Structured Loss Amplification for U-Net-based Human-scale Dark-field CT Streak Reduction

D. Frey*[1,2], T. Dorosti*[1,2,3], J. B. Thalhammer[1,2,3,4], J. F. Hilmer[1,2], P. Bleuel[1,2], T. Hiu[1,2], S. Peterhansl[1,2], J. McGinnis[5,6], T. Koehler[4,7], D. Pfeiffer[3,4], F. Pfeiffer[1,2,3,4], D. Rueckert[5,6,8], F. Schaff[1,2]

***Abstract*—Streak artifacts remain a substantial challenge for the clinical translation of human-scale dark-field computed tomography. Supervised learning approaches using convolutional neural networks have shown promise in artifact suppression, but typically omit domain-specific prior knowledge. In this work, we present a structured loss formulation that integrates spatial and frequency-informed bias into U-Net-based streak reduction by selectively amplifying pixel-wise loss, guiding the model towards high-fidelity predictions during training. Performance was analyzed across independently trained models and quantified via contrast-to-noise ratio (CNR) and full-width-at-half-maximum (FWHM) of the line spread function at the air-tissue interface. As a substitute for clinical patient data, test metrics were computed from a scan of ventilated ex vivo porcine lungs. Using center- and high-frequency amplification, CNR was enhanced by 24 % for L_1 and 33 % for L_2 loss while preserving adequate image resolution. These results motivate the use of structured loss amplification as a flexible design choice for neural network architectures and demonstrate its effectiveness for advanced streak reduction in dark-field imaging.**

***Index Terms*—Computed tomography, convolutional neural networks, dark-field imaging, deep learning, streak artifact reduction**

I. Introduction

X-RAY dark-field computed tomography (DFCT) measures small-angle scattering as complementary information to conventional attenuation-based imaging [1]. Recently, DFCT imaging has been translated to the human scale, but reconstruction quality remains limited by prominent streak artifacts intensified by the mechanically unstable environment of a rotating grating-based interferometer [2]. While these vibrations enable phase retrieval without an explicit stepping mechanism, fluctuations in the relative grating motion induce artifacts during sliding-window processing [3]. Previous approaches to streak reduction have focused on advancing conventional processing algorithms, for example by suppressing vibrational modes during sample reconstruction [4].

Convolutional neural networks (CNNs) have demonstrated their worth for a wide range of medical imaging tasks. Although predominantly applied to image segmentation, the structure-preserving nature of U-Net architectures makes them well-suited for regression-based image restoration [5]. In the context of attenuation-based CT reconstruction, U-Nets have been successfully applied to the mitigation of undersampling artifacts in low-dose sparse-view conditions [6], [7]. Metal artifact reduction faces a similar problem, for which some solutions extend the traditional supervised formula via self-supervised CNNs or prior-guided algorithms trained on the sinogram domain [8]–[10].

While previous work in these closely related fields provides an inspiring foundation for the development of advanced streak reduction algorithms, the complexity of DFCT artifacts impedes direct translation and dictates careful consideration of design choices, such as the loss function. Approaches such as feature-aware perceptual loss or spectral methods like EAGLE provide alternatives to the traditional loss computation between terminal network prediction and ground truth (GT) [11], [12]. Recently, a normalized loss formulation adopted as part of a dual-frame 3D U-Net has proven feasible for enhanced DFCT reconstructions by accounting for the lower intensity of lung tissue compared to bone [13].

Building on this idea, our work focuses on modulating loss functions tailored to DFCT streak reduction. Contrary to previous strategies, we modify loss on a pixel-wise level by selectively amplifying errors guided by prior knowledge in spatial and frequency domain. In doing so, a structured bias is imposed to purposefully constrain the solution space.

II. Methods

A. Dataset

The training data consisted of 89 reconstructions of CT scans acquired with a human-scale dark-field CT prototype scanner at the Munich Institute of Biomedical Imaging. Samples included a mix of different phantom materials as well as

*Equal contribution.

