DCDiff: Dual-Domain Conditional Diffusion for CT Metal Artifact Reduction

Ruochong Shen[1], Xiaoxu Li[2], Yuan-Fang Li[1], Chao Sui[2], Yu Peng[2], Qiuhong Ke[1,⋆⋆]

[1] Monash University, Melbourne, Australia
{ruochong.shen, yuanfang.li, qiuhong.ke}@monash.edu
[2] CurveBeam AI, Melbourne, Australia
{shawn.li, yu.peng, chao.sui}@curvebeamai.com

Abstract. Metallic implants in X-ray Computed Tomography (CT) scans can lead to undesirable artifacts, adversely affecting the quality of images and, consequently, the effectiveness of clinical treatment. Metal Artifact Reduction (MAR) is essential for improving diagnostic accuracy, yet this task is challenging due to the uncertainty associated with the affected regions. In this paper, inspired by the capabilities of diffusion models in generating high-quality images, we present a novel MAR framework termed Dual-Domain Conditional Diffusion (DCDiff). Specifically, our DCDiff takes dual-domain information as the input conditions for generating clean images: 1) the image domain incorporating raw CT image and the filtered back project (FBP) output of the metal trace, and 2) the sinogram domain achieved with a new diffusion interpolation algorithm. Experimental results demonstrate that our DCDiff outperforms state-of-the-art methods, showcasing its effectiveness for MAR.

Keywords: Computed tomography (CT) · Metal artifact reduction (MAR) · Diffusion models

1 Introduction

Computed Tomography (CT) is a crucial medical imaging method for obtaining detailed internal body pictures, commonly employed in diagnosis. Nevertheless, various types of noise can compromise CT image quality. Notably, metal artifacts pose a severe challenge arising from metal implants such as surgical instruments, pacemakers, and orthopedic devices. These metallic objects can induce substantial distortions in the images, hindering accurate diagnosis.

Numerous algorithms, spanning traditional methods rooted in mathematics and physics (e.g., linear interpolation (LI) [5], normalized MAR (NMAR) [11]), and advanced deep learning-based approaches(e.g. CNNMAR[21], DuDoNet[7], cGANMAR[16], ADN[6], DSCMAR[20], InDuDoNet[14], OSCNet[15]), have been proposed for MAR. Despite these endeavors, the development of effective algorithms to sufficiently reduce artifacts remains a challenge. This challenge stems from the inherent uncertainty associated with the nature and extent of these artifacts. They can manifest as a wide range of image distortions, including streaks, shading, and blurring. Additionally, the intricate interactions between X-rays and metal objects further contribute to the uncertainties, making it challenging to accurately predict and correct the impact of metal artifacts.

Recently, Denoising Diffusion Probabilistic Models (DDPMs) [4], also known as diffusion models, have emerged as an effective tool for high-quality image generation [2], surpassing previous state-of-the-art methods such as GANs [3]. The diffusion models exhibit the ability to generate samples conforming to a specified data distribution (e.g., natural images) by iteratively removing noise from random, indeterminate inputs [4]. The concept of progressive denoising employed by diffusion models intuitively bridges the gap between highly uncertain and determinate distributions, breaking it down into smaller intermediate steps. This approach facilitates the model's smooth convergence towards generating samples in the target distributions. Motivated by the robust capabilities of DDPMs, we propose leveraging diffusion models for Metal Artifact Reduction (MAR), which also involves handling uncertainty and indeterminacy of the metal artifacts. However, the utilization of diffusion models for MAR presents non-trivial challenges: 1) Traditional diffusion models primarily focus on transforming data from a simple Gaussian distribution to a target image distribution. However, MAR demands the model to take specific images containing metal artifacts as input, deviating from the conventional approach. 2) Traditional diffusion models usually focus on a singular image modality, yet MAR necessitates the inclusion of other essential modalities. For instance, the sinogram has demonstrated effectiveness in understanding physical constraints and enhancing image contents for improved MAR [7]. This highlights the imperative for a more comprehensive multi-modal approach.

