Large-Volume Conditioned 3D Latent Diffusion Models for CT Metal Artifact Suppression

Xabier Moreno Casado[1,2][0009−0009−2532−3504], Jef Vandemeulebroucke[1,2,3][0000−0001−5714−3254], and Jakub Ceranka[1,2][0000−0002−0241−7737]

[1] Department of Electronics and Informatics (ETRO.RDI), Vrije Universiteit Brussel (VUB), Pleinlaan 2, 1050 Brussels, Belgium
[2] imec, Kapeldreef 75, 3001 Leuven, Belgium
[3] Department of Radiology, Universitair Ziekenhuis Brussel (UZ Brussel), Vrije Universiteit Brussel (VUB), Laarbeeklaan 101, 1090 Brussels, Belgium

Abstract. Metal artifacts in computed tomography (CT) significantly degrade anatomical visibility and can compromise downstream image analysis tasks, including segmentation, registration, implant migration assessment, and image-guided surgical planning. Many effective metal artifact reduction methods use projection-domain or dual-domain information, but raw sinograms are often unavailable in retrospective or public clinical CT datasets. In this work, we present the first 3D image-domain framework for large-volume CT metal artifact suppression using conditional latent diffusion models (LDM). First, a VQ-VAE-GAN compresses CT volumes to latent representations. Secondly, a 3D conditioned diffusion U-Net operating in a latent space on crops of $448 \times 448 \times 256$ suppresses CT metal artifacts by synthesizing non-artifacted versions of the input. We compare two conditional models: an anatomy conditioned LDM and an anatomy-metadata conditioned LDM with extra information on anatomical region, laterality, and implant material. Training is performed on paired synthetic data generated by an anatomy-aware metal artifact generation pipeline that inserts implant masks into clean CT volumes, simulates polychromatic projection artifacts, and reconstructs artifact-corrupted volumes with paired non-artifacted image. On 50 test volumes, the anatomy-metadata conditioned model achieved the strongest structural and perceptual metrics (RMSE 0.021, PSNR 39.430 dB, SSIM 0.992, LPIPS 0.028), while anatomy-only conditioning showed slightly stronger visual artifact suppression. On postoperative real CTs with metal artifacts, the method successfully reduced visible artifacts in qualitative examples. The results support large-volume 3D image-domain LDMs as a promising direction for CT metal artifact suppression.

Keywords: Metal Artifact Suppression · Computed Tomography · Latent Diffusion Models · Large-Volume 3D Synthesis · Image-Domain

Corresponding author: xabier.moreno.casado@vub.be

1 Introduction

The clinical prevalence of computed tomography (CT) images with metal artifacts is increasing year to year [1]. Metallic implants in CT, such as hip prostheses, spinal screws, plates, clips, and dental metal objects, produce streaking, shading, cupping, and apparent implant enlargement through beam hardening, photon starvation, scatter, and reconstruction nonlinearities [2]. These artifacts obscure implant-adjacent anatomy and reduce the reliability of downstream image computing and computer-assisted intervention workflows, affecting both human interpretation and automated processing [3].

Metal artifact reduction (MAR) methods are commonly grouped into image-domain, projection-domain, and dual-domain approaches [3]. Image-domain methods restore already reconstructed artifact-corrupted CT images; projection-domain methods correct metal traces before reconstruction; and dual-domain methods combine projection and image-domain information.

A classic projection-domain example is NMAR [4], which replaces corrupted metal traces using interpolation after normalization with a prior image. More recent projection-domain methods use denoising diffusion probabilistic models (DDPMs) to inpaint missing sinogram data [5]. These methods can improve reconstruction quality, but they require accurate metal-trace identification as errors in the trace or sinogram completion can propagate to secondary artifacts.

Dual-domain methods combine sinogram and image information. CNN-MAR [6] uses synthetic metal insertion and CNN-based priors for MAR, while InDuDoNet+ [7] optimizes sinogram and image-domain corrections through physics-informed iterative updates. Recent diffusion methods such as DCDiff [8] and Dual Domain Diffusion Guidance [9] further exploited learned generative priors while maintaining projection-domain constraints. A practical limitation of dual-domain methods is that they require sinograms or scanner-geometry information, which is often unavailable in retrospective clinical datasets. Training with both sinogram and image data also increases computational cost [9].

