Metal Inpainting in CBCT Projections Using Score-based Generative Model

Siyuan Mei¹, Fuxin Fan¹*, Andreas Maier¹

¹ Pattern Recognition Lab, Friedrich-Alexander-University Erlangen-Nuremberg, Germany

ABSTRACT

During the orthopaedic surgery, the insertion of metallic implants or screws is often performed under mobile C-arm systems. However, due to the high attenuation of metals, severe metal artifacts occur in 3D reconstructions, which degrade the image quality significantly. Therefore, many metal artifact reduction algorithms have been developed to reduce the artifacts, and metal inpainting in the projection domain is an essential step. In this work, a score-based generative model is trained on simulated knee projections, and the inpainted images are obtained by removing the noise in the conditional resampling process. The result implies that the inpainted images by the score-based generative model have more detailed information and achieve the lowest mean absolute error of 0.069 and the highest peak-signal-to-noise ratio of 43.07 compared with the inverse distance weighting interpolation method and the mask pyramid network. Besides, the score-based model can also recover projections with large circular and rectangular masks, showing its generalization in inpainting tasks.

* Email: fuxin.fan@fau.de

Index Terms— Knee projection inpainting, score-based generative model, unsupervised learning

1. INTRODUCTION

A Mobile C-arm X-ray system is a medical imaging device frequently used in current interventional surgery, which supports the accurate placement of metallic implants or screws. However, some physical effects like photon starvation and beam hardening occur when x-rays pass through metals, resulting in the bright and dark streak artifacts, which reduce the quality of the reconstructed image [1]. The conventional metal artifacts reduction (MAR) algorithms tackle the problem by sinogram completion [2, 3] and iterative reconstruction [4, 5]. Deep learning-based methods are also applied in MAR [6, 7, 8, 9, 10], and the models are trained on paired data in a supervised way. For the C-arm system, since only the central slice in cone-beam geometry can be represented as a sinogram and other artifacts like truncation artifacts also exist in reconstructions, image inpainting in the projection domain is the only feasible method for MAR of cone-beam computed tomography (CBCT).

Score-based generative models are applied in computer vision recently [11, 12], and with the mechanism of stepwise noise perturbation and removal, they have shown superiority over generative adversarial networks on image synthesis [13]. Such models are successfully applied in image inpainting task [11, 14]. In Ref. [14], a score-based generative model called RePaint is trained, and it can generate restored images with high fidelity under different masks. Besides, score-based generative models are used in medical imaging processing, such as CT and MRI reconstruction [15, 16]. Inspired by the research above, a score-based model is trained on knee projections, and this is the first study to apply such model in metal inpainting in CBCT projections.

2. MATERIALS AND METHOD

2.1. Score-based generative models

Score-based generative models through stochastic differential equation (SDE), which can explicitly learn the score of the probability distribution of the noisy data set, consist of three steps: perturbation process, reverse process and resampling. Depending on the way of variance changes during the perturbation process, SDEs are classified into variance exploding (VE), variance preserving (VP), and sub-VP SDE [11].

In this paper, VE-SDE is used. The continuous noise scale equation is

$$\sigma(t) = \sigma_{\text{min}}(\frac{\sigma_{\text{max}}}{\sigma_{\text{min}}})^t, \quad t \in [0, 1].$$

The forward SDE has the form of

$$f(t) = 0, \quad g(t) = \sigma(t) \sqrt{2 \log \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}}},$$

$$dx_f = f(t)dt + g(t)dw,$$

where $f(t)$ is the drift coefficient, $g(t)$ is the diffusion coefficient and the $w$ denotes a standard Wiener process. The associated reverse SDE is formulated as

$$dx_r = \{f(t) - g(t)^2 \nabla_x log p_t(x_t)\}dt + g(t)dw,$$

where $\nabla_x log p_t(x_t)$ is the score function of $p_t(x_t)$, $dt$ is the infinitesimal negative time step, and $\bar{w}$ denotes the reverse Wiener process.
Then the forward diffusion process $x_{t+\Delta t}$ can be obtained from the original data $x_0$ by

$$x_{t+\Delta t} = x_t + dx_f \approx x_0 + \sigma(t)z, \quad \Delta t = \frac{1}{N},$$

where $\Delta t$ denotes the small time step size for $N$ discretizations and $z$ is a normal distribution. By starting the reverse process from sample $x_1 \sim \sigma(1)z$, the new sample can be updated by

$$x_{t-\Delta t} = x_t + dx_r \approx x_t + g(t)^2 s_\theta(x, t)\Delta t + g(t)z,$$

where $s_\theta(x, t)$ denotes the score estimated by training a score-based neural network.

