Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 16, No. 2, 2026

Research on Low-Dose CT Image Denoising Methods Based on Self-Supervised Learning

An Ge

Yunnan Normal University, Kunming, Yunnan, China

Abstract: Low-dose CT images are affected by noise and artifacts, compromising diagnostic reliability. While deep learning-based denoising methods have shown promise, most rely on paired normal-dose CT images, which are difficult to obtain clinically. To address this, this study proposes a self-supervised denoising method using pseudo-labels generated by guided image filtering, along with an Attention-Gate-enhanced REsidual Denoising network (AG-REDCNN). Experiments demonstrate that the method effectively reduces noise and preserves structural details without requiring normal-dose CT images during training, outperforming several existing approaches.

Keywords: Low-dose CT; Image Denoising; Self-supervised Learning; Guided Image Filtering; Attention Gate; Residual Network.

1. Introduction

Computed Tomography (CT) is indispensable in modern diagnostics, but its use of ionizing radiation raises concerns, driving the need for Low-Dose CT (LDCT)[1]. LDCT reduces patient exposure at the cost of increased image noise and artifacts, which hinder diagnosis. While deep learning methods have advanced denoising, they typically require perfectly paired low-dose and normal-dose CT images for training—a major clinical limitation due to the impracticality and ethical issues of duplicate scanning[2-3].

To overcome this, we propose a novel self-supervised denoising framework. It eliminates the need for normal-dose images by using Guided Image Filtering to generate edge-preserving pseudo-labels directly from LDCT scans. Building on this, we design AG-REDCNN, an attention-enhanced network that adaptively focuses on critical anatomical structures. Evaluated on a public dataset, our method outperforms existing techniques in both quantitative metrics and visual quality, offering a practical, data-efficient solution for improving LDCT image quality.

2. Related Works

2.1. Low-Dose CT and Noise

Computed Tomography is vital for diagnosis, but its ionizing radiation necessitates Low-Dose CT. While reducing patient exposure, LDCT severely degrades image quality by introducing complex, structured-dependent noise and artifacts that are intertwined with anatomical details, making denoising uniquely challenging. Traditional methods—such as projection-domain filtering, iterative reconstruction, and image post-processing—struggle to optimally balance noise suppression with the preservation of critical clinical structures[4]. This fundamental dose-quality trade-off and the specific noise characteristics of LDCT form the basis for developing advanced, deep learning-based denoising solutions.

2.2. Convolutional Neural Network

Convolutional Neural Networks (CNNs) have become a cornerstone for Low-Dose CT denoising due to their powerful, data-driven ability to learn the complex mapping from noisy to clean images. Unlike traditional methods, CNNs can automatically extract hierarchical features to separate noise from true anatomical structures[5]. Architectures like encoder-decoders with skip connections (e.g., RED-CNN) are particularly effective, as they compress and then reconstruct images while preserving fine details. A key innovation is residual learning, where the network learns the noise residual (the difference between the noisy and clean image) rather than the clean image directly. This approach stabilizes training and improves structure preservation[6]. Furthermore, this residual framework aligns perfectly with self-supervised learning, enabling the network to be trained using only noisy LDCT images by learning the difference between the input and a generated pseudo-clean target, offering a practical solution for clinical data scarcity.

2.3. Pseudo-labelling and Guided Image Filtering

Guided Image Filtering (GIF)[7] is an edge-preserving smoothing technique. It assumes a local linear relationship between a guidance image (G) and the filtered output (O) within a small window. The linear coefficients are determined by minimizing a cost function that considers both fidelity to the input image (I) and a regularization term. A key property is that the filter's behavior adapts based on local variance: in flat regions (low variance), it performs strong averaging to suppress noise; near edges (high variance), it preserves the original structure. This makes it highly suitable for medical images where maintaining anatomical boundaries is critical. In this study, GIF is applied in a self-guided manner (where $G = I$, the noisy LDCT image itself) to generate pseudo-labels. These pseudo-labels provide approximate clean targets that retain essential structures while reducing noise, forming a reliable foundation for the subsequent self-supervised training of the deep denoising network without needing any real normal-dose CT images.