Image-domain diffusion methods avoid sinogram dependence and can be applied retrospectively to reconstructed volumes. DiffMAR [10] learns an image-domain diffusion restoration process for CT MAR; however, pixel-space DDPMs remain computationally expensive, slow at inference, and may leave fine residual streak artifacts.

Latent diffusion models (LDMs) reduce diffusion cost by denoising compressed image representations rather than full-resolution voxels [11]. MAISI [12] and MedLoRD [13] showed that high-dimensional 3D CT synthesis up to $512 \times 512 \times 768$ voxels can be performed with feasible resources by combining image compression and 3D latent diffusion. Since diffusion models have shown image synthesis quality competitive with or superior to GANs [14], applying this generative prior to MAR may help achieve strong artifact suppression and texture preservation compared with earlier CNN/GAN-based approaches [8].

LDMs have recently been used for MAR: Choi *et al.* [15] proposed a conditional LDM for dental CBCT MAR. However, this approach is limited to 2D slice-wise CBCT images and does not exploit 3D volumetric context.

In this work, we propose the first large-volume 3D image-domain latent diffusion framework for CT metal artifact suppression using low-resource latent backbone. We adapt the principle of conditioned CT synthesis to paired image-domain MAR, targeting large 3D volumes around $448 \times 448 \times 256$ voxels. The method was trained on a synthetic paired artifact-generation pipeline with anatomy-aware implant insertion, full 3D projection simulation, and paired non-corrupted ground truth images; and evaluated on paired synthetic data and qualitative real-case testing. The implementation, synthetic data generation framework and trained models are available at: `https://github.com/ETRO-MIT/3D-LDMs-for-CT-MAR`.

2 Methods

2.1 Data and Synthetic Artifact Generation

The dataset contains 867 3D-CT volumes from five open-source datasets: CLINIC ($n = 103$) and CLINIC-metal ($n = 75$) from CTPelvic1K [16], KiTS19 ($n = 299$) [17], and the liver ($n = 200$) and colon ($n = 190$) tasks of the Medical Segmentation Decathlon (MSD) [18]. CLINIC, KiTS19, MSD liver, and MSD colon contain clean volumes for synthetic paired data generation; CLINIC-metal contains 75 real postoperative CTs with metal implants for qualitative validation only. Volumes were reoriented to RAS, resampled to 1 mm isotropic spacing, and stored as NIfTI.

Paired and spatially aligned clinical CT volumes acquired before and after implant placement are generally unavailable. We therefore generate supervised pairs synthetically. `TotalSegmentator` is used to obtain organ and bone masks for anatomy-aware implant placement [19]. An implant library of 29 binary masks is built from high-intensity components extracted from real CLINIC-metal volumes using a 2500 HU threshold and from voxelized public STL meshes. The library contains hip implants ($n = 11$), pelvic screws ($n = 6$), plates ($n = 3$), spine implants ($n = 2$), and spine screws ($n = 7$).

For each clean volume, a region-plausible implant is inserted. Titanium and stainless steel are simulated with 3000 HU and 4000 HU values, respectively. The implant-only ground truth is created by replacing voxels inside the metal mask while leaving acquisition artifacts absent. Artifacted images are synthesized using a 3D ASTRA [20] cone-beam projection and reconstruction workflow inspired by CNN-MAR [6]. Water, bone, and metal components are forward-projected, a polychromatic Beer-Lambert signal is simulated, Poisson noise is added, and the projection data are reconstructed. The metal-induced reconstruction difference is added to the clean CT inside the body mask. Each sample stores the artifacted CT, implant-only ground truth, binary metal mask, and JSON file with metadata. The supervised paired MAR set contains 668 training pairs, 74 validation pairs, and 50 held-out test pairs.

`https://github.com/wasserth/totalsegmentator`
`https://grabcad.com/library/tag/hip`

2.2 Large-Volume Latent Diffusion

The proposed framework follows the MedLoRD two-stage design [13] (see Fig. 1). First, a VQ-VAE-GAN encodes CT volumes into a compact latent representation and decodes restored latents back to image space. Intensities are clipped to $[-1000, 4000]$ HU and scaled to $[-1, 1]$. The VQ-VAE-GAN is trained on 128^3 patches with random small-angle rotations and flips. Its encoder has two 3D convolutional downsampling levels with 128 and 256 channels, followed by residual layers. The spatial downsampling factor is 4, producing an $8 \times 32 \times 32 \times 32$ latent for a 128^3 patch, and a codebook of 16,384 embeddings of dimension 8.