In this paper, we propose DCDiff, a novel dual-domain conditional diffusion-based framework, to address the challenges for effective MAR. DCDiff incorporates two conditional diffusion models, each utilizing CT scans with metal artifacts in either the image or sinogram domains as input conditions during the reverse diffusion process. This strategy guides the models to remove artifacts effectively. In the sinogram domain, we depart from the conventional diffusion model approach by forcing the model to generate contents exclusively on the metal trace during inference. This ensures the preservation of original contents without artifacts. In the image domain, in addition to utilizing raw CT images and the filtered back projection (FBP) of the sinogram model outputs, we introduce the FBP of the metal trace as a condition. This addition provides valuable information about the artifact structure, thereby enhancing overall performance. While dual-domain diffusion-based framework is also proposed for CT MAR in some recent works[1,8], our approach differs from others by supervised training two diffusion models in both the sinogram and image domains.

Our method has demonstrated efficacy through extensive experiments conducted on the DeepLesion[19] dataset. Our contributions can be summarized as follows: (1) We present a pioneering diffusion-based framework for MAR, marking one of the first explorations of leveraging supervised conditional diffusion models tailored for MAR tasks. (2) We introduce a novel dual-domain conditional design, enabling diffusion models to leverage multi-modal data with artifacts for generating clean images. (3) Our method achieves new state-of-the-art performance.

2 Revisiting Diffusion Models

DDPM[4] comprises two processes: a forward diffusion process and a reverse diffusion process. In the forward process, given a sample x_0 from some prior distribution $x_0 \sim q(x_0)$, independent Gaussian noises are added T times, step by step, to produce latent variables $x_1, \ldots, x_T$ and these x_t make up a Markov chain. For time t, the noising process is defined as follows:

$$q(x_t|x_{t-1}) = \mathcal{N}(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t \cdot \mathbf{I}), \forall t \in \{1, \ldots, T\}. \tag{1}$$

The hyperparameters $\beta_t \in [0, 1)$ denote the variance schedule across diffusion steps, respectively. $\mathbf{I}$ is the identity matrix and $\mathcal{N}(x; \mu, \sigma)$ represents the normal distribution of mean μ and covariance σ. Let $\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$, one can derive as follow:

$$q(x_t|x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t}x_0, \beta_t \cdot \mathbf{I}); x_t = \sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, \epsilon \sim \mathcal{N}(0, I). \tag{2}$$

For the reverse process, the following can be derived with the property of Gaussian distribution and Bayesian equation: $p_\theta(x_{t-1}|x_t, x_0) = \mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \sigma_t^2 I)$, where $\mu_\theta(x_t, t)$ is predicted by a network, named *denoising network*, with θ denoting its parameters. During the process, the inference x_0 is unknown, but it can be estimated by training the denoising network θ to predict $x_0 = f_\theta(x_t, t)$. We can then write the estimated μ_θ to the subject of f_θ as $\mu_\theta(x_t, t) = \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} f_\theta + \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t} x_t$. The model is trained by optimizing a variational lower bound of $\log p_\theta(x)$. This lower bound can be simplified as $L_{simple} = ||x_0 - f_\theta||^2$.

3 Methodology

As shown in Figure 1, our proposed framework, DCDiff, comprises two conditional diffusion models responsible for generating images and sinograms, respectively. Below, we describe the details.

3.1 Model Architecture and Training

As shown in Figure 1 (A), we start with $s_0 = S_{gt}$ and $x_0 = X_{gt}$ for the forward process of the sinogram and the image diffusion models, respectively. For any time step t uniformly chosen from 1 to the largest step $T = 1000$, the diffused sinogram s_t can be calculated as $s_t = \sqrt{\bar{\alpha}_t^{(s)}} s_0 + \sqrt{1 - \bar{\alpha}_t^{(s)}} \epsilon_1$ and diffused image