2.4. Supervised Learning

Supervised LDCT denoising relies on scarce, perfectly-paired normal-dose labels. We employ a self-supervised approach using pseudo-labels generated via Guided Image

Filtering, which preserves edges while suppressing noise[8]. A network is trained to predict this pseudo-label from the LDCT input, learning the residual noise distribution rather than merely replicating the filter. This eliminates the need for normal-dose data, providing a practical training framework, though performance is inherently limited by pseudo-label quality.

2.5. Gated Attention

The Attention Mechanism allows deep learning models to focus on task-relevant information. This is critical for LDCT denoising, as it helps the network enhance diagnostically vital regions—like lesion boundaries and organ edges—within a noisy image. This work utilizes the Attention Gate, a lightweight module that generates a spatial attention map to weight input features, amplifying important structural details while suppressing noise[9-10]. Integrated into a skip connection of an encoder-decoder network, the AG selectively propagates crucial shallow features to deeper layers. Compared to heavier mechanisms like self-attention, AG provides a better balance between detail recovery and computational efficiency, which is essential for processing high-resolution medical images[11].

3. Methods

3.1. Self-Supervised Denoising Framework Using Pseudo-Labels

The proposed method addresses the clinical unavailability of paired normal-dose CT images by adopting a self-supervised learning strategy. Instead of relying on explicit labels, it generates supervision directly from the available LDCT data itself. The core idea is to use Guided Image Filtering to create pseudo-labels from the noisy LDCT inputs. GIF is an edge-preserving filter that smooths uniform regions (suppressing noise) while preserving areas with high variance (maintaining edges and structures). This property makes it highly suitable for generating approximate clean targets that retain crucial anatomical information.

Formally, for an LDCT image I, the pseudo-label $\hat{I}$ is generated as $\hat{I} = GIF(I)$, where GIF uses the image itself as the guide. This pseudo-label is not the final output but serves as the training target. A deep neural network is then trained to predict this pseudo-label from the original LDCT input by minimizing the mean squared error between its output and I^. Crucially, the network learns the residual mapping—the difference (noise/artifacts) between the input and the pseudo-label. This allows the model to potentially outperform the simple filter by learning a more complex, data-driven denoising function. The entire framework operates using only LDCT images, eliminating the dependency on clinically scarce normal-dose data and providing a practical training paradigm. Key filter parameters (window size, regularization) are empirically chosen to balance noise suppression and detail preservation in the pseudo-labels, which is critical for the framework's effectiveness.

3.2. Self-Supervised Denoising Model Construction

The proposed self-supervised framework trains a denoising network using only Low-Dose CT images. The training target is a pseudo-label generated from each input LDCT scan via Guided Image Filtering, eliminating the need for paired normal-dose data. The network is optimized by minimizing the Mean Squared Error (MSE) between its output and this pseudo-label. Crucially, the network learns the residual mapping—the complex noise and artifact distribution—rather than merely replicating the GIF output. This allows the deep network to perform advanced feature modeling, potentially surpassing the quality of the initial pseudo-labels.

We can formulate the self-supervised model as:

$$\min_{\theta} E_l \left[\sum_i L_{MSE} \left(f^{\theta}(l_i), GIF_{r,\epsilon}(l_i) \right) \right]$$

The MSE loss is given by:

$$L_{MSE} = \frac{1}{N} \sum_{n=1}^{N} \left(f^{\theta}(l_i) - GIF_{r,\epsilon}(l_i) \right)^2$$

3.3. Self-Supervised Residual Denoising Network with Integrated Attention

The proposed AG-REDCNN enhances the established RED-CNN denoising architecture. While RED-CNN effectively preserves details via skip connections, it uniformly processes all image regions. This is suboptimal for medical images, where critical structures like lesions occupy small areas. To address this, we integrate an Attention Gate module into a key skip connection. The AG automatically learns to weight feature maps, amplifying responses from important anatomical regions while suppressing irrelevant or noisy areas. This focused attention mechanism, within the self-supervised framework, guides the network to improve detail recovery and noise suppression.

3.4. AG-REDCNN Network Design

The proposed AG-REDCNN network enhances the RED-CNN architecture by integrating an Attention Gate (AG) mechanism. The core idea is to enable the network to adaptively focus on diagnostically critical regions (e.g., organ boundaries, lesions) while suppressing irrelevant noise during the feature reconstruction process.

The AG module is embedded specifically into the third skip connection of the symmetric encoder-decoder structure. It takes feature maps from both the encoder and decoder streams, processes them through convolutions and a sigmoid activation to generate a spatial attention weight map. This map is then used to perform element-wise weighting on the features being passed via the skip connection, thereby amplifying important structural information and attenuating noise before they reach the decoder.