Second, LDMs are trained in the VQ-VAE-GAN latent space. For a target latent z_0, cosine-scheduled diffusion samples

$$z_t = \sqrt{\bar{\alpha}_t} z_0 + \sqrt{1 - \bar{\alpha}_t}\epsilon, \tag{1}$$

where z_t is the noisy latent at diffusion step t, $\epsilon \sim \mathcal{N}(0, I)$, and $\bar{\alpha}_t$ is the cumulative signal coefficient of the cosine noise schedule. The 3D U-Net predicts the velocity target

$$v_t = \sqrt{\bar{\alpha}_t}\epsilon - \sqrt{1 - \bar{\alpha}_t} z_0, \tag{2}$$

using a Huber loss [21],

$$\mathcal{L}_{LDM} = \text{Huber}\left(v_\theta(z_t, c, t), v_t\right). \tag{3}$$

The condition c denotes the artifacted-image latent and, for the metadata-conditioned model, the metadata cross-attention context.

The denoiser uses feature channels 64, 128, 256, and 512 with attention in the two deepest levels. T=500 steps was selected for the conditional models based on validation performance. Training and inference use crops of $448 \times 448 \times 256$ voxels.

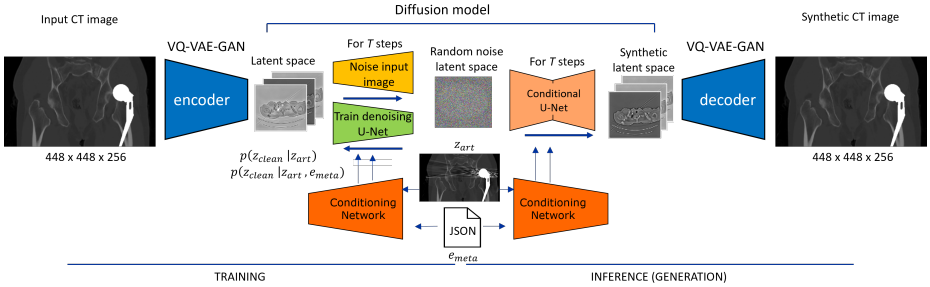

Fig. 1. Architecture of the proposed conditional large-volume 3D latent diffusion for image-domain MAR. The VQ-VAE-GAN encoder compresses CT volumes into latent space, where the conditional U-Net is trained to denoise target latents. In inference, the denoised latent is decoded back to the CT image space. Both conditioning variants use the artifacted CT image; the metadata-conditioned model additionally injects implant/anatomy information.

2.3 Conditioning Strategies

The two proposed conditioning strategies (see Fig. 1) share the same VQ-VAE-GAN and 3D diffusion U-Net with a difference in the type of information provided to the denoiser.

Anatomy conditioned LDM. The artifacted CT is encoded as z_{art}. During each denoising step, the U-Net receives the concatenation $[z_t, z_{\text{art}}]$ and predicts v_t, learning $p(z_0 \mid z_{\text{art}})$. In the implemented configuration, this gives a 16-channel LDM input and an 8-channel latent output.

Anatomy-metadata conditioned LDM. The second model uses the same spatial condition and additionally encodes non-spatial metadata as e_{meta}: anatomical region `[unknown, spine, hip, knee, shoulder]`, side `[unknown, left, right, midline, bilateral]`, and metal name `[unknown, titanium, steel]`. Simulation parameters are excluded because they are constant across synthetic cases. Each categorical metadata value is mapped to a learned embedding and passed through a multi-layer perceptron to form the cross-attention context for the U-Net, learning $p(z_0 \mid z_{\text{art}}, e_{\text{meta}})$.

2.4 Evaluation and Computational Setup

Artifact-suppressed test cases are evaluated against implant-only ground truths using MAE, RMSE, PSNR, SSIM and LPIPS [22,23,24]. MAE and RMSE are computed on normalized $[-1, 1]$ volumes. PSNR uses dynamic range $L = 2$. SSIM and LPIPS are computed on central axial, coronal, and sagittal slices.