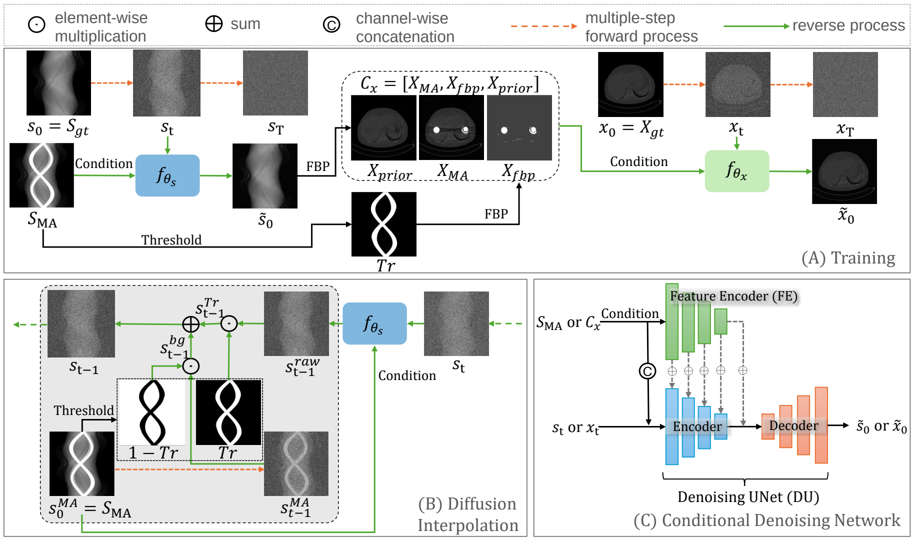

Fig. 1: Overview of the DCDiff framework. (A) depicts the training process of the dual-domain conditional diffusion models. (B) illustrates the proposed diffusion interpolation for sinogram generation during the testing phase. (C) displays the structures of the conditional denoising network.

x_t as $x_t = \sqrt{\bar{\alpha}_t^{(x)}} x_0 + \sqrt{1 - \bar{\alpha}_t^{(x)}} \epsilon_2$ where $\epsilon_1, \epsilon_2 \sim \mathcal{N}(0, I)$ denotes the Gaussian noises, and α_t is the hyperparameter related to the variance schedule across the diffusion steps, $\bar{\alpha}_t = \prod_{s=0}^{t} \alpha_s$. $\alpha_t^{(s)}$ and $\alpha_t^{(x)}$ denote that hyperparameter of the sinogram and the image diffusion model, respectively. To enable the models to remove artifacts for given CT scan inputs, we incorporate the inputs as conditions of the diffusion models in the reverse process.

Specifically, for the sinogram diffusion, we take the raw sinogram with metal trace, S_{MA}, as the condition: $C_s = [S_{MA}]$ of the conditional denoising network f_{θ_s}. The architecture of f_{θ_s} is shown in Figure 1 (C). Inspired by the denoising module in Diff-UNet[18], f_{θ_s} includes a UNet encoder, named feature encoder (FE), and a denoising UNet (DU). In this design, the input s_t is first channel-wise concatenated with the condition C_s before being fed into the DU's encoder. Simultaneously, C_s is inputted into FE, which mirrors the structure of DU's encoder. Consequently, the multi-scale features outputted by FE match in number and size with those extracted from the DU's encoder. We sum the features correspondingly to get the fused features, which is then fed to the DU's decoder to obtain the final output. Notably, we compel the sinogram denoising model to estimate the clean starting sinogram $\tilde{s}_0 = f_{\theta_s}(s_t, t, C_s)$ via minimising a simplified version of the variational evidence lower bound of DDPM, which is represented as $L_{simple_s} = ||s_0 - \tilde{s}_0||^2 = ||(S_{gt} - f_{\theta_s}(s_t, t, C_s)||^2$.

For the denoising network condition of the image diffusion model, we introduce not only the raw CT image with metal artifacts, X_{MA}, but also a prior image X_{prior} and an image condition X_{fbp}. In the training phase, the prior image is derived from the FBP output of $\tilde{s}_0$ obtained from our sinogram denoising network: $Xprior = \text{FBP}(\tilde{s}_0)$. X_{fbp} represents the FBP output of the metal trace: $X_{fbp} = \text{FBP}(Tr)$. The metal trace is derived by thresholding from S_{MA}. Although not commonly used in prior MAR works, X_{fbp} encompasses knowledge about the artifact structure, contributing to the enhancement of the output images. The denoising network condition is formed by concatenating the three images channel-wise: $C_x = [X_{MA}, X_{fbp}, X_{prior}]$. The architecture of the image conditional denoising network f_{θ_x} mirrors that of f_{θ_s}. It takes x_t as input with the condition C_x, producing the estimated starting clean image $\tilde{x}_0 = f_{\theta_x}(x_t, t, C_x)$ by minimizing $L_{simple_x} = ||x_0 - \tilde{x}_0||^2 = ||(X_{gt} - f_{\theta_x}(x_t, t, C_x)||^2$.

