MAGGNet: Guided Volumetric Metal Artefact Synthesis for Computed Tomography Scans

Jakub Drzewiecki, Igor Hołowacz, Bartosz Lewandowski, Filip Fit, and Urszula Markowska-Kaczmar

Wrocław University of Science and Technology (Politechnika Wrocławska), Wrocław, Poland
kuba.drzewiecki4@gmail.com; igor.holowacz@gmail.com;
barteklew123@gmail.com; filip.fit1302@gmail.com;
urszula.markowska-kaczmar@pwr.edu.pl

Abstract. Metal artefacts remain a major challenge in CT, especially near metallic implants, where they degrade image quality and obscure anatomy. The lack of paired CT scans with and without artefacts forces most metal artefact reduction methods (MAR) to rely on synthetic artefacts, causing domain gaps, particularly in 3D. This work introduces a StyleGAN-based data generation framework for realistic, volumetrically consistent CT metal artefact synthesis. A variational content encoder and procedurally generated metal masks guide artefact patterns, while adversarial training enforces realism and diversity. The framework rapidly produces high-quality CT volumes with controllable, anatomically plausible artefacts that remain coherent across slices. Volumetric consistency, assessed with a gradient-based cosine similarity, surpasses random artefact generation, providing an efficient data source for training and evaluating MAR methods.

Keywords: Computed Tomography · Metal Artefact Reduction · Artefact Synthesis · Deep Learning · GAN

1 Introduction

Most prior work on computed tomography (CT) artefacts has focused on metal artefact reduction (MAR), using either classical reconstruction techniques or deep learning models that map corrupted scans to clean ones [3, 5, 1]. However, supervised MAR requires paired CT scans with and without metal, which are rarely available, so training data are often generated synthetically, typically in 2D, introducing domain gaps and limiting robustness. Recent methods model artefacts directly for training and benchmarking, including unsupervised disentanglement approaches such as ADN [7] and 3D generative frameworks that emphasise volumetric consistency over slice-wise artefact injection [14, 11].

A key challenge in CT artefact synthesis is preserving volumetric consistency. Since CT data are inherently three-dimensional, metal artefacts should

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_27

vary smoothly across adjacent slices. Slice-wise generation that ignores inter-slice context can therefore introduce discontinuities, unrealistic textures, and structural misalignments.

In this work, a guided generative framework is proposed for controllable, volumetrically consistent artefact synthesis. The method is presented as a proof of concept for anatomically consistent artefacts coherent across CT slices, without yet aiming to model highly realistic metal shapes tailored to specific body parts. The key contributions of this work are:

- a volumetrically consistent artefact generation strategy that preserves inter-slice coherence across entire CT volumes;
- a controllable synthesis mechanism that disentangles anatomical content from artefact appearance using structured metal and streak guidance;
- a gradient-based metric for quantitatively assessing slice-to-slice structural consistency in generated CT volumes.

2 MAGGNet Method

MAGGNet addresses the data bottleneck by treating artefacts as controllable image components and focusing on guided artefact synthesis that preserves anatomy while modelling realistic artefact variability. This allows consistent data augmentation and the creation of surrogate paired data for MAR training, narrowing the gap between simplistic synthetic artefacts and clinical distributions.

MAGGnet Architecture. The proposed MAGGNet (Metal Artefact Guided Generation network) extends StyleGAN [6] with physically guided, artefact-aware conditioning and volumetric control to synthesise CT images with metal artefacts that respect their physical origin, spatial structure and inter-slice coherence. A procedurally generated metal-artefact guidance mask encodes implant locations and expected streak propagation, enforcing a direct correspondence between metal and artefacts.

As illustrated in Fig. 1, MAGGNet employs a variational artefact encoder E_A to learn a latent code for artefact appearance and a content encoder E_C to preserve anatomy from clean CTs and the guidance mask; the generator G_A combines these to produce anatomically faithful, artefact-corrupted images. After adversarial training, the discriminator is discarded and the model can rapidly generate diverse, controllable and volumetrically coherent metal artefacts for clean CT volumes, providing synthetic paired data for supervised MAR. Below, we shortly describe each part of the MAGGNet architecture.