In addition to quantitative metrics, six reviewers with medical image processing expertise performed a blinded visual review. For each test case, reviewers inspected the artifacted input, the implant-only ground truth, and anonymized MAR model outputs in axial, coronal, and sagittal views. Conditioning strategies were randomized and decoded only after review. Each output was scored from 0 to 5 using the criteria in Table 1.

Table 1. Visual-review scoring scale for artifact suppression, anatomical preservation, and overall image quality.

Score	Artifact Reduction	Anatomy Preservation	Overall Quality
5	None residual	Preserved	Target-like
4	Small residual	Minor changes	Good
3	Moderate residual	Noticeable changes	Acceptable
2	Strong residual	Partial distortion/removal	Poor
1	Very strong residual	Severe distortion	Very poor
0	No reduction	Distorted/unreliable	Dominated by noise/blur

Experiments used Python, PyTorch, and MONAI [25] on the anonymized HPC cluster. The training and inference used a NVIDIA H200 GPU 140 GB. VQ-VAE-GAN was trained for 222 epochs with a batch size of 15, while LDM

was trained for 250 epochs with a batch size of 2. Test inference jobs averaged 3.64 min per subject for the anatomy-conditioned LDM and 3.81 min per subject for the anatomy-metadata conditioned LDM.

3 Results

Synthetic Paired Test Set As a preliminary validation of the latent compression stage, the VQ-VAE-GAN reconstruction was evaluated on the test set across the original CT, synthetic artifacted CT, and implant-only ground truth images, achieving a low combined MAE of (0.0068 ± 0.0020). The lowest error was obtained for original CT images (0.0060 ± 0.0009), while the highest error was observed for artifacted images (0.0085 ± 0.0024), supporting the use of the learned latent representation for downstream diffusion-based MAR.

Table 2 reports results on the 50 paired synthetic test cases and 300 reviewer-case ratings per model. Both conditional models improved RMSE, PSNR, SSIM and LPIPS over the artifacted input and qualitatively suppressed metal artifacts (Fig. 2). The anatomy-metadata conditioned model achieved the best RMSE, PSNR, SSIM and LPIPS. The anatomy conditioned model had slightly lower MAE among learned models, while the raw artifacted input had the lowest MAE overall. This is expected for a global voxel-wise metric because most voxels are unaffected by metal artifacts and remain identical to the target.

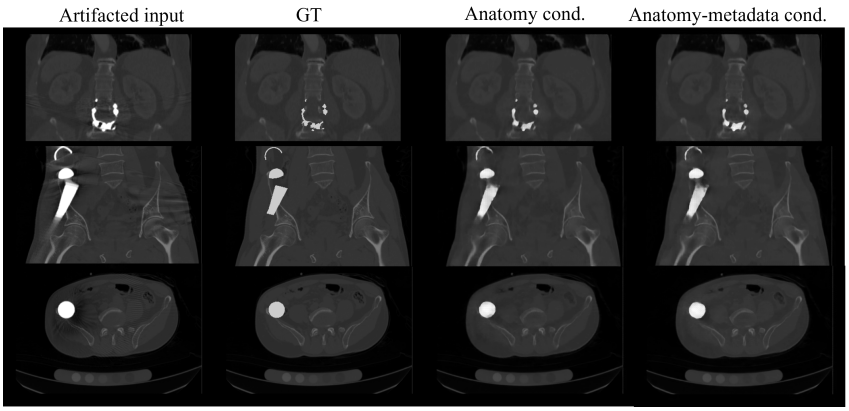

Fig. 2. Paired synthetic MAR examples, with columns showing artifacted input, implant-only target without artifacts, anatomy conditioned output, and anatomy-metadata conditioned output. Each row corresponds to a different simulated patient case.

Normality of paired differences was tested with the Shapiro–Wilk. Paired t-tests for normal quantitative comparisons, Wilcoxon signed-rank tests for non-normal quantitative comparisons and visual-review scores, and Bonferroni correction within each metric were used. Both conditional models significantly improved over the raw artifacted input for RMSE, PSNR, SSIM, and LPIPS ($p_{\mathrm{adj}} < 0.001$). In the direct comparison, the metadata-conditioned model was

Table 2. Final paired synthetic test results on 50 volumes. Quantitative metrics are median [bootstrap 95% CI]. Visual-review scores are mean ± standard deviation on a 0–5 scale; the raw artifacted input was shown to reviewers but not scored as a model output.