The two conditional diffusion model f_{θ_s} and f_{θ_x} are trained jointly by optimizing the final loss:

$$L = L_{simple_s} + L_{simple_x} = ||(S_{gt} - f_{\theta_s}(s_t, t, C_s)||^2 + ||(X_{gt} - f_{\theta_x}(x_t, t, C_x)||^2. \quad (3)$$

3.2 Model Inference with Diffusion Interpolation

Once the model is trained, intuitively, we can take Gaussian noises and the metal-affected CT scan X_{MA} and sinogram S_{MA} as input to conduct inference via the reverse process, allowing us to generate a clean image.

In the sinogram domain, the sampling process commences with Gaussian noise $s_T \sim \mathcal{N}(0, I)$ and proceeds iteratively through denoising, transforming s_T to s_0. To predict s_{t-1} from s_t $(1 \leq t \leq T)$, we depart from the conventional approach of utilizing the entire sinogram and following the DDPM sampling algorithm. Instead, drawing inspiration from recent advancements in image inpainting[9], we introduce a novel algorithm termed diffusion interpolation (DI), as illustrated in Figure 1 (B).

This idea stems from the observation that pixels in the background region, excluding the metal trace (referred to as the background region, s^{bg}, for simplicity), can be considered uncorrupted. This implies $S_{gt} \odot (1 - Tr) = S_{MA} \odot (1 - Tr)$, where Tr is the binary image of the metal trace, and $\odot$ denotes element-wise multiplication. Thus at each step of the denoising process, we use the intermediate sinogram s_{t-1}^{MA} generated from the forward diffusion process starting from S_{MA} to construct the background region to preserve its accurate characteristics.

Concretely, to predict s_{t-1}, we first calculate $s_{t-1}^{MA} = \sqrt{\bar{\alpha}_t^{(s)}} s_0^{MA} + (1 - \bar{\alpha}_t^{(s)})\epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$ and $s_0^{MA} = S_{MA}$. Meanwhile, s_t and the condition S_{MA} are inputted into the trained sinogram denoising network f_{θ_s} and make a raw prediction by $s_{t-1}^{raw} = \frac{\sqrt{\bar{\alpha}_{t-1}^{(s)}}(1 - \alpha_t^{(s)})}{1 - \bar{\alpha}_t^{(s)}} f_{\theta_s}(s_t, t, C_s) + \frac{\sqrt{\alpha_t^{(s)}}(1 - \bar{\alpha}_{t-1}^{(s)})}{1 - \bar{\alpha}_t^{(s)}} s_t + \sigma_t z$, where $z \sim \mathcal{N}(0, I)$, $\sigma_t = \sqrt{\frac{1 - \bar{\alpha}_{t-1}^{(s)}}{1 - \bar{\alpha}_t^{(s)}}(1 - \alpha_t^{(s)})}$ and the condition $C_s = [S_{MA}]$. Then, the binary metal trace Tr is estimated by thresholding S_{MA}. The pixel values of the background region are then obtained as $s_{t-1}^{bg} = s_{t-1}^{MA} \odot (1 - Tr)$, and those of the metal trace region as $s_{t-1}^{Tr} = s_{t-1}^{raw} \odot Tr$. The final prediction for s_{t-1} is a combination of these results: $s_{t-1} = s_{t-1}^{bg} + s_{t-1}^{Tr}$. This process is iterated, and the predicted sinogram is sampled as $S_{pred} = s_0$. Note that while the output of $f_{\theta_s}(s_t, t, C_s)$ is intended to estimate the clean input during training for model optimization, during inference, we adopt an iterative sampling approach for the final result rather than directly estimating s_0 from s_t. This strategy is employed to fully leverage the model's capabilities and attain a high-quality output.