No live experiments were performed. Ethical approval was not required. The authors declare no conflicts of interest.

[1]Chair of Biomedical Physics, Technical University of Munich (TUM), Germany
[2]Munich Institute of Biomedical Engineering, TUM, Germany
[3]Institute for Diagnostic and Interventional Radiology, TUM University Hospital, TUM, Germany
[4]TUM Institute for Advanced Study, TUM, Germany
[5]Chair for AI in Healthcare and Medicine, TUM, Germany
[6]Munich Center for Machine Learning (MCML), TUM, Germany
[7]Philips GmbH Innovative Technologies, Hamburg, Germany
[8]Department of Computing, Imperial College London, U.K.

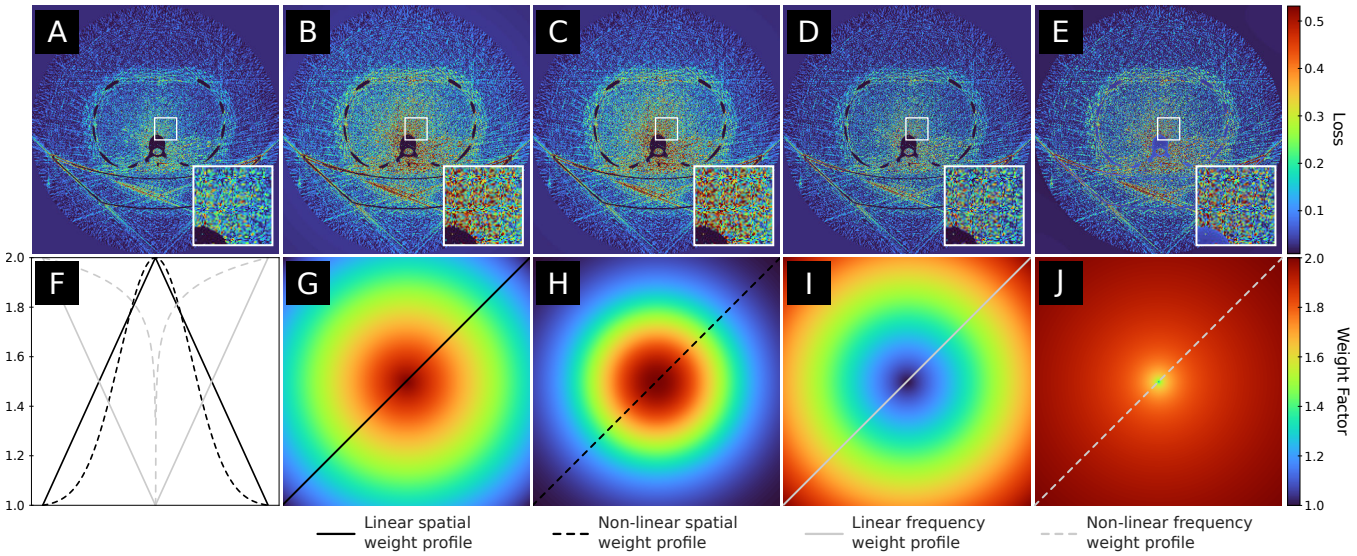

Fig. 1. Loss amplification strategies in spatial and frequency domain. For illustration, the pixel-wise loss map (A) is computed using an L_1 loss between a normalized single-rotation input slice and its multi-rotation GT. Spatially weighted loss maps (B, C) increase contributions towards the image center with linear and non-linear weights (G, H). Frequency-weighted loss maps (D, E) enhance high-frequency components by multiplication with linear and non-linear weights (I, J) in Fourier space. Designed weighting functions are highlighted from line profiles (F). Images are windowed consistently for each row.

ex vivo porcine lungs ventilated inside an anthropomorphic "Lungman" thorax phantom (Kyoto Kagaku, Tokyo, Japan). Experimental conditions were maintained at $80\,\mathrm{kVp}$, $550\,\mathrm{mA}$ and $1.5\,\mathrm{s}$ per full revolution.