Metric	Raw	Anatomy	Anatomy-metadata
MAE ↓	**0.006[0.005,0.007]**	0.008[0.008,0.009]*	0.009[0.008,0.009]
RMSE ↓	0.031[0.028,0.033]	0.023[0.021,0.024]	**0.021[0.021,0.023]*****
PSNR ↑	36.29[35.62,37.20]	38.89[38.47,39.81]	**39.43[38.94,39.72]*****
SSIM ↑	0.981[0.974,0.985]	0.991[0.988,0.992]	**0.992[0.991,0.993]*****
LPIPS ↓	0.062[0.040,0.081]	0.037[0.029,0.043]	**0.028[0.025,0.034]*****
Artifact suppression ↑	–	**3.543 ± 1.107*****	3.397 ± 1.103
Anatomical preservation ↑	–	3.477 ± 0.966	**3.633 ± 0.988*****
Overall quality ↑	–	3.387 ± 1.007	**3.433 ± 0.994**$^{\text{n.s.}}$

Superscripts report the direct comparison between the two LDMs: * $p_{\text{adj}} < 0.05$, ** $p_{\text{adj}} < 0.01$, *** $p_{\text{adj}} < 0.001$, n.s. not significant after Bonferroni correction.

significantly better for RMSE, PSNR, SSIM, and LPIPS ($p_{\text{adj}} < 0.001$), while MAE favored the anatomy conditioned model ($p_{\text{adj}} = 0.012$). The visual review analysis showed a complementary trade-off: the anatomy conditioned model had slightly higher artifact suppression, whereas metadata conditioning improved anatomical preservation and slightly improved overall quality (see Fig. 3).

Real Metal Case Evaluation The 75 CLINIC-metal volumes contain real postoperative implants but no paired artifact-free targets, so they were used for qualitative evaluation only. We applied the selected metadata conditioned model to representative real metal cases (Fig. 4). The model reduces visible streaking and shading around implants, with residual artifacts near high-density metal, supporting qualitative transfer from synthetic training to clinical artifacts.

4 Discussion

We propose the first image-domain 3D LDM framework for CT metal artifact reduction. The MAR results prove that conditional latent diffusion can successfully reduce metal artifacts while preserving important surrounding anatomy (see Fig. 2). Both conditional models improved quantitatively compared with the raw artifacted input. MAE was the exception, with the raw input obtaining the lowest value, since the diffusion model is introducing small intensity changes outside the main artifact region.

The anatomy-metadata conditioned model achieved the best quantitative performance, supporting the idea that metadata provides useful context beyond the image anatomy itself. However, the visual review showed a trade-off between the two conditioning strategies. The anatomy conditioned model stronger suppresses artifacts, whereas adding metadata improved anatomical preservation (see Fig. 3). The anatomy conditioned model removes metal artifacts more aggressively but may alter fine anatomical details, while the metadata-conditioned model leaves some residual streaking but better maintains local anatomy. The

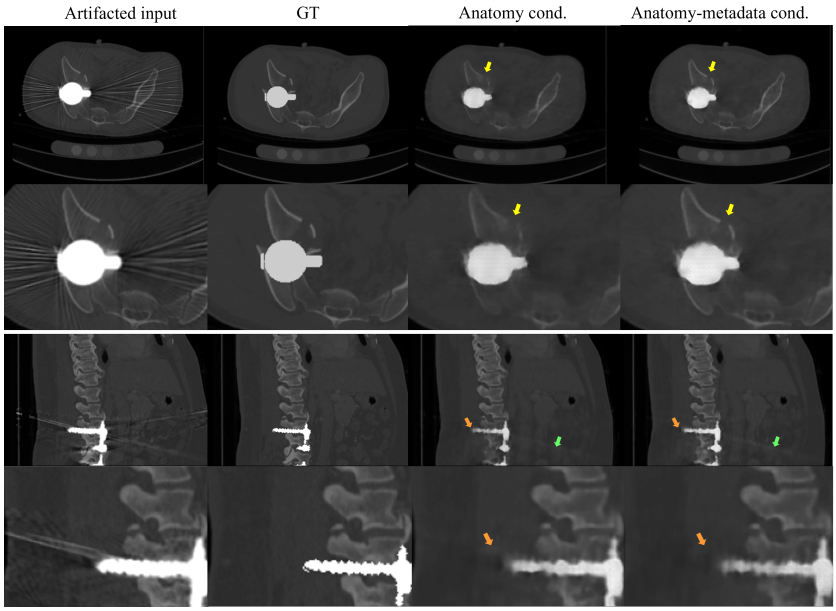

Fig. 3. Additional examples of LDM MAR result images. **Top:** CT image where better anatomical preservation of anatomy-metadata conditioning is present (yellow arrows). **Bottom:** CT image with visible remaining artifacts more evident for anatomy-metadata conditioning method (green arrows - residual streak artifacts, orange arrows - residual shading/cupping artifacts).