For projection inpainting, the missing pixels in the metal area should be restored, and the pixels in the background serve as conditional information in the resampling process. The pipeline for projection inpainting is shown in Fig. 1. We denote the projection to be inpainted as $y$ and the binary metal mask with ones in the background as $m$. $y_t$ is obtained by forward SDE as Eq. 5 and has the form of

$$y_t = y + \sigma(t)z,$$

providing the background information. $x'_t$ follows the process of reverse SDE as Eq. 6

$$x'_t = x_{t+\Delta t} + dx_r,$$

and it predicts the inpainted pixels. Then the restored projection $x_t$ at time point $t$ can be obtained as

$$x_t = y_t \odot m + x'_t \odot (1-m),$$

where $\odot$ means pixel-wise multiplication.

Hence, the actual score we use during this conditional sampling process changes into the estimated score of the posterior probability $s_\theta(x, t) \approx \nabla_x \log p_t(x_t | y_t)$, because the trained score-based model estimates the score of $x_t$, which is conditioned on $y_t$.

2.2. Data generation

The knee CT volumes are selected from the whole body CT volumes from the SICAS medical image repository [17]. In total, there are 50 volumes with a single leg for each, and all volumes are rescaled to a voxel side length of 0.5 mm. All the volumes are forward projected to generate projections by CONRAD [18]. 60 projections are generated for each volume, with angular difference of $6^\circ$ between adjacent projections. 2700 of them are used for model training and 300 projections are used for test. Some implants like K-wires, screws, plates with holes are randomly selected and placed in different 3D volumes, and the metal masks are obtained by forward projecting these multi-metal volumes.

2.3. Network structures and training

Since the mask pyramid network (MPN) from Ref. [7] has good performance in projection inpainting, its results are used for comparison. The projection to be inpainted and the corresponding metal mask are the inputs for MPN, both of which have the size of $256 \times 256$. According to Ref. [11], the network structure for the score-based generative model has high flexibility. As shown in Fig. 2, the structure of the MPN without downsampling layers from metal mask serves as the backbone for the score-based generative model in this
work. Besides, the Gaussian random features [19] are generated at time step \( t \) and these features are summed up with the corresponding blocks from the network, which are labeled in red in Fig. 2.

In the training process, the objective function for the score-based model is the simplified denoising score matching [20]:

\[
\theta^* = \arg\min_\theta E_t \left\{ \frac{1}{2} E_{x \sim p_{data}, z \sim N(0, I)} \left[ \| s_\theta(x, t) + \frac{z}{\sigma(t)} \|_2^2 \right] \right\}.
\] (10)

For the defined VE SDE in this work, the minimum and maximum variances are 0.01 and 128, respectively. During the re-sampling process, the image of Gaussian noise with the variance of \( \sigma_{max} \) is the initializer, and the inpainted projection is generated by the predictor-corrector (PC) sampler [11] for 1000 steps. Other training strategies for both models keep the same; we use a batch size of 64 and the Adam optimizer [21] with learning rate \( 2 \times 10^{-4} \). All experiments are conducted on Nvidia A100 GPU, and we find that both models reach convergence at about the same rate.