Each scan comprised multiple (up to 20) rotations with 2400 shots each to create a stack of full-view projections. Input data at clinically feasible dose was gathered per sliding-window phase retrieval and filtered back-projection (FBP). Higher-quality GT data with significantly diminished artifacts and noise were obtained by virtual phase-stepping across all available rotations and FBP. From each volume, only the central axial slice (800×800 pixels) was used to limit training stochasticity.

B. Model and Training

Throughout all experiments, a fixed model architecture and consistent hyperparameters (learning rate $1\cdot10^{-3}$, weight decay $1\cdot10^{-1}$, Adam optimizer, linear learning rate scheduler with step size 3 and $\gamma = 0.9$, batch size 32) were used to isolate the effect of loss design. Hyperparameters were selected via grid-search using an L_1 loss. The validation split was set to $20\,\%$.

We employed a standard four-block 2D U-Net architecture with bilinear upsampling, batch normalization and ReLU as activation function in a residual-learning formulation. Each layer doubled the number of channels from 64 to 1024 at the deepest level. A global residual connection added the network prediction to the input image. During training, intensities of single- and multi-rotation data were clipped consistently to emphasize streaks and normalized to the range $[0, 1]$. For each input, a random patch (128×128 pixels) was picked to improve generalization; weights were cropped to the same patch during amplification.

Inference on the validation split, as well as on the hold-out test set containing scans of porcine lungs from a different specimen than those used for training, was performed on full-sized slices. Data augmentation comprised horizontal and vertical flips and in-plane rotations with a maximum angle of $\pm15°$ at $50\,\%$ chance. Training was terminated using early stopping with a patience of 20 epochs and a maximum of 100 epochs. Experiments were implemented in Python (v3.12.10) with PyTorch (v2.5.1, CUDA 12.4) and executed on an NVIDIA A100 GPU with $80\,\mathrm{GB}$ VRAM. While training and inference can be performed with substantially reduced VRAM, we chose to maximize batch sizes to reduce run-to-run variance given the limited dataset size.

C. Structured Loss Design

We based our design choices on previous knowledge collected from human-scale DFCT data. Pixel-wise loss maps $\mathcal{L} \in \mathbb{R}^{H\times W}$ for slices of height H and width W were multiplied element-wise with amplification weights $\mathcal{A}$. Weights were constrained to a range of $[1, 2]$ to mitigate the risk of vanishing or exploding gradients. Fig. 1 illustrates how this strategy visually enhanced structurally relevant components.

Focus towards the image center was enforced by increasing spatial weights $\mathcal{A}_S$ towards the iso-center to penalize errors more heavily in sample regions of interest (ROIs):

$$\mathcal{L}_S = \mathcal{A}_S \cdot \mathcal{L}. \tag{1}$$

This approach also emphasized regions where streaks typically appear stronger. The non-linear radial profile provided a smoother decay towards the image edges and a more distinct transition between ROI and background.

Frequency-space weights $\mathcal{A}_F$ were designed to amplify high-frequency contributions and multiplied with the Fourier-transformed loss map. The modulated loss follows as the real part of the inverse transformation after weighting:

$$\mathcal{L}_F = \mathrm{Re}\left\{\mathcal{F}_{2\mathrm{D}}^{-1}\left\{\mathcal{A}_F \cdot \mathcal{F}_{2\mathrm{D}}\left\{\mathcal{L}\right\}\right\}\right\}, \tag{2}$$

where $\mathcal{F}_{2\mathrm{D}}$ and $\mathcal{F}_{2\mathrm{D}}^{-1}$ denote the 2D Fourier transform and its inverse, and Re the real part. This approach was fine-tuned to the typical signature of DFCT streaks by further increasing the weight of mid- and high-frequency bands in the non-linear version. Concatenated weighting was implemented as consecutive application of (1) and (2) or their reversed order.

D. Performance Evaluation

For a systematic assessment of our loss amplification approaches, we trained all possible experimental configurations for ten repetitions at individually fixed random seeds. Configurations combined the following aspects:

- L_1 vs. L_2 as the baseline loss,
- linear vs. non-linear weighting functions, and
- spatial (S) vs. frequency (F) amplification, as well as their combined application.