E_C – *Content encoder.* Let $x_c \in [-1,1]^{1\times H\times W}$ be a clean CT slice, $g \in [0,1]^{1\times H\times W}$ the artefact guidance map and $m \in \{0,1\}^{1\times H\times W}$ a binary metal mask. We form a synthetically metallized input $x_m = \mathrm{clip}(x_c + m,\, -1,\, 1)$ and encode the concatenation $[x_m, g]$ into a bottleneck feature map

$$z_c = E_C([x_m, g]) \in \mathbb{R}^{256\times H/4\times W/4}. \tag{1}$$

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_27

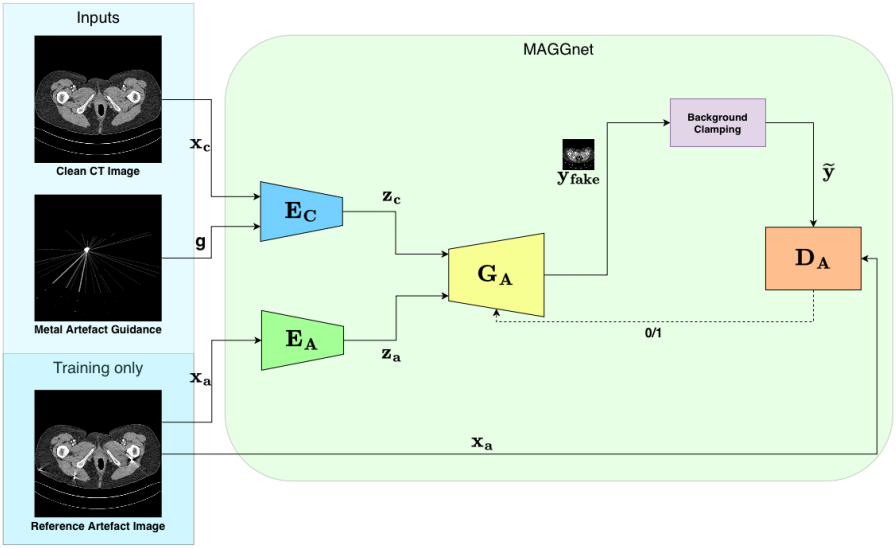

Fig. 1: **MAGGnet architecture.** G_A and D_A denote the artefact generator and discriminator, respectively. E_C is the content encoder and E_A is the variational artefact encoder, which produce latent codes z_c and z_a. x_c denotes a clean CT slice, x_a a metal–corrupted slice, and g the artefact guidance map.

E_A – *Variational artefact encoder.* Given a real artefact image $x_a \in [-1,1]^{1\times H\times W}$, E_A defines a diagonal Gaussian

$$q(z_a \mid x_a) = \mathcal{N}\big(\mu_{\text{artifact}}, \text{diag}(\sigma^2_{\text{artifact}})\big), \tag{2}$$

and samples a 128-D latent code via

$$z_a = \mu_{\text{artifact}} + \sigma_{\text{artifact}} \odot \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I). \tag{3}$$

G_A – *Artefact generator.* The generator produces an artefact-corrupted image

$$y_{\text{fake}} = G_A(z_c, z_a) \in [-1,1]^{1\times H\times W}. \tag{4}$$

Background clamping. With a body mask $b \in \{0,1\}^{1\times H\times W}$ derived from x_c, the final image is

$$\tilde{y} = b \odot y_{\text{fake}} + (1-b) \odot x_c. \tag{5}$$

MAGGnet Training. Given a clean slice x_c, guidance map g, and metal mask m, the generator produces a synthetic artefact image $y_{\text{fake}} = G_A(z_c, z_a)$ with $z_c = E_C([x_m, g])$ and $z_a \sim q(z_a \mid x_a)$, and the final image used for training is obtained as $\tilde{y}$ after background clamping. The whole model is trained in an adversarial fashion using a least-squares GAN (LSGAN) objective.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_27

