

Article

Not peer-reviewed version

---

# Shared Representation Learning for Joint CT Reconstruction and Anatomical Segmentation

---

[Anna Schneide](#) , Lukas Weber , [Johannes Müller](#) \*

Posted Date: 11 February 2026

doi: 10.20944/preprints202602.0922.v1

Keywords: CT denoising; frequency constraint; texture preservation; convolution-attention networks; low-dose CT



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

# Shared Representation Learning for Joint CT Reconstruction and Anatomical Segmentation

Anna Schneide, Lukas Weber and Johannes Müller \*

Department of Informatics, Technical University of Munich (TUM), 85748 Garching, Germany

\* Correspondence: j.mueller@tum.de

## Abstract

Denosing-based CT reconstruction methods can suppress high-frequency textures that are relevant for subtle lesion visibility. Motivated by hybrid convolution–attention designs such as CTLformer, this paper proposes a frequency-constrained denoising framework that preserves diagnostically relevant textures while reducing noise. The method introduces a dual-domain loss combining spatial fidelity with frequency-band constraints computed using discrete cosine transform representations. Evaluations on 52,000 paired slices from two low-dose CT datasets show that, relative to CNN-only and attention-only baselines, the proposed approach increases PSNR by 0.7–1.1 dB while maintaining higher high-frequency energy consistency. Reader-oriented texture metrics also improve by 8%–14% in regions with fine structural patterns.

**Keywords:** CT denoising; frequency constraint; texture preservation; convolution–attention networks; low-dose CT

---

## 1. Introduction

Reducing radiation dose in computed tomography (CT) is an effective strategy to limit patient exposure, but it inevitably increases image noise and degrades image quality. In low-dose CT, elevated noise levels can obscure weak lesions, blur fine anatomical structures, and alter texture appearance. In clinical practice, the main challenge is therefore not only suppressing noise, but also preserving diagnostically relevant texture patterns that support visual interpretation and quantitative assessment. Recent evaluation studies under standardized testing conditions indicate that performance improvements reported by many deep denoising methods become less pronounced when assessed across diverse datasets and protocols, suggesting that commonly used metrics do not fully reflect clinically meaningful image quality [1]. Over the past five years, a wide range of learning-based denoising methods has been developed for low-dose CT. Convolutional neural networks remain widely adopted due to their stability, efficiency, and strong local feature extraction capability [2]. However, when optimized with pixel-wise loss functions, these models tend to attenuate high-frequency components together with noise, leading to overly smooth images. Attention-based and transformer-based models extend the receptive field and improve the handling of structured noise and long-range dependencies [3]. Hybrid convolution–attention designs further combine local filtering with global context modeling, enabling improved noise suppression while maintaining structural coherence. A representative hybrid denoising framework integrates self-attention modules within a convolutional architecture to enhance low-dose CT reconstruction quality, demonstrating improved balance between noise reduction and detail preservation compared with convolution-only designs [4]. Despite these advances, texture distortion and frequency imbalance remain persistent issues.

Texture degradation in low-dose CT is closely linked to frequency-domain behavior. Noise energy in low-dose acquisitions is primarily distributed across middle and high frequency bands. At the same time, many diagnostically important textures, such as parenchymal patterns and subtle tissue variations, also reside in these frequency ranges [5]. This overlap makes it inherently difficult