Underperforming models were included to assess the robustness of each configuration.

After training, model performance was evaluated on test data and quantified for a single example slice via contrast-to-noise ratio (CNR) incorporating both signal (lung tissue) and background (air) noise:

$$\mathrm{CNR} = \frac{\mu_{\mathrm{lung}} - \mu_{\mathrm{air}}}{\sqrt{\sigma_{\mathrm{lung}}^2 + \sigma_{\mathrm{air}}^2}}, \tag{3}$$

where μ denotes the mean and σ the standard deviation of each ROI. Additionally, we analyzed the trade-off with image resolution by computing the full-width-at-half-maximum (FWHM) of the edge spread function derivative. ROIs (30×30 pixels) were picked for lung tissue as signal region and air as background reference close to the image center. The line profile for FWHM computation was drawn between the centers of both ROIs.

III. Results and Discussion

A. Quantitative Comparison of Model Performances

A comprehensive comparison of test performance metrics across differently amplified models is presented in Fig. 2. Median values with their standard deviations are listed in Tab. I for L_1 and Tab. II for L_2, maximum CNR values per weighting scheme are highlighted in bold font. Using L_2, loss amplification leads to modest improvements in CNR while

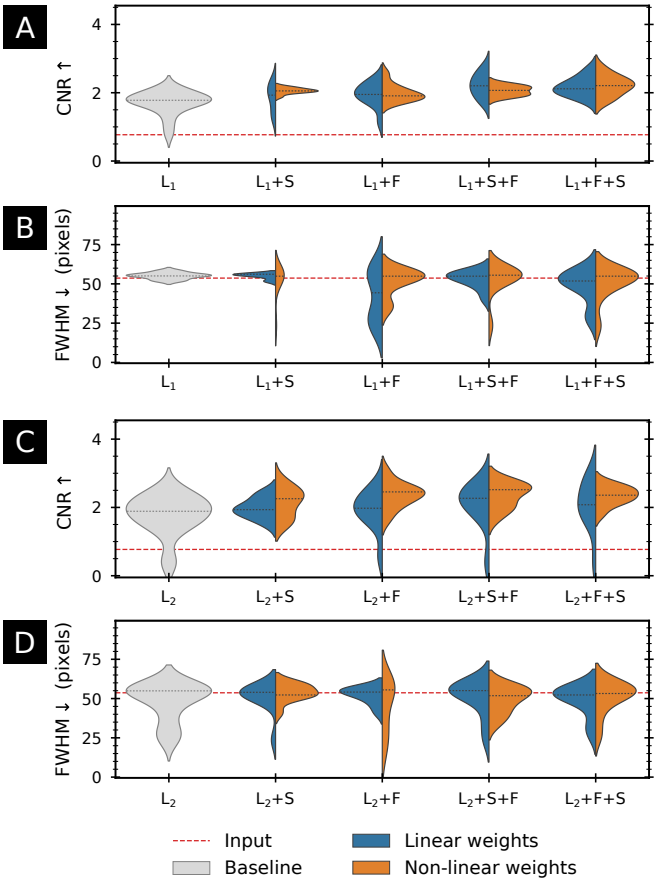

Fig. 2. Test performance for ten independently trained models in each experimental configuration. Violin plots illustrate the distributions of contrast-to-noise ratio (CNR) and full-width-at-half-maximum (FWHM) of the air-tissue edge for the original input image and predictions from models trained on amplified L_1 (A, B) and L_2 (C, D). Dashed lines indicate median values. Loss amplifications were realized with linear (blue) and non-linear weightings (orange). Abbreviations denote spatial weights (S), frequency weights (F), and their concatenations (S+F and F+S).

largely preserving FWHM, with the strongest gains observed for concatenated weighting schemes. Relative improvement in CNR reaches up to 24 % ($\mathrm{CNR} = 2.21 \pm 0.30$) compared to the baseline model ($\mathrm{CNR} = 1.78 \pm 0.34$). Non-linear weighting appears more robust against suboptimal outlier models, reflected by a slightly reduced variance across repetitions.