Compared to the baseline RED-CNN, the primary advantages of AG-REDCNN are twofold: First, the AG provides adaptive feature refinement, allowing the network to selectively enhance crucial anatomical details rather than processing all areas uniformly. Second, the AG mechanism offers a favorable balance between performance and computational efficiency, being more lightweight than global self-attention modules while effectively improving the network's focus and detail recovery capabilities for the LDCT denoising task.

4. Experiments

4.1. Datasets

Experiments were conducted using the public NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset. It contains paired abdominal CT scans: full-dose images (120 kV, 200 mAs) and simulated quarter-dose images (120 kV, 50

mAs). For training, 2,036 slices from 8 patients were used. The validation and test sets comprised 214 slices and 128 slices from one patient each, respectively. All images had a slice thickness of 3 mm. For display and comparison, the window width/level was set to -160 HU / 240 HU.

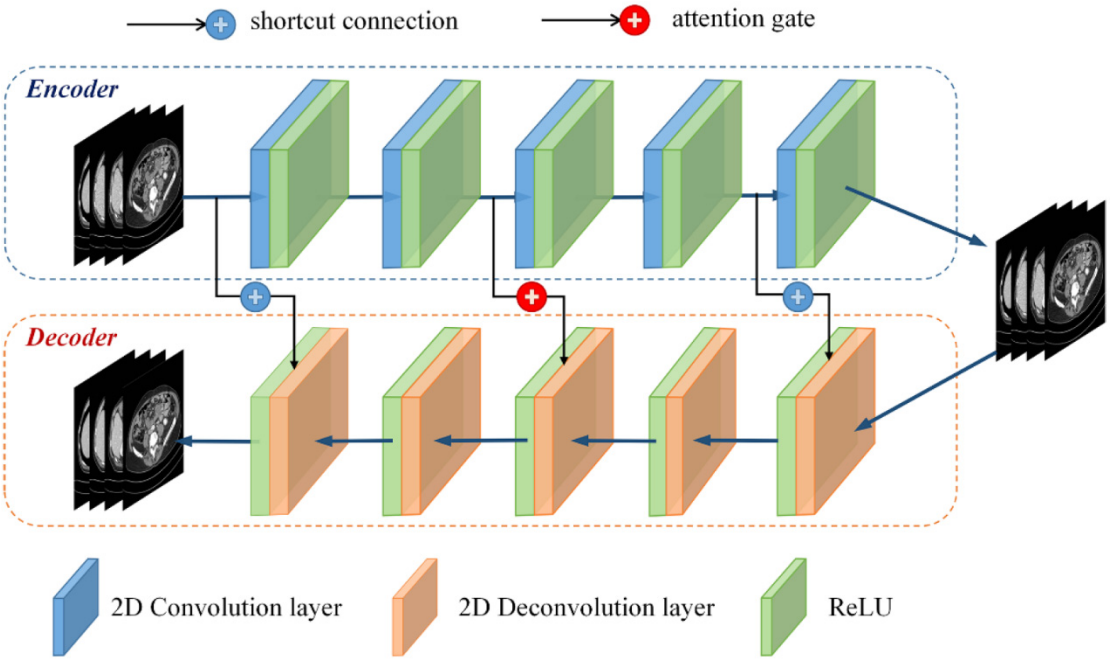

Fig 1. Network Structure of AG-REDCNN

4.2. Parameters

All experiments were performed on a workstation equipped with an NVIDIA RTX A6000 GPU using the PyTorch deep learning framework. Network parameters were initialized with a Gaussian distribution (mean=0, std=1e-2). Training used a batch size of 16, with input patches of 64×64 pixels randomly extracted from images. The initial learning rate was 1e-4, halved every 20 epochs. This configuration balanced training stability, memory usage, and the model's ability to learn local structural features.

4.3. Evaluation Criteria

Denoising performance was quantitatively evaluated using Peak Signal-to-Noise Ratio (PSNR), Root Mean Square Error (RMSE), and the Structural Similarity Index (SSIM). The proposed AG-REDCNN was compared against several representative methods: the unsupervised adversarial models WGAN-VGG and CycleGAN(which require full-dose data), the CTformer(Transformer-based, trained only on low-dose images), and the classical filter BM3D. In line with a self-supervised setup, both AG-REDCNN and CTformer were trained using only low-dose CT images.