2.4. Hyperparameter optimization

Some hyperparameters in the sampling process are considered, including the signal-to-noise ratio (SNR) \( \eta \) for the corrector and the number of discretization steps \( N \) for reverse-time SDE. The SNR determines the step size \( \epsilon \) in Langevin dynamics, and \( N \) corresponds to the noise scales. The quantitative results for 300 projections with metal masks under ancestral sampling predictor and Langevin corrector are listed in Tab. 1. It demonstrates that the score-based generative model with \( \eta \) of 0.4 (marked by star in Tab. 1) can achieve better performance than others. Time consumption is an important issue in clinical applications. According to Tab. 1, \( N \) is proportional to time usage and it has limited performance improvement from 1000 to 2000 steps. Therefore, the parameters in our experiments are chosen as \( \eta = 0.4 \), \( N=1000 \).

3. RESULTS AND DISCUSSION

In this work, result evaluation is compared among inpainted projections generated by the inverse distance weighting interpolation with 4 closest points and the MPN and the score-based generative model. Three representative inpainted projections with different metal masks are shown in Fig. 3. From Fig. 3 (a1) to (c1), more pixels are missing. The interpolation method relies on the intensity of existing pixels outside the metal area. Therefore, the inpainted projections by interpolation method have no semantic connections to the bones or soft tissue, as shown in Fig. 3 (a2)-(c2). From the perspective of semantic performance, the inpainted projections by the MPN can restore more details in the metal area. However, when the size of the metal area increases, the prediction gets blurred, which can be observed inside the red boxes in Fig. 3 (c3). In the same region of Fig. 3(c4), the inpainted projection by the score-based generative model has more detailed information. The quantitative evaluation results for all 300 projections with metal masks are shown in Tab. 2. The score-based generative model has the lowest mean absolute error (MAE) of 0.069 in metal regions and the highest mean peak-signal-to-noise ratio (PSNR) of 43.07 dB. Meanwhile, we back-project these inpainted projections using the CONRAD[18] reconstruction algorithm and evaluate the reconstructed in the reconstruction domain. The score-based model also performs best in the corresponding reconstructions.

Except for the metal masks, some circular and rectangular masks are also generated to test the generalization of these methods. As shown in Fig. 3 (d1)-(f1), a circle, horizontal and vertical rectangles mask out the leg projections. The results of...
Table 1. Hyperparameter sweep on SNR and sampling steps under same PC sampler.

| Metric     | $N=500$ | $N=1000$ | $N=2000$ |
|------------|---------|----------|----------|
| $\eta=0.20$ | 0.0935  | 0.0789   | 0.0736   |
| $\eta=0.40^*$| 0.0749  | 0.0694   | 0.0674   |
| $\eta=0.60$ | 0.0787  | 0.0824   | 0.0958   |
| time per slice(s) | 2.50    | 4.38     | 9.79     |

Table 2. Quantitative results comparison.

| Domain   | Interpolation | MPN | Score-based |
|----------|---------------|-----|-------------|
| Metric   | MAE (dB)      | MAE (dB) | MAE (dB)    |
| Metal mask | 0.147        | 0.121       | 0.069       |
| Circle   | 0.328        | 0.176       | 0.120       |
| Horizontal rectangle | 0.064 | 0.086 | 0.044 |
| Vertical rectangle | 0.299 | 0.268 | 0.164 |
| Reconstruction | Metal mask | 21.02 HU | 20.13 HU | 12.34 HU |

4. CONCLUSION

This work applies the score-based generative model in metal inpainting for knee CBCT projections. By predicting more detailed information under metal masks, the proposed unsupervised method has the best performance, which is supposed to benefit MAR algorithms. Moreover, the score-based generative model is able to restore the knee projections when faced with much more enormous circular and rectangular masks, showing its robustness in the CBCT projection inpainting tasks.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human subject data made available in open access by the SICAS medical image repository [17]. Ethical approval was not required as confirmed by the license attached with the open access data.

6. ACKNOWLEDGMENTS

The authors gratefully acknowledge the scientific support and High Performance Computing (HPC) resources provided by the Erlangen National High Performance Computing Center (NHR@FAU) of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). The hardware is funded by the German Research Foundation (DFG).

7. REFERENCES

[1] Masaki Katsura, Jiro Sato, Masaaki Akahane, Akira Kunimatsu, Osamu Abe, et al., “Current and novel techniques for metal artifact reduction at ct: practical guide
for radiologists,” *Radiographics*, vol. 38, no. 2, pp. 450–461, 2018.