L_2-based models exhibit more pronounced performance gains from amplification. All configurations yield higher CNR values, achieving at maximum a relative increase of 33 % ($\mathrm{CNR} = 2.52 \pm 0.37$) from baseline ($\mathrm{CNR} = 1.89 \pm 0.58$), again observed for the concatenated configuration. Non-linear L_2 amplification consistently outperforms linearly weighted models. For the highest CNR gains, variance is reduced as well.

Loss amplification preserves spatial resolution at comparable levels to the baseline in exchange for substantial CNR

TABLE I
TEST METRICS FOR L_1 ($n = 10$, MEDIAN $\pm$ STANDARD DEVIATION)

	CNR ↑		FWHM ↓ (pixels)	
Experiment	Linear	Non-linear	Linear	Non-linear
L_1	1.78 ± 0.34		55 ± 3	
L_1+S	1.93 ± 0.41	2.05 ± 0.08	56 ± 2	55 ± 11
L_1+F	1.95 ± 0.36	1.91 ± 0.25	44 ± 16	55 ± 9
L_1+S+F	$\mathbf{2.20 \pm 0.35}$	2.07 ± 0.14	55 ± 6	56 ± 11
L_1+F+S	2.12 ± 0.28	$\mathbf{2.21 \pm 0.30}$	52 ± 11	55 ± 11

TABLE II
TEST METRICS FOR L_2 ($n = 10$, MEDIAN $\pm$ STANDARD DEVIATION)

	CNR ↑		FWHM ↓ (pixels)	
Experiment	Linear	Non-linear	Linear	Non-linear
L_2	1.89 ± 0.58		55 ± 12	
L_2+S	1.93 ± 0.30	2.26 ± 0.43	54 ± 10	52 ± 6
L_2+F	1.98 ± 0.58	2.45 ± 0.37	55 ± 6	56 ± 17
L_2+S+F	$\mathbf{2.27 \pm 0.66}$	$\mathbf{2.52 \pm 0.37}$	56 ± 12	52 ± 9
L_2+F+S	2.08 ± 0.74	2.36 ± 0.27	53 ± 10	53 ± 11

gains. The observed differences between configurations suggest that spatial and frequency-based structuring contribute complementary information during training.

B. *Qualitative Comparison of Model Performances*

Fig. 3 illustrates input, GT, and prediction slices from the test dataset for the assessment of streak reduction efficacy with non-linearly amplified L_2 loss. The single-rotation reconstruction exhibits pronounced streaks of varying intensity superimposed on the lung signal. These artifacts are largely suppressed in the multi-rotation reconstruction, as phase information is sampled more densely from the surplus of information. Although dose-wise clinically unfeasible, the quality of this reconstruction emphasizes the inherent potential of human-scale DFCT for pulmonary imaging.

Inference with the U-Net trained using baseline L_2 loss achieves only partial streak reduction. Residual artifacts appear as circular bands of alternating intensity. In contrast, all loss-amplified predictions are capable of mitigating these residual streaks globally.

Spatial loss amplification (L_2+S) enhances CNR while simultaneously improving resolution, as evidenced by sharper structural detail within the lung parenchyma. This observation suggests that emphasizing central regions during training is critical for effective streak mitigation. Frequency loss amplification (L_2+F) further increases CNR at the cost of FWHM, enhancing contrast at boundaries near large air cavities. Notably, this configuration treats intensity variations in homogeneous regions more conservatively than spatial amplification, which may help prevent the misinterpretation of residual artifacts as anatomical features.

Concatenated weighting (shown for L_2+S+F) prioritizes contrast enhancement over resolution, yielding a visually balanced trade-off between the two approaches. Lung tissue appears smoother at reasonable resolution. Although none of the predictions fully match the sharpness of the GT, CNR is clearly superior, as even multi-rotation reconstructions retain fine streak artifacts. In general, the observed gains in CNR can be attributed to reduced noise in both tissue and air regions at similar signal levels.