The discriminator D_A learns to distinguish real artefact images from generated ones:

$$\mathcal{L}_D = \frac{1}{2}\mathbb{E}_{x_a}\big[(D_A(x_a)-1)^2\big] + \frac{1}{2}\mathbb{E}_{\tilde{y}}\big[D_A(\tilde{y})^2\big], \tag{6}$$

while the generator is trained to fool D_A:

$$\mathcal{L}_{GAN} = \mathbb{E}_{\tilde{y}}\big[(D_A(\tilde{y})-1)^2\big]. \tag{7}$$

To obtain anatomically faithful yet realistically corrupted images, the generator combines the adversarial loss with several task-specific regularizers. The total loss is

$$\begin{aligned}\mathcal{L}_G = \lambda_{GAN}\mathcal{L}_{GAN} + \lambda_{content}\mathcal{L}_{content} + \lambda_{edge}\mathcal{L}_{edge} \\ + \lambda_{bg}\mathcal{L}_{bg} + \lambda_{metal}\mathcal{L}_{metal} + \lambda_{KLD}\mathcal{L}_{KLD},\end{aligned} \tag{8}$$

where $\lambda_\cdot$ are scalar weights.

We encourage that the generated image matches the clean CT outside the guided streak regions. With body mask b and

$$c_m = b \odot \mathbb{1}[g < 0.2], \tag{9}$$

the content and background terms read

$$\mathcal{L}_{content} = \big\|c_m \odot (\tilde{y} - x_c)\big\|_1, \tag{10}$$

$$\mathcal{L}_{bg} = \big\|(1-b) \odot (\tilde{y} - x_c)\big\|_2^2. \tag{11}$$

To promote sharp, streak-like structures where guidance is present, we use an edge-based term. With Sobel operators S_x, S_y and $\nabla_x u = S_x * u$, $\nabla_y u = S_y * u$,

$$\mathcal{L}_{edge} = -\mathbb{E}\left[\sqrt{\nabla_x(\tilde{y})^2 + \nabla_y(\tilde{y})^2 + \varepsilon} \;\odot\; \mathbb{1}[g > 0.1]\right]. \tag{12}$$

We additionally constrain intensities inside metal regions and regularise the artefact latent space:

$$\mathcal{L}_{metal} = \big\|m \odot (\tilde{y} - \mathbf{1})\big\|_2^2, \tag{13}$$

$$\mathcal{L}_{KLD} = \mathrm{KL}(q(z_a \mid x_a) \,\|\, \mathcal{N}(0, I)). \tag{14}$$

Training alternates one gradient step on (E_C, E_A, G_A) using $\mathcal{L}_G$ with one step on D_A using $\mathcal{L}_D$, analogous to standard LSGAN training.

MAGGnet – Generation of Volumetrically Consistent CT Scans. A key goal of this work is to generate volumetrically consistent CT slices. Given a clean volume $X_c = \{x_c^1, \ldots, x_c^N\}$, we construct a volumetrically consistent metal mask $M = \{m^1, \ldots, m^N\}$ and process the slices sequentially using inter-slice guidance. For each slice i, the content code $z_c^i = E_C([x_c^i + m^i,\, g^i])$ is combined with an artefact latent code $z_a^i \sim q(z_a \mid x_a^i)$ to generate the corrupted slice $\tilde{y}^i = G_A(z_c^i, z_a^i)$. Repeating this for all slices yields the artefact-corrupted volume $\tilde{Y} = \{\tilde{y}^1, \ldots, \tilde{y}^N\}$.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_27

3 Experiments

Training and experiments were conducted on two computing platforms: a local workstation equipped with an NVIDIA RTX 4070 Super GPU and an Intel i7-14700K CPU, and the Helios supercomputer, comprising 440 GH200 chips with a theoretical peak performance of 36 PFLOPS.