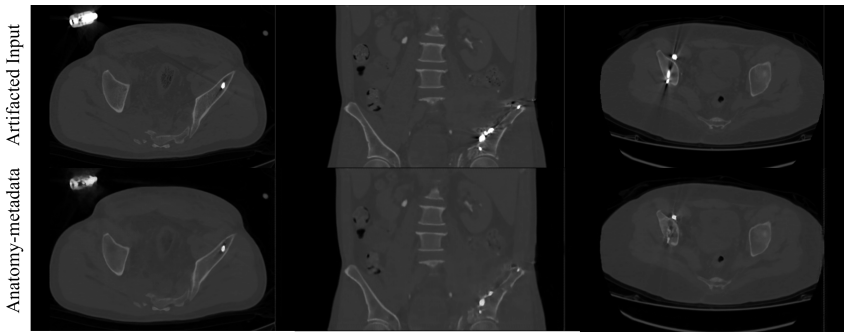

Fig. 4. Qualitative inference on three different real CLINIC-metal cases (columns). Each column compares a real artifacted CT slice (top) with the metadata-conditioned LDM output (bottom).

anatomy-metadata conditioned model was selected for real-case inference as it gave stronger overall quantitative result and better anatomical preservation.

Some residual artifacts persist, especially near high-density metal, and the proposed method can alter the implant geometry, which may complicate image interpretation. Future work can explore post-processing metal-mask constraints or anatomy-aware (metal segmentation mask) conditioning to distinguish between artifact patterns that should be removed and anatomical/implant structures that should be preserved. MAR-specific metrics, such as artifact-region or metal-adjacent ROI measures, should also be considered. To assess whether an LDM trained solely on synthetic artifacted data can generalize to real postoperative CT acquisitions, we evaluated the anatomy-metadata conditioned model qualitatively on CT volumes with real metal artifacts (see Fig. 4). The results show that the model can suppress visible artifacts in real postoperative CT images despite being trained only on synthetic paired data. Overall, this work shows the potential of image-domain 3D conditional latent diffusion for CT metal artifact reduction. By improving the readability of CT images affected by metallic implants, this approach may support future diagnostic assessment, radiological planning, and computer-assisted workflows.

Acknowledgments. This work was funded by VLAIO project HBC.2024.0696, AINIMO. Computational resources were provided by the Vlaams Supercomputer Centrum (VSC).

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

References

1. Wang, Q.F., Tang, Y.C., Liao, H.R., Lei, M., Dong, W., Liu, Z.Y., Hao, J., Hu, Z.M.: Prevalence of metal implants among US adults aged 40 years and older. Scientific Reports **15**(1), 584 (2025). https://doi.org/10.1038/s41598-024-84340-0
2. Boas, F.E., Fleischmann, D.: CT artifacts: Causes and reduction techniques. Imaging in Medicine **4**(2), 229–240 (2012). https://doi.org/10.2217/iim.12.13
3. Gjesteby, L., De Man, B., Jin, Y., Paganetti, H., Verburg, J., Giantsoudi, D., Wang, G.: Metal artifact reduction in CT: Where are we after four decades? IEEE Access **4**, 5826–5849 (2016). https://doi.org/10.1109/ACCESS.2016.2608621
4. 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). https://doi.org/10.1118/1.3484090
5. Wu, Z., Zhong, X., Lyu, T., Xi, Y., Ji, X., Zhang, Y., Xie, S., Yu, H., Chen, Y.: PRAISE-Net: Deep projection-domain data-consistent learning network for CBCT metal artifact reduction. IEEE Transactions on Instrumentation and Measurement **74**, 1–13 (2025). https://doi.org/10.1109/TIM.2025.3551446
6. 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). https://doi.org/10.1109/TMI.2018.2823083