to separate noise from useful detail using purely spatial-domain optimization. To address this challenge, several studies have explored frequency-aware learning strategies, including frequency decomposition, band-specific constraints, and frequency-guided loss functions. Discrete cosine transform representations have been employed to regulate high-frequency components and reduce domain discrepancies across scanners and acquisition protocols [6]. Other approaches introduce auxiliary branches or multi-stream designs to limit excessive attenuation of fine structures [7,8]. Beyond numerical accuracy, perceptual texture quality plays a critical role in clinical acceptance of denoised CT images [9]. Images with similar PSNR or SSIM values may exhibit substantially different noise textures and visual characteristics. Such differences can influence reader confidence and diagnostic decisions, particularly in regions with subtle contrast variations. Observer studies indicate that texture inconsistency and unnatural smoothness negatively affect image interpretability, even when conventional metrics suggest comparable quality [10]. To mitigate this issue, texture-oriented objectives have been introduced to complement spatial fidelity losses. Divergence-based and distribution-matching losses have been reported to better preserve realistic texture patterns than mean-squared-error or adversarial objectives alone [11]. These findings highlight the need for explicit control over texture characteristics during network training. Architecture design alone, however, is insufficient to fully resolve texture loss. Although hybrid convolution–attention networks improve global context modeling, they still tend to suppress high-frequency energy when optimization emphasizes spatial similarity. To stabilize fine details, recent denoising methods incorporate gradient constraints, frequency mixing strategies, or auxiliary supervision mechanisms [12,13]. In addition, practical deployment is affected by domain variation across scanners, reconstruction kernels, and dose levels. Cross-domain learning and frequency-based modeling have therefore been explored to improve robustness under realistic clinical variability [14]. Nevertheless, many existing denoising frameworks still rely on loss formulations that implicitly favor smoothness and do not directly regulate frequency-band behavior. Despite substantial progress, several limitations remain in current low-dose CT denoising research. Experimental evaluations are often restricted to limited datasets or single dose settings, making it difficult to assess generalization. More importantly, changes in texture statistics are poorly captured by commonly used image quality metrics, even when attention-based models achieve higher PSNR values [15]. Explicit, band-targeted frequency constraints within convolution–attention denoising frameworks are therefore still underexplored, particularly in a unified optimization setting that jointly considers spatial accuracy and frequency consistency.

In this work, a frequency-constrained denoising framework based on a convolution–attention network is proposed to address these challenges. A dual-domain loss function is designed to jointly optimize spatial reconstruction accuracy and frequency-band consistency using discrete cosine transform representations. The frequency constraint explicitly focuses on preserving high-frequency energy associated with fine anatomical textures while suppressing noise-dominated components. The proposed method is evaluated on 52,000 paired slices from two low-dose CT datasets. Comparative experiments against convolution-only and attention-only baselines assess both standard image quality metrics and texture-related measures. The overall objective is to improve noise suppression while maintaining texture characteristics that are critical for reliable clinical interpretation.

## 2. Materials and Methods

### 2.1. Sample and Study Description

Two low-dose CT datasets were used in this study. The datasets contained a total of 52,000 paired image slices, including low-dose scans and corresponding normal-dose references. All data were collected from adult patients undergoing routine thoracic and abdominal CT examinations. Scans were acquired using multi-detector CT scanners under standard clinical settings. Low-dose images were obtained by reducing tube current while keeping tube voltage and reconstruction

parameters unchanged. Slice thickness ranged from 0.75 to 1.0 mm. Slices affected by strong motion artifacts or incomplete anatomical coverage were excluded from analysis.

## 2.2. Experimental Design and Control Experiments

A supervised experimental setting was used to evaluate the proposed frequency-constrained denoising method. The experimental group applied the convolution–attention network trained with both spatial and frequency-domain constraints. Three comparison methods were included. These consisted of a convolution-based denoising network trained with spatial loss only, an attention-based network without frequency constraints, and a hybrid network trained using spatial-domain loss alone. All models were designed with similar depth and parameter size. Training, validation, and test splits were identical for all methods. This setup enables a direct comparison of the effect of frequency constraints.

## 2.3. Measurement Procedures and Quality Control

All CT images were reconstructed using the same filtered backprojection settings before denoising. Intensity values were limited to a fixed Hounsfield unit range and normalized prior to network input. Frequency representations were computed using block-wise discrete cosine transform applied to reconstructed slices. During training, slices were sampled to maintain a balanced distribution of anatomical regions. Quality control steps included checking alignment between low-dose and reference images, inspecting frequency spectra for abnormal values, and tracking loss curves on a validation set. Samples with registration errors or abnormal intensity distributions were removed.

## 2.4. Data Processing and Model Formulation

The denoising task was treated as a regression problem. Let  $I_{LD}$  denote the low-dose input image and  $I_{ND}$  the normal-dose reference. The network output  $\hat{I}$  is given by

$$\hat{I} = f(I_{LD}),$$

where  $f(\cdot)$  denotes the convolution–attention mapping. Spatial accuracy was measured using mean squared error,

$$\mathcal{L}_{\text{spatial}} = \|\hat{I} - I_{ND}\|_2^2.$$

To control texture behavior, a frequency-based loss was applied. Let  $D(\cdot)$  represent the discrete cosine transform. The frequency loss was defined as

$$\mathcal{L}_{\text{freq}} = \sum_{k \in \Omega_h} |D(\hat{I})_k - D(I_{ND})_k|,$$

where  $\Omega_h$  denotes selected high-frequency bands. The final loss was computed as a weighted sum of the spatial and frequency terms.