The model was implemented in PyTorch and trained for 100 epochs using the Adam optimiser with a learning rate of 1×10^{-4} for both the generator and discriminator ($\beta_1 = 0.5$, $\beta_2 = 0.999$), using alternating $1:1$ updates. The following weights were used for the training objective: $\lambda_{GAN} = 1$. $\lambda_{content} = 30$, $\lambda_{edge} = 1.5$, $\lambda_{bg} = 50$, $\lambda_{metal} = 15$, $\lambda_{KLD} = 0.005$. All images were normalised to $[-1, 1]$.

3.1 Experiment 1. Guidance Mask Generation Comparison

Plenty of guidance mask generation methods were compared during the development of the proposed approach, out of which the following 2 showed the most promising results:

- **Method 1:** 1–2 metal sources with 30–120 metal streaks of constant intensity equal to 0.2.
- **Method 2:** 1–2 metal sources with 30–120 metal streaks of constant intensity equal to 0.2, augmented with thicker streaks of intensity 1 connecting the metal sources and extending to the body outline.

Fig. 2 presents a qualitative comparison of the generated results for each guidance mask type. Each row corresponds to one guidance method, following the order listed above. From left to right, the images show the clean CT input, the guidance mask, the artefact-corrupted style image, and the generated artefact-corrupted output. Method 1 produces dense streak patterns. Method 2 additionally enforces strong connections between metal sources. These configurations yield more pronounced artefacts that better resemble streak patterns observed in clinical CT scans.

3.2 Experiment 2. Evaluation of Volumetric Consistency in CT Scan Generation

As illustrated in Fig. 3, the proposed method is capable of generating volumetrically consistent artefacts across adjacent slices. However, qualitative visual inspection alone is insufficient for objective assessment, and a quantitative measure is required to evaluate inter-slice consistency in a principled manner. To the best of the authors' knowledge, no prior work has introduced a dedicated metric for assessing volumetric consistency in medical image generation tasks.

To address this limitation, we propose a gradient-based consistency measure GCM that evaluates the agreement of local structural changes between neighbouring slices. Specifically, inter-slice consistency is quantified using the cosine

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_27

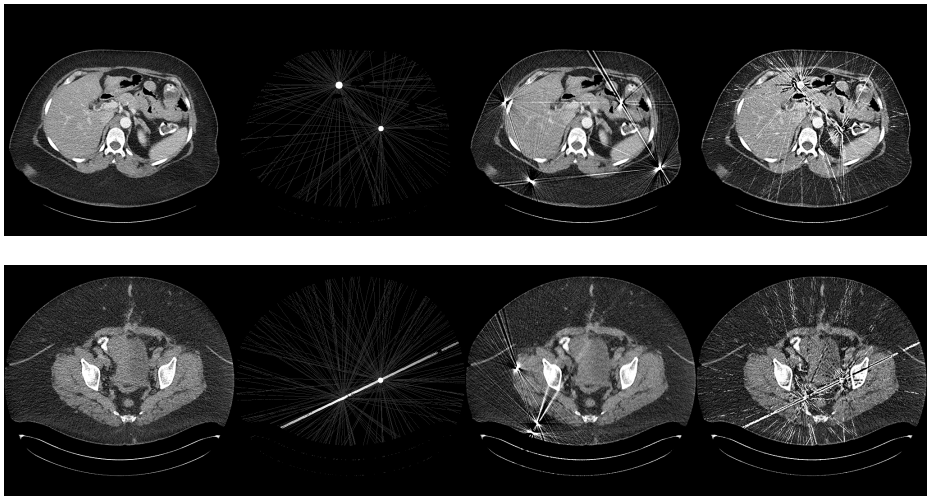

Fig. 2: **Visual comparison of image generation for different guidance-mask strategies.** Each row shows: clean CT input, guidance mask, artefact-corrupted style image, and generated output.

similarity of image gradients between adjacent slices ∇I_z and ∇I_{z+1}, defined as

$$GCM = \cos(\nabla I_z, \nabla I_{z+1}) = \frac{\nabla I_z \cdot \nabla I_{z+1}}{\|\nabla I_z\| \, \|\nabla I_{z+1}\| + \epsilon}, \tag{15}$$

where ϵ is a small constant introduced for numerical stability.