7. Wang, H., Li, Y., Zhang, H., Meng, D., Zheng, Y.: InDuDoNet+: A deep unfolding dual domain network for metal artifact reduction in CT images. Medical Image Analysis **85**, 102729 (2023). https://doi.org/10.1016/j.media.2022.102729
8. Shen, R., Li, X., Li, Y.F., Sui, C., Peng, Y., Ke, Q.: DCDiff: Dual-domain conditional diffusion for CT metal artifact reduction. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024, Lecture Notes in Computer Science, vol. 15007, pp. 223–232. Springer (2024). https://doi.org/10.1007/978-3-031-72104-5_22
9. Choi, Y., Kwon, D., Baek, S.J.: Dual domain diffusion guidance for 3D CBCT metal artifact reduction. In: IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 7950–7959 (2024). https://doi.org/10.1109/WACV57701.2024.00778
10. Cai, T., Li, X., Zhong, C., Tang, W., Guo, J.: DiffMAR: A generalized diffusion model for metal artifact reduction in CT images. IEEE Journal of Biomedical and Health Informatics **28**(11), 6712–6724 (2024). https://doi.org/10.1109/JBHI.2024.3439729
11. Rombach, R., Blattmann, A., Lorenz, D., Esser, P., Ommer, B.: High-resolution image synthesis with latent diffusion models. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10684–10695 (2022)
12. Guo, P., Zhao, C., Yang, D., Xu, Z., Nath, V., Tang, Y., Simon, B., Belue, M., Harmon, S., Turkbey, B., et al.: MAISI: Medical AI for synthetic imaging. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4430–4441. IEEE (2025)
13. Seyfarth, M., Dar, S.U.H., Ayx, I., Fink, M.A., Schoenberg, S.O., Kauczor, H.U., Engelhardt, S.: MedLoRD: A medical low-resource diffusion model for high-resolution 3D CT image synthesis (2025). https://doi.org/10.48550/arXiv.2503.13211
14. Dhariwal, P., Nichol, A.: Diffusion models beat GANs on image synthesis. Advances in neural information processing systems **34**, 8780–8794 (2021)
15. Choi, D.i., Yun, S., Hyun, S., Cho, S.: Metal artifact reduction algorithm with conditional latent diffusion model for dental cone-beam CT. Journal of Applied Clinical Medical Physics **26**(11), e70317 (2025). https://doi.org/10.1002/acm2.70317
16. Liu, P., Han, H., Du, Y., Zhu, H., Li, Y., Gu, F., Xiao, H., Li, J., Zhao, C., Xiao, L., Wu, X., Zhou, S.K.: Deep learning to segment pelvic bones: Large-scale CT datasets and baseline models. International Journal of Computer Assisted Radiology and Surgery **16**(5), 749–756 (2021). https://doi.org/10.1007/s11548-021-02363-8
17. Heller, N., Sathianathen, N., et al.: The KiTS19 challenge data: 300 kidney tumor cases with clinical context, CT semantic segmentations, and surgical outcomes (2019)
18. Antonelli, M., Reinke, A., Bakas, S., et al.: The medical segmentation decathlon. Nature Communications **13**, 4128 (2022). https://doi.org/10.1038/s41467-022-30695-9
19. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: TotalSegmentator: Robust segmentation of 104 anatomic structures in CT images. Radiology: Artificial Intelligence **5**(5), e230024 (2023). https://doi.org/10.1148/ryai.230024
20. van Aarle, W., Palenstijn, W.J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K.J., Sijbers, J.: Fast and flexible X-ray tomography using the ASTRA toolbox. Optics Express **24**(22), 25129–25147 (2016). https://doi.org/10.1364/OE.24.025129

21. Meyer, G.P.: An alternative probabilistic interpretation of the Huber loss. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5261–5269 (2021). https://doi.org/10.48550/arXiv.1911.02088
22. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing **13**(4), 600–612 (2004). https://doi.org/10.1109/TIP.2003.819861
23. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018). https://doi.org/10.1109/CVPR.2018.00068
24. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in Neural Information Processing Systems. vol. 30 (2017)
25. Cardoso, M.J., Li, W., Brown, R., Ma, N., Kerfoot, E., Wang, Y., Murrey, B., Myronenko, A., Zhao, C., Yang, D., et al.: MONAI: An open-source framework for deep learning in healthcare. arXiv preprint arXiv:2211.02701 (2022)