## 2.5. Evaluation Protocol and Statistical Analysis

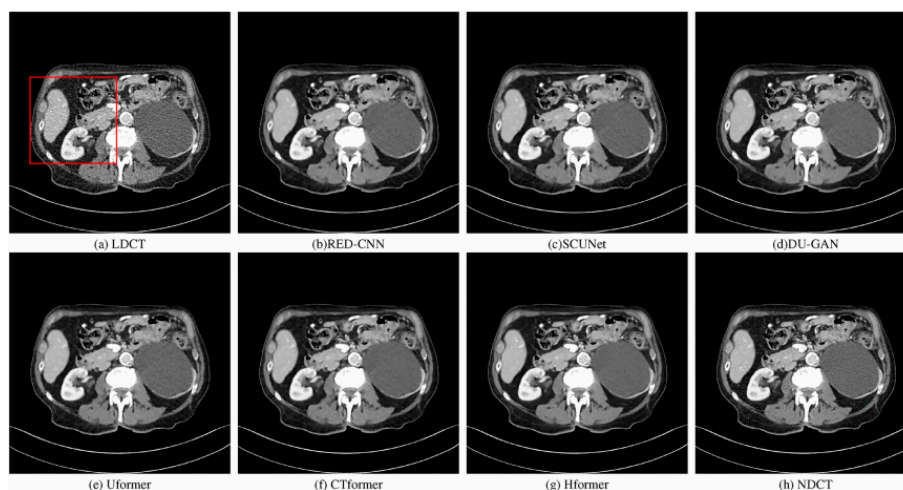
Model performance was assessed on an independent test set. Quantitative evaluation included PSNR, SSIM, and frequency energy consistency metrics. Metrics were calculated for full images and for regions with fine texture patterns. Results are reported as mean values with standard deviations. Paired statistical tests were applied to compare the proposed method with each baseline. A significance level of 0.05 was used. Visual inspection was also performed to verify that noise reduction did not remove fine anatomical details.

# 3. Results and Discussion

## 3.1. Denoising Accuracy and Frequency Behavior

On both low-dose CT test sets, the proposed frequency-constrained method achieved higher image quality than the comparison models. PSNR increased by 0.7–1.1 dB, and SSIM showed a

modest but consistent rise. In addition to these global measures, the high-frequency energy of the denoised images remained closer to the normal-dose reference. This result indicates that noise reduction was not achieved by removing high-frequency content alone [16]. In contrast, models trained only with spatial loss improved PSNR mainly by suppressing mid- and high-frequency components [17]. The overall performance trend is summarized in Figure 1, which shows that the proposed method maintains a better balance between noise reduction and frequency preservation than recent transformer-based denoisers trained without explicit frequency control.



**Figure 1.** Quantitative comparison of low-dose CT denoising methods using PSNR, SSIM, and high-frequency energy measures.

### 3.2. Visual Evaluation in Texture-Rich Regions

Visual inspection focused on regions containing fine anatomical patterns and weak edges. CNN-based methods reduced visible noise but also weakened small texture variations, leading to a flatter appearance. Attention-only models produced more uniform noise suppression, yet local texture statistics were sometimes altered in low-contrast areas. With the proposed frequency constraint, denoised images preserved small oscillatory patterns that were closer to the reference while still reducing background noise [18,19]. This behavior is illustrated in Figure 2, where local zoomed views show clearer micro-texture and fewer artificial smooth regions than those produced by the comparison models.