This metric captures the consistency of local structural orientation changes while remaining invariant to absolute intensity scaling and contrast variations. Higher values indicate coherent slice-to-slice structural evolution, whereas lower values correspond to disrupted continuity, as commonly observed in the presence of metal-induced artefacts.

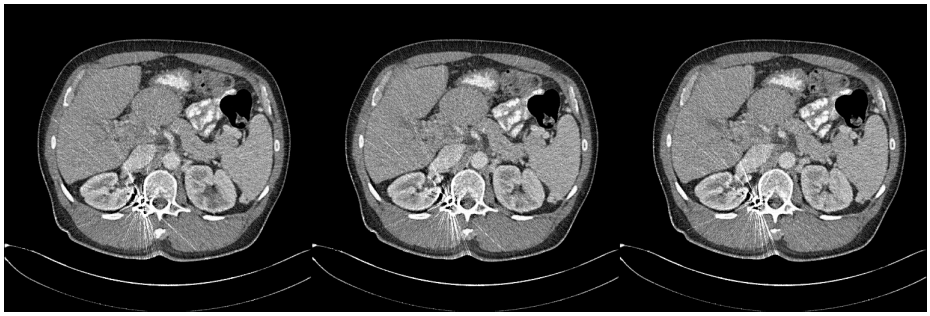

Fig. 3: **Metal artefacts generated on three consecutive CT slices with only minor slice-to-slice variations in the guidance mask.**

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_27

Table 1: *GCM* (Cosine similarity of image gradients) w.r.t. clean CT.

	Baseline	Vol. Consistent	Random
Mean ± Std	0.6901 ± 0.0309	0.5853 ± 0.0304	0.3856 ± 0.0385

For quantitative evaluation, 74 clean CT scans from *DeepLesion* [12] were processed, yielding 5261 images, each with volumetrically consistent or randomly placed artefacts. As shown in Table 1, clean scans achieve the highest mean similarity (0.6901 ± 0.0309), indicating strong inter-slice gradient consistency. Volumetrically consistent artefacts reduce the mean to 0.5853 ($\approx 15\%$ drop) with little change in standard deviation, showing that structured artefacts partially preserve coherence. Random artefacts further lower the mean to 0.3856 with increased variability (± 0.0385), reflecting their irregularity. Thus, volumetrically consistent artefacts maintain substantially more structural continuity than random ones.

4 Conclusions

We propose a guided generative framework that disentangles anatomy from artefacts, using procedural metal guidance to synthesise controllable, volumetrically consistent CT artefacts more efficiently than projection-based simulators like CatSim.

Additionally, we introduce the Gradient-based Continuity Metric (GCM), a cosine-similarity measure for slice-to-slice structural coherence. To our knowledge, this is the first quantitative metric dedicated to volumetric consistency in medical artefact generation.

MAGGNet currently operates purely in the image domain and has been evaluated mainly on structural coherence. Future work will study its impact on MAR training and integrate anatomically informed metal placement.

Acknowledgments. We gratefully acknowledge Polish high-performance computing infrastructure PLGrid (HPC Center: ACK Cyfronet AGH) for providing computer facilities and support within computational grant no. PLG/2025/019073

This work uses the AAPM CT–MAR Grand Challenge datasets [4, 10], generated with a hybrid simulation framework [8] that combines public clinical images [13, 2] with metal objects simulated by CatSim in the open-source XCIST toolkit [9].