[2] Esther Meyer, Rainer Raupach, Michael Lell, Bernhard Schmidt, and Marc Kachelrieß, “Normalized metal artifact reduction (nmart) in computed tomography,” *Med Phys*, vol. 37, no. 10, pp. 5482–5493, 2010.

[3] Esther Meyer, Rainer Raupach, Michael Lell, Bernhard Schmidt, and Marc Kachelrieß, “Frequency split metal artifact reduction (fsmar) in computed tomography,” *Med Phys*, vol. 39, no. 4, pp. 1904–1916, 2012.

[4] Ge Wang, Donald L Snyder, Joseph A O’Sullivan, and Michael W Vannier, “Iterative deblurring for ct metal artifact reduction,” *IEEE Trans Med Imaging*, vol. 15, no. 5, pp. 657–664, 1996.

[5] Xiaomeng Zhang, Jing Wang, and Lei Xing, “Metal artifact reduction in x-ray computed tomography (ct) by constrained optimization,” *Med Phys*, vol. 38, no. 2, pp. 701–711, 2011.

[6] Yanbo Zhang and Hengyong Yu, “Convolutional neural network based metal artifact reduction in x-ray computed tomography,” *IEEE Trans Med Imaging*, vol. 37, no. 6, pp. 1370–1381, 2018.

[7] 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 *Proc. MICCAI*. Springer, 2019, pp. 77–85.

[8] Tristan M Gottschalk, Björn W Kreher, Holger Kunze, and Andreas Maier, “Deep learning based metal inpainting in the projection domain: Initial results,” in *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer, 2019, pp. 125–136.

[9] Tao Wang, Wenjun Xia, Yongqiang Huang, Huaiqiang Sun, Yan Liu, Hu Chen, Jiliu Zhou, and Yi Zhang, “Dual-domain adaptive-scaling non-local network for ct metal artifact reduction,” in *Proc. MICCAI*. Springer, 2021, pp. 243–253.

[10] 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 Trans Med Imaging*, vol. 41, no. 4, pp. 869–880, 2021.

[11] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole, “Score-based generative modeling through stochastic differential equations,” in *ICLR*, 2020.

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel, “Denoising diffusion probabilistic models,” *Adv Neural Inf Process Syst.*, vol. 33, pp. 6840–6851, 2020.

[13] Prafulla Dhariwal and Alexander Nichol, “Diffusion models beat gans on image synthesis,” *Adv Neural Inf Process Syst.*, vol. 34, pp. 8780–8794, 2021.

[14] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool, “Repaint: Inpainting using denoising diffusion probabilistic models,” in *Proc. CVPR*, 2022, pp. 11461–11471.

[15] Yang Song, Liyue Shen, Lei Xing, and Stefano Ermon, “Solving inverse problems in medical imaging with score-based generative models,” in *ICLR*, 2021.

[16] Hyungjin Chung and Jong Chul Ye, “Score-based diffusion models for accelerated mri,” *Med Image Anal.*, p. 102479, 2022.

[17] Michael Kistler, Serena Bonaretti, Marcel Pfahrer, Roman Niklaus, Philippe Büchler, et al., “The virtual skeleton database: an open access repository for biomedical research and collaboration,” *J Med Internet Res.*, vol. 15, no. 11, pp. e2930, 2013.

[18] Andreas Maier, Hannes G Hofmann, Martin Berger, Peter Fischer, Chris Schwemmer, Haibo Wu, Kerstin Müller, Joachim Hornegger, Jang-Hwan Choi, Christian Riess, et al., “CONRAD—a software framework for cone-beam imaging in radiology,” *Med Phys.*, vol. 40, no. 11, pp. 111914, 2013.

[19] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng, “Fourier features let networks learn high frequency functions in low dimensional domains,” *Adv Neural Inf Process Syst.*, vol. 33, pp. 7537–7547, 2020.

[20] Pascal Vincent, “A connection between score matching and denoising autoencoders,” *Neural computation*, vol. 23, no. 7, pp. 1661–1674, 2011.

[21] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.