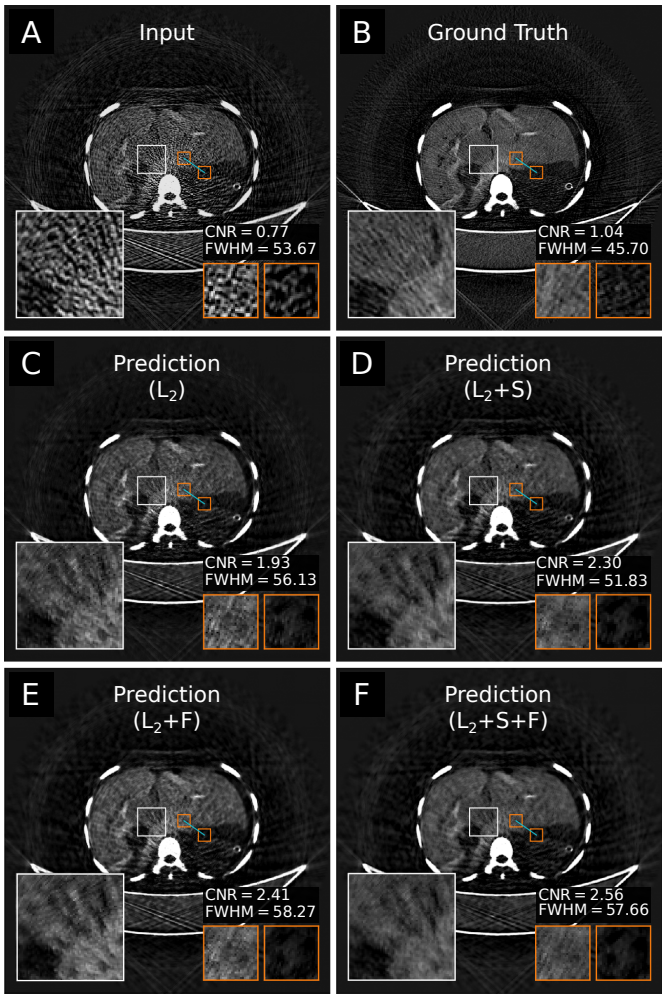

Fig. 3. Improved streak reduction of an axial DFCT slice from the test dataset with non-linear L_2 amplification. Inference tested on single-rotation input data (A) with a multi-rotation GT reconstruction for reference (B). Predictions retrieved using the baseline (C), spatially amplified (D), frequency-amplified (E), and concatenated spatially and frequency-amplified model (F). Out of all ten options per configuration, the prediction with CNR closest to each respective median is shown as representative slice. ROIs (orange) and line profile (cyan) for contrast-to-noise ratio (CNR) and full-width-at-half-maximum (FWHM) computation are indicated.

IV. CONCLUSION

This work demonstrates that incorporating domain knowledge directly into pixel-wise loss computation provides a powerful and interpretable mechanism for guiding deep learning based artifact suppression. Structured loss design offers a pathway toward narrowing the gap between domain knowledge and data-driven learning, ultimately supporting

the development of physics-informed models for clinical imaging. The direct comparison between linear and non-linear weighting strategies further indicates that fine-tuning loss maps can unlock additional performance gains, even when applied to comparatively simple U-Net architectures.

Notably, these improvements were achieved despite very limited training data, suggesting that further gains can be expected with continued data acquisition. Beyond the 2D setting, extending loss amplification to 3D U-Net architectures might further improve training based on global contextual CT information. While we originally aimed at removing DFCT streaks, our results indicate that CNNs also suppress other types of noise, an effect that should be examined more broadly in future work.

As DFCT streaks appear non-local and directional, more sophisticated weighting schemes incorporating masks adapted to streak orientations may be more effective than isotropic weighting functions, and should therefore be examined. Considering that even the multi-rotation GT exhibits fine streaks, a simulation study could provide further insights into how U-Net training is affected, while also offering a more concise basis for comparative evaluation. Additionally, unbiased reader studies will be essential for fine-tuning informed loss design choices.