4.4. Experiment Results Analysis

Table 1 presents the quantitative results of different methods on the test set. Overall, the proposed AG-REDCNN achieved the best performance across all three metrics, with a PSNR of 43.7051 dB, RMSE of 0.0040, and SSIM of 0.9743, outperforming WGAN-VGG, CycleGAN, CTformer, and BM3D.

Specifically, WGAN-VGG and CycleGAN yielded lower PSNR values of 42.5230 dB and 42.9943 dB, respectively. Although CTformer's PSNR (43.5352 dB) was close to ours, it was slightly inferior in RMSE and SSIM. BM3D also underperformed relative to AG-REDCNN. These results demonstrate that our method delivers superior comprehensive performance compared to existing unsupervised/self-supervised deep learning approaches and traditional post-processing techniques. The visual comparison in Figure 2 further confirms AG-REDCNN's advantage in both error reduction and structural preservation.

Table 1. Comparison of Quantitative Evaluation Results for Different Methods

Methods	PSNR	RMSE	SSIM
LDCT	40.8827	0.0056	0.9431
WGAN-VGG	42.5230	0.0046	0.9649
CycleGAN	42.9943	0.0043	0.9653
CTformer(self-supervised)	43.5352	0.0041	0.9720
BM3D	42.6850	0.0045	0.9685
AG-REDCNN	**43.7051**	**0.0040**	**0.9743**

To further compare the subjective visual quality, representative test slices are visualized in Figures 3. and 4. Overall, the proposed method demonstrates better performance in noise suppression, streak artifact removal, and structural detail preservation.

The results show that while all comparison methods improve upon the original low-dose image, each has visible shortcomings: WGAN-VGG and CycleGAN leave residual streak artifacts and noise; BM3D causes over-smoothing and loss of detail; and CTformer retains some local noise. In contrast, the proposed AG-REDCNN more effectively suppresses noise and artifacts while maintaining sharper organ boundaries and structural continuity, indicating its stronger capability for feature refinement and structural recovery.