For the image domain, the sampling process starts with a Gaussian noise $x_T \sim \mathcal{N}(0, I)$ and generates the metal-reduced image X_{MAR} by the conditional reverse process. The condition here is also channel-wise concatenation of X_{MA}, X_{prior} and X_{fbp}. But different to the C_x in the training phase, the prior image X_{prior} is the FBP output of the predicted clean sinogram S_{pred} from the testing stage of the sinogram model: $X_{prior} = \text{FBP}(S_{pred})$. Given x_t, the condition $C_x = [X_{MA}, X_{fbp}, X_{prior}]$ and the trained image denoising network f_{θ_x}, the intermediate image x_{t-1} is derived as $x_{t-1} = \frac{\sqrt{\bar{\alpha}_{t-1}^{(x)}}(1-\alpha_t^{(x)})}{1-\bar{\alpha}_t^{(x)}} f_{\theta_x}(x_t, t, C_x) + \frac{\sqrt{\alpha_t^{(x)}}(1-\bar{\alpha}_{t-1}^{(x)})}{1-\bar{\alpha}_t^{(x)}} x_t + \sigma_t z$, where $z \sim \mathcal{N}(0, I)$ and $\sigma_t = \sqrt{\frac{1-\bar{\alpha}_{t-1}^{(x)}}{1-\bar{\alpha}_t^{(x)}}(1-\alpha_t^{(x)})}$. Iteratively, the metal-reduced CT image X_{MAR} can be sampled as: $X_{MAR} = x_0$.

4 Experiments and Results

4.1 Datasets and Experiment settings

Dataset: Following the simulation protocol in InDuDoNet[14], we randomly select a subset from the DeepLesion [19] to synthesize metal artifact data. The metal masks are from CNNMAR[21], which contain 100 metallic implants with different shapes and sizes. We combine 1,000 images, 800 for training and 200 for validation, with 90 metal masks to synthesize the training and validation samples. The additional 200 CT images from 12 patients are paired with the remaining 10 metal masks to generate 2,000 images for testing. The sizes of these 10 implants are [2061, 890, 881, 451, 254, 124, 118, 112, 53, 35] in pixels. Consistent with [20], each 2 adjacent sizes are grouped for MAR performance evaluation. The CT images are resized to 416 × 416 pixels, 640 projection views are uniformly spaced in 360 degrees, and thus sinograms are of size 641 × 640.

Metrics: Peak signal-to-noise ratio (PSNR) and structured similarity index (SSIM) are selected for evaluation.

Implementation Details: Based on NVIDIA RTX A4000 GPUs, we implement our network with PyTorch and differential operations in MATLAB. We adopt the AdamW optimizer. The initial learning rate is 5×10^{-4} with weight decay as 1×10^{-3} and a linear warmup scheduler followed by a cosine annealing schedule. The total epoch is 500. Similar to [14], we randomly select an image and a metal mask in each training iteration to synthesize a metal-affected sample. We refer to Diff-UNet to develop the diffusion model and UNet. The parameter number is the sum of parameter numbers in the two denoising UNet, which is around 26M. The total timestep T of both diffusion models is 1000, and β_t are

Table 1: Performance comparison of different MAR approaches on the synthesized DeepLesion[19] dataset. (PSNR(dB)/SSIM)

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[5]	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[11]	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[21]	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[7]	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[20]	34.04/0.9343	33.10/0.9362	33.37/0.9384	32.75/0.9393	32.77/0.9395	33.21/0.9375
DAN-Net[17]	30.82/0.9750	31.30/0.9796	33.39/0.9852	35.02/0.9883	43.61/0.9950	34.83/0.9846
DuDoNet++[10]	36.17/0.9784	38.34/0.9891	40.32/0.9913	41.56/0.9919	42.08/0.9921	39.69/0.9886
InDuDoNet[14]	36.74/0.9801	39.32/0.9896	41.86/0.9931	44.47/0.9942	45.01/0.9948	41.48/0.9904
DICDNet[13]	37.19/0.9853	39.53/0.9908	42.25/0.9941	44.91/0.9953	45.27/0.9958	41.83/0.9923
OSCNet[15]	37.70/0.9883	**39.88**/0.9902	42.92/0.9950	45.04/0.9958	45.45/0.9962	42.19/0.9931
Ours	**39.03/0.9903**	38.90/**0.9925**	**44.06/0.9958**	**46.53/0.9966**	**46.30/0.9966**	**42.96/0.9943**

evenly spaced numbers over $[0.0001, 0.02]$, where $\beta_1 = 0.0001$, $\beta_T = 0.02$ and $\alpha_t = 1 - \beta_t$. The denoising diffusion implicit models (DDIM) [12] sampling algorithm is employed to replace the DDPM sampling process. By skipping certain intermediate steps during the DDIM sampling process, only 50 steps are required to generate the result, which could greatly accelerate the inference procedure. The testing time is around 15 times faster than DDPM.