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

References

1. Agrawal, H., Hietanen, A., Särkkä, S.: Metal artifact reduction in cone-beam extremity images using gated convolutions. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). pp. 1573–1577. IEEE (2021).

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_27

https://doi.org/10.1109/ISBI48211.2021.9434163, https://ieeexplore.ieee.org/document/9434163
2. Avery, J., Goren, N., Mackle, E., Dowrick, T., Witkowska-Wrobel, A., Holder, D.: The uclh stroke eit dataset – subjects (2017). https://doi.org/10.5281/zenodo.836841, https://doi.org/10.5281/zenodo.836841
3. Barrett, J.F., Keat, N.: Artifacts in ct: recognition and avoidance. Radiographics : a review publication of the Radiological Society of North America, Inc **24 6**, 1679–91 (2004), https://api.semanticscholar.org/CorpusID:46736196
4. Haneda, E., Peters, N., Zhang, J., Karageorgos, G., Xia, W., Paganetti, H., Wang, G., Guo, Y., Ma, J., Park, H.S., Jeon, K., Fan, F., Thies, M., De Man, B.: Aapm ct metal artifact reduction grand challenge. Medical Physics **52**(10), e70050 (2025). https://doi.org/https://doi.org/10.1002/mp.70050, https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.70050
5. Jin, X., Yang, G., Wang, Q., Chen, Y., Zhang, N., Gu, Y.: Metal artifact reduction in ct images. Measurement **178**, 109359 (2021). https://doi.org/10.1016/j.measurement.2021.109359, https://www.sciencedirect.com/science/article/pii/S0263224121009519
6. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks (2019), https://arxiv.org/abs/1812.04948
7. 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 (Mar 2020). https://doi.org/10.1109/tmi.2019.2933425, http://dx.doi.org/10.1109/TMI.2019.2933425
8. Peters, N., Haneda, E., Zhang, J., Karageorgos, G., Xia, W., Verburg, J., Wang, G., Paganetti, H., De Man, B.: A hybrid training database and evaluation benchmark for assessing metal artifact reduction methods for x-ray ct imaging. Medical Physics **52**(10), e70020 (2025). https://doi.org/https://doi.org/10.1002/mp.70020, https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.70020
9. Wu, M., FitzGerald, P., Zhang, J., Segars, W.P., Yu, H., Xu, Y., De Man, B.: XCIST: an open access x-ray/ct simulation toolkit. Physics in Medicine & Biology **67**(19), 194001 (2022), https://pubmed.ncbi.nlm.nih.gov/36096127/
10. XCIST Team: AAPM CT Metal Artifact Reduction (CT-MAR) Grand Challenge Benchmark Tool. https://github.com/xcist/example/tree/main/AAPM_datachallenge/ (2023), accessed: 2026-04-06
11. Xia, J., Zhou, Y., Wang, H., Deng, W., Kang, J., Wu, W., Qi, M., Zhou, L., Ma, J., Xu, Y.: Pnd-net: Physics-based non-local dual-domain network for metal artifact reduction. arXiv preprint arXiv:2305.17778 (2023). https://doi.org/10.48550/arXiv.2305.17778, https://arxiv.org/abs/2305.17778
12. Yan, K., Wang, X., Lu, L., Summers, R.M.: Deeplesion (2018), https://nihcc.app.box.com/v/DeepLesion/
13. 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 (2018). https://doi.org/10.1117/1.JMI.5.3.036501, https://www.spiedigitallibrary.org/journals/journal-of-medical-imaging/volume-5/issue-3/036501/DeepLesion--automated-mining-of-large-scale-lesion-annotations-and/10.1117/1.JMI.5.3.036501.full
14. Zihao, W., Clair, V., Thomas, D., Dan, G., Charles, R., Nicolas, G., Herve, D.: Inner-ear augmented metal artifact reduction with simulation-based 3d generative adversarial networks (2021), https://arxiv.org/abs/2104.12510

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_27