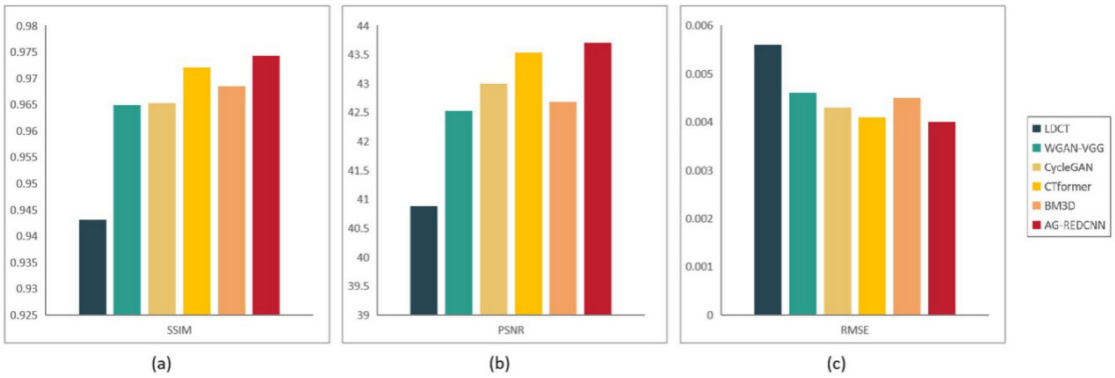

Fig 2. The Graph of Comparison of Quantitative Evaluation Results

4.5. Human Visual Analysis

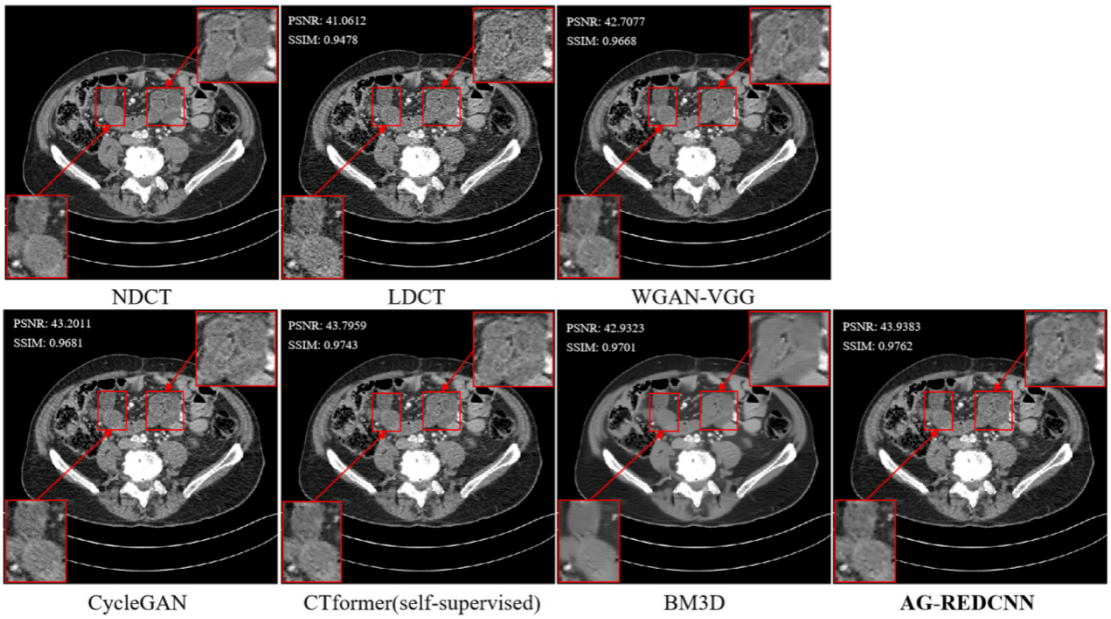

Fig 3. Denoising Results and Evaluation on Slice 1

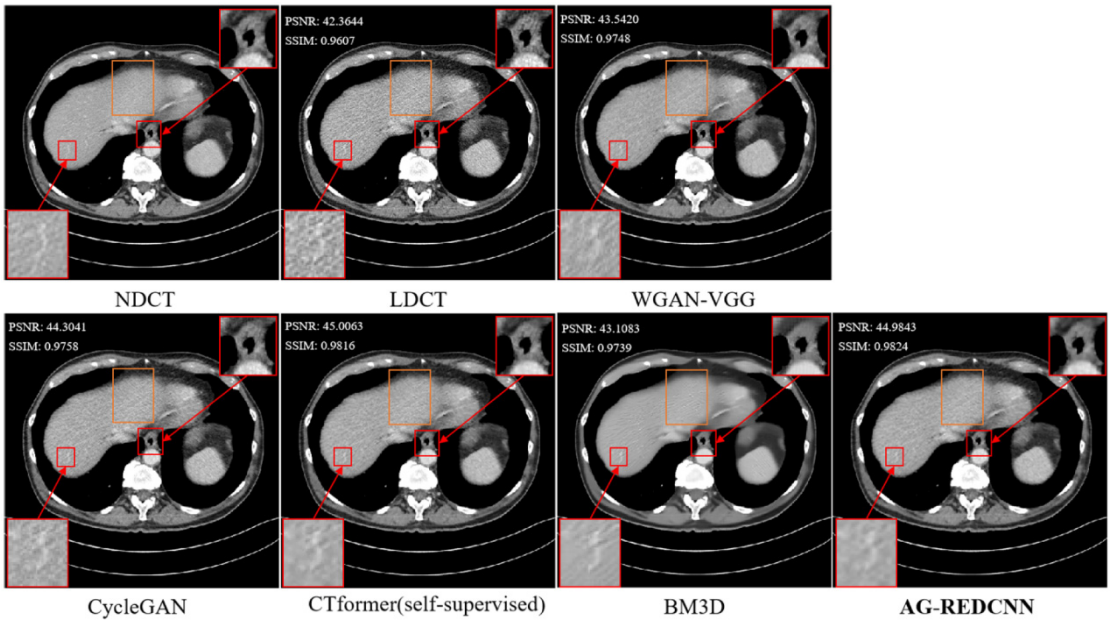

Fig 4. Denoising Results and Evaluation on Slice 2

4.6. Analysis of Absolute Error Maps

To provide a more intuitive pixel-level error analysis, Figure 5 presents the absolute error maps between the denoised results and the reference NDCT image. Brighter (yellow) areas in these maps indicate larger errors.