ACKNOWLEDGMENT

Financial support through the European Research Council (ERC Smart Detectors for Darkfield X-ray Imaging, SyG 101167328), and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the States, as well as by the Technical University of Munich – Institute for Advanced Study.

REFERENCES

[1] F. Pfeiffer, M. Bech, O. Bunk, P. Kraft, E. F. Eikenberry, C. Brönnimann, C. Grünzweig, and C. David, "Hard-X-ray dark-field imaging using a grating interferometer," *Nature Materials*, vol. 7, no. 2, pp. 134–137, 2008.

[2] M. Viermetz, N. Gustschin, C. Schmid, J. Haeusele, M. v. Teuffenbach, P. Meyer, F. Bergner, T. Lasser, R. Proksa, T. Koehler, and F. Pfeiffer, "Dark-field computed tomography reaches the human scale," *Proceedings of the National Academy of Sciences*, vol. 119, no. 8, 2022.

[3] C. Schmid, M. Viermetz, N. Gustschin, W. Noichl, J. Haeusele, T. Lasser, T. Koehler, and F. Pfeiffer, "Modeling vibrations of a tiled Talbot-Lau interferometer on a clinical CT," *IEEE Transactions on Medical Imaging*, vol. 42, no. 3, pp. 774–784, 2023.

[4] J. Haeusele, C. Schmid, M. Viermetz, N. Gustschin, T. Lasser, T. Koehler, and F. Pfeiffer, "Robust sample information retrieval in dark-field computed tomography with a vibrating Talbot-Lau interferometer," *IEEE Transactions on Medical Imaging*, 2024.

[5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Springer International Publishing, 2015, pp. 234–241.

[6] J. Thalhammer, M. Schultheiß, T. Dorosti, T. Lasser, F. Pfeiffer, D. Pfeiffer, and F. Schaff, "Improving automated hemorrhage detection at sparse-view CT via U-net–based artifact reduction," *Radiology: Artificial Intelligence*, vol. 6, no. 4, p. e230275, 2024.

[7] A. Ries, T. Dorosti, J. Thalhammer, D. Sasse, A. Sauter, F. Meurer, A. Benne, T. Lasser, F. Pfeiffer, F. Schaff, and D. Pfeiffer, "Improving image quality of sparse-view lung tumor CT images with U-Net," *European Radiology Experimental*, vol. 8, no. 1, p. 54, 2024.

[8] L. Gjesteby, Q. Yang, Y. Xi, B. Claus, Y. Jin, B. De Man, and G. Wang, "Reducing metal streak artifacts in CT images via deep learning: Pilot results," in *The 14th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 14, no. 6, 2017, pp. 611–614.

[9] B. Kim, H. Shim, and J. Baek, "A streak artifact reduction algorithm in sparse-view CT using a self-supervised neural representation," *Medical Physics*, vol. 49, no. 12, pp. 7497–7515, 2022.

[10] L. Yu, Z. Zhang, X. Li, and L. Xing, "Deep sinogram completion with image prior for metal artifact reduction in CT images," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 228–238, 2021.

[11] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Springer International Publishing, 2016, pp. 694–711.

[12] Y. Sun, Y. Huang, Z. Yang, L.-S. Schneider, M. Thies, M. Gu, S. Mei, S. Bayer, F. G. Zöllner, and A. Maier, "EAGLE: An edge-aware gradient localization enhanced loss for CT image reconstruction," *Journal of Medical Imaging*, vol. 12, no. 1, pp. 014 001–014 001, 2025.

[13] T. Kumschier, J. Thalhammer, C. Schmid, J. Haeusele, T. Koehler, F. Pfeiffer, T. Lasser, and F. Schaff, "Streak artefact removal in X-ray dark-field computed tomography using a convolutional neural network," *Medical Physics*, vol. 51, no. 10, pp. 7404–7414, 2024.