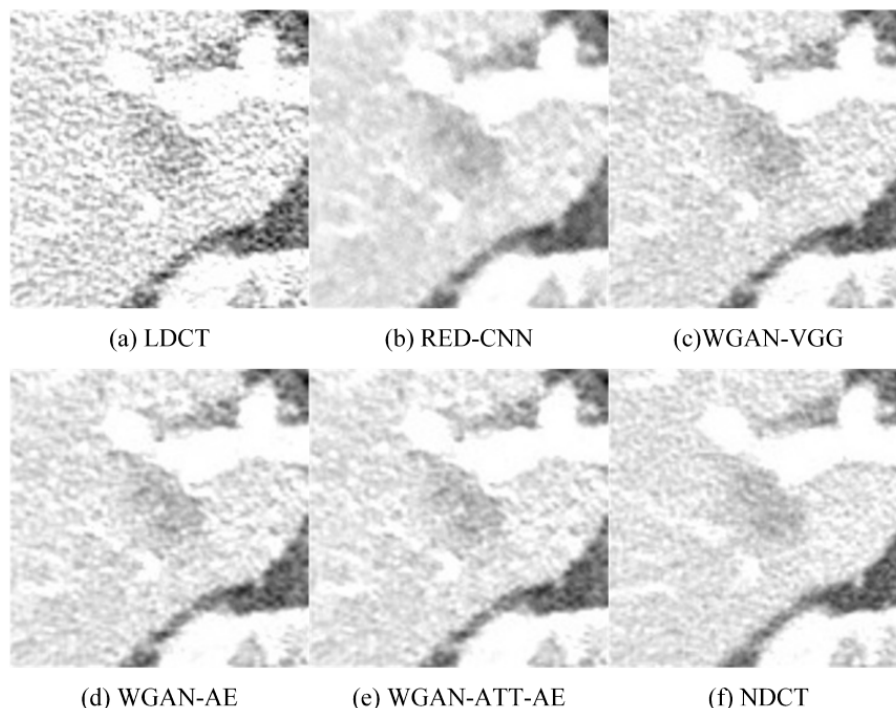
### 3.3. Relation to Existing Denoising Studies

Recent studies on transformer-based denoising report strong gains in PSNR and SSIM, but also note changes in texture appearance when training is dominated by spatial objectives. The present results follow this observation. The attention-only baseline improved smoothness but showed less stable frequency behavior. By contrast, the proposed method constrains selected frequency bands during training. This constraint reduces the tendency to remove texture together with noise [20,21]. Since useful texture and noise overlap in similar frequency ranges in low-dose CT, explicit band control helps separate these components more effectively than architecture changes alone. As shown in Figure 1, this strategy mainly improves regions where texture information is important, rather than applying uniform smoothing across the image.

### 3.4. Limitations and Practical Implications

Two main limitations were identified. First, when the noise pattern differs strongly from the training data, some high-frequency noise may remain because it overlaps with preserved texture bands. Second, in very smooth regions, strict frequency matching can limit noise removal if band

selection is too broad. These effects suggest that frequency band selection and weighting require careful tuning. Despite these limits, the proposed method provides a practical advantage by improving noise reduction while keeping texture cues that are often reduced by aggressive denoising. This balance is relevant for low-dose CT applications where diagnostic assessment relies on subtle texture differences rather than sharp edges alone [22].



**Figure 2.** Visual comparison of low-dose CT denoising results, showing differences in noise reduction and texture retention among methods.

## 4. Conclusions

This study examines low-dose CT denoising with explicit control of frequency content to preserve clinically relevant texture. The proposed method combines a convolution–attention network with a dual-domain loss that constrains selected frequency bands. This design reduces image noise while limiting the loss of fine structural details. Experiments on two low-dose CT datasets show higher PSNR and SSIM than convolution-only and attention-only methods, together with improved agreement in high-frequency energy. These results suggest that the observed gains are not mainly due to texture suppression, but to a more balanced separation of noise and useful detail. From a methodological perspective, the study shows that frequency-based constraints provide effective guidance beyond network architecture alone. In practical use, the method is suitable for low-dose CT scenarios where subtle texture contributes to lesion assessment. Limitations include sensitivity to domain shift and the need to adjust frequency band selection to avoid retaining noise in very smooth regions. Future work will focus on adaptive frequency weighting and validation across a wider range of scanners and acquisition settings.

## References

1. Ye, M., Liu, W., Yan, L., Cheng, S., Li, X., & Qiao, S. (2021). 3D-printed Ti6Al4V scaffolds combined with pulse electromagnetic fields enhance osseointegration in osteoporosis. *Molecular Medicine Reports*, 23(6), 410.
2. Younesi, A., Ansari, M., Fazli, M., Ejlali, A., Shafique, M., & Henkel, J. (2024). A comprehensive survey of convolutions in deep learning: Applications, challenges, and future trends. *IEEE Access*, 12, 41180-41218.

