 2023. 1, 5, 6, 7, 8

[30] Hong Wang, Minghao Zhou, Dong Wei, Yuexiang Li, and Yefeng Zheng. Mepnet: A model-driven equivariant proximal network for joint sparse-view reconstruction and metal artifact reduction in ct images. *arXiv preprint arXiv:2306.14274*, 2023. 1

[31] 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. 1

[32] Chuanqing Zhuang, Zhengda Lu, Yiqun Wang, Jun Xiao, and Ying Wang. Acdnet: Adaptively combined dilated convolution for monocular panorama depth estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 3653–3661, 2022. 5, 6, 7, 8