4.2 Performance Evaluation

We report the numerical results of the proposed model on the synthesized DeepLesion dataset in Table 1. The deep learning-based methods achieve better performance than the conventional methods like LI and NMAR. Meanwhile, compared to all the SOTAs, our model achieves the highest SSIM score across metals of different sizes, as well as the highest PSNR on average. This suggests that the diffusion model possesses the inherent capability to outperform conventional methods and other baseline deep learning models and achieve superior results.

Figure 2 depicts the reconstruction results of different models. A large portion of the output from LI and CNNMAR lacks smoothness, whereas DSCMAR, InDuDoNet and OSCNet fail to completely eliminate the artifacts between the metal implants. The proposed DCDiff exhibits less shading and streak artifacts, and reconstructs the gap between different organs, even though they are occluded by the metal artifact.

4.3 Ablation Study

To further evaluate the effectiveness of different components in our methods, we first compare our diffusion interpolation (DI) method with the intuitive sinogram diffusion model without DI to justify the superiority of our approach and then investigate the performance of different conditions in the image denoising network. Both experiments are conducted on the synthesized dataset.

Effectiveness of Diffusion Interpolation: We establish a sinogram diffusion model that performs Filtered Back Projection (FBP) on the sinogram

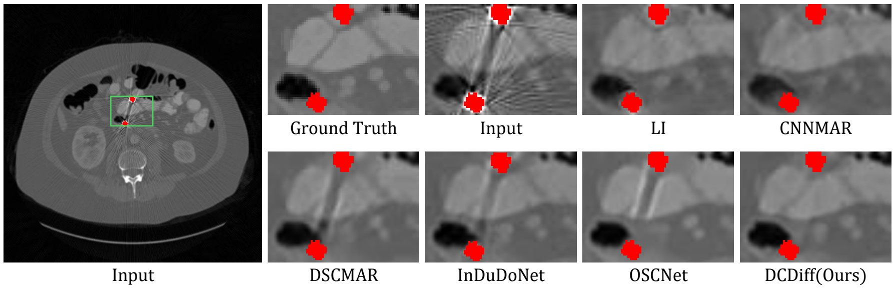

Fig. 2: Visual results comparison of different models on the synthesized DeepLesion[19] dataset. The red pixels denote metal implants.

output directly to generate a clean metal-free image without the proposed DI method, yielding a PSNR/SSIM of 31.59 dB/0.9124. In contrast, applying FBP on the final sinogram output by the model with DI, the PSNR/SSIM performance could significantly improve to 35.47 dB/0.9390, demonstrating the effectiveness of the proposed DI method.

Features concatenated in the condition We also evaluate the effect of different components, i.e., the original image X_{ma}, the FBP result of the metal trace X_{fbp}, and the FBP output of the sinogram diffusion model X_{prior}, as conditions of the image diffusion model. The results are listed in the Table 2. Compared to the most straightforward condition of X_{ma} only, either treating X_{fbp} or X_{prior} as another addition would increase PSNR/SSIM to 41.41 dB/0.9913 or 42.11 dB/0.9930, indicating the sinogram domain embeds essential information for metal artifact removal. Further, incorporating the three conditions together could increase the PSNR/SSIM to 42.96 dB/0.9943.

Table 2: Effect of different features of the image diffusion model condition.

Condition			PSNR (dB)	SSIM
X_{prior}	X_{ma}	X_{fbp}		
✗	✓	✗	40.78	0.9895
✗	✓	✓	41.41	0.9913
✓	✓	✗	42.11	0.9930
✓	✓	✓	**42.96**	**0.9943**

5 Conclusion

This paper presents DCDiff, a pioneer study based on diffusion models for metal artifact reduction of CT images. Our DCDiff framework absorbs dual-domain information as the input conditions for generating metal-free images, incorporating the raw CT image and the filtered back project image of metal trace from the image domain, and the prior image obtained via a new diffusion interpolation algorithm from the sinogram domain. Extensive experiments on the DeepLesion dataset indicate the superior performance of our method. We believe that our work holds significant implications in exploring the application of deep generative models in the field of MAR.