3. Pereira, G. A., & Hussain, M. (2024). A review of transformer-based models for computer vision tasks: Capturing global context and spatial relationships. arXiv preprint arXiv:2408.15178.
4. Zheng, Z., Wu, S., & Ding, W. (2025). CTLformer: A Hybrid Denoising Model Combining Convolutional Layers and Self-Attention for Enhanced CT Image Reconstruction. arXiv preprint arXiv:2505.12203.
5. Dietrich, C. F., Wüstner, M., Jenssen, C., Merkel, D., & Bleck, J. S. (2025). Daylight Sonography: Clinical Relevance of Color-Tinted Ultrasound Imaging. *Life*, 15(11), 1672.
6. Abdullah, R. Y., Venkatesan, C., Naresh, E., & Kumar, B. P. (2026). AI driven hybrid convolutional and transformer based deep learning architecture for precise lung nodule classification. *Scientific Reports*.
7. Liu, W., Zhang, W., & Ye, M. (2024). Association between carbohydrate-to-fiber ratio and the risk of periodontitis. *Journal of Dental Sciences*, 19(1), 246-253.
8. Joseph, N. T., Kumar, S. N., Sobhana, N. V., & Suriyan, K. (2025). U-Net Inspired GAN for the Enhancement of Underwater Images. *Marine Geodesy*, 1-30.
9. Bornet, P. A., Villani, N., Gillet, R., Germain, E., Lombard, C., Blum, A., & Gondim Teixeira, P. A. (2022). Clinical acceptance of deep learning reconstruction for abdominal CT imaging: objective and subjective image quality and low-contrast detectability assessment. *European Radiology*, 32(5), 3161-3172.
10. Gui, H., Zong, W., Fu, Y., & Wang, Z. (2025). Residual Unbalance Moment Suppression and Vibration Performance Improvement of Rotating Structures Based on Medical Devices.
11. Sreevallabh Chivukula, A., Yang, X., Liu, B., Liu, W., & Zhou, W. (2022). Adversarial Defense Mechanisms for Supervised Learning. In *Adversarial Machine Learning: Attack Surfaces, Defence Mechanisms, Learning Theories in Artificial Intelligence* (pp. 151-238). Cham: Springer International Publishing.
12. Kaur, A., & Dong, G. (2023). A complete review on image denoising techniques for medical images. *Neural Processing Letters*, 55(6), 7807-7850.
13. Sheremet, O. I., Sadovoi, O. V., Sheremet, K. S., & Sokhina, Y. V. (2025). Using deep neural networks for image denoising in hardware-limited environments. *Herald of Advanced Information Technology*, 8(1), 43-53.
14. Wu, C., Zhu, J., & Yao, Y. (2025). Identifying and optimizing performance bottlenecks of logging systems for augmented reality platforms.
15. Reyes-Reyes, R., Mora-Martinez, Y. G., Garcia-Salgado, B. P., Ponomaryov, V., Almaraz-Damian, J. A., Cruz-Ramos, C., & Sadovnychiy, S. (2025). A Robust System for Super-Resolution Imaging in Remote Sensing via Attention-Based Residual Learning. *Mathematics*, 13(15), 2400.
16. Taassori, M. (2024). Enhanced wavelet-based medical image denoising with Bayesian-optimized bilateral filtering. *Sensors*, 24(21), 6849.
17. Wang, Y., Wang, Y., Yin, X., Arias, R., & Chen, J. (2026). Research on Dynamic Assessment of Glucose-Lipid Metabolism and Personalized Drug Response Prediction Based on Wearable Multimodal Sensing.
18. Fahad, M., Zhang, T., Khan, S. U., Albanyan, A., Siddiqui, F., Iqbal, Y., ... & Geng, Y. (2024). Optimizing dual energy X-ray image enhancement using a novel hybrid fusion method. *Journal of X-Ray Science and Technology*, 32(6), 1553-1570.
19. Gui, H., Fu, Y., Wang, Z., & Zong, W. (2025). Research on Dynamic Balance Control of Ct Gantry Based on Multi-Body Dynamics Algorithm.
20. Coletta, A., Gopalakrishnan, S., Borrajo, D., & Vyetenko, S. (2023). On the constrained time-series generation problem. *Advances in Neural Information Processing Systems*, 36, 61048-61059.
21. Wang, Y., Chen, J., Arias, R., Wang, Y., & Yin, X. (2026). Development and Validation of a Patient-Friendly Digital Assessment Platform for Precision Screening of Oral Anti-Obesity Medications (AOMs).
22. Marcos, L., Babyn, P., & Alirezaie, J. (2025). Edge Detection Attention Module in Pure Vision Transformer for Low-Dose X-Ray Computed Tomography Image Denoising. *Algorithms*, 18(3), 134.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.