Acknowledgments. This work is supported by the CurveBeam AI - Monash University PhD Scholarship.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Choi, Y., Kwon, D., Baek, S.J.: Dual domain diffusion guidance for 3d cbct metal artifact reduction. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7965–7974 (2024)
2. Dhariwal, P., Nichol, A.: Diffusion models beat gans on image synthesis. Advances in Neural Information Processing Systems **34**, 8780–8794 (2021)
3. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial networks. Communications of the ACM **63**(11), 139–144 (2020)
4. Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems **33**, 6840–6851 (2020)
5. Kalender, W.A., Hebel, R., Ebersberger, J.: Reduction of ct artifacts caused by metallic implants. Radiology **164**(2), 576–577 (1987)
6. Liao, H., Lin, W.A., Zhou, S.K., Luo, J.: Adn: Artifact disentanglement network for unsupervised metal artifact reduction. IEEE Transactions on Medical Imaging **39**(3), 634–643 (2019)
7. Lin, W.A., Liao, H., Peng, C., Sun, X., Zhang, J., Luo, J., Chellappa, R., Zhou, S.K.: Dudonet: Dual domain network for ct metal artifact reduction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10512–10521 (2019)
8. Liu, X., Xie, Y., Diao, S., Tan, S., Liang, X.: Unsupervised ct metal artifact reduction by plugging diffusion priors in dual domains. IEEE Transactions on Medical Imaging (2024)
9. Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: Repaint: Inpainting using denoising diffusion probabilistic models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11461–11471 (2022)
10. Lyu, Y., Lin, W.A., Liao, H., Lu, J., Zhou, S.K.: Encoding metal mask projection for metal artifact reduction in computed tomography. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part II 23. pp. 147–157. Springer (2020)
11. Meyer, E., Raupach, R., Lell, M., Schmidt, B., Kachelrieß, M.: Normalized metal artifact reduction (nmar) in computed tomography. Medical physics **37**(10), 5482–5493 (2010)

12. Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)
13. Wang, H., Li, Y., He, N., Ma, K., Meng, D., Zheng, Y.: Dicdnet: deep interpretable convolutional dictionary network for metal artifact reduction in ct images. IEEE Transactions on Medical Imaging **41**(4), 869–880 (2021)
14. Wang, H., Li, Y., Zhang, H., Chen, J., Ma, K., Meng, D., Zheng, Y.: 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. pp. 107–118. Springer (2021)
15. Wang, H., Xie, Q., Li, Y., Huang, Y., Meng, D., Zheng, Y.: Orientation-shared convolution representation for ct metal artifact learning. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 665–675. Springer (2022)
16. Wang, J., Zhao, Y., Noble, J.H., Dawant, B.M.: Conditional generative adversarial networks for metal artifact reduction in ct images of the ear. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I. pp. 3–11. Springer (2018)
17. Wang, T., Xia, W., Huang, Y., Sun, H., Liu, Y., Chen, H., Zhou, J., Zhang, Y.: Dan-net: Dual-domain adaptive-scaling non-local network for ct metal artifact reduction. Physics in Medicine & Biology **66**(15), 155009 (2021)
18. Xing, Z., Wan, L., Fu, H., Yang, G., Zhu, L.: Diff-unet: A diffusion embedded network for volumetric segmentation. arXiv preprint arXiv:2303.10326 (2023)
19. Yan, K., Wang, X., Lu, L., Summers, R.M.: Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. Journal of medical imaging **5**(3), 036501–036501 (2018)
20. Yu, L., Zhang, Z., Li, X., Xing, L.: Deep sinogram completion with image prior for metal artifact reduction in ct images. IEEE transactions on medical imaging **40**(1), 228–238 (2020)
21. Zhang, Y., Yu, H.: Convolutional neural network based metal artifact reduction in x-ray computed tomography. IEEE transactions on medical imaging **37**(6), 1370–1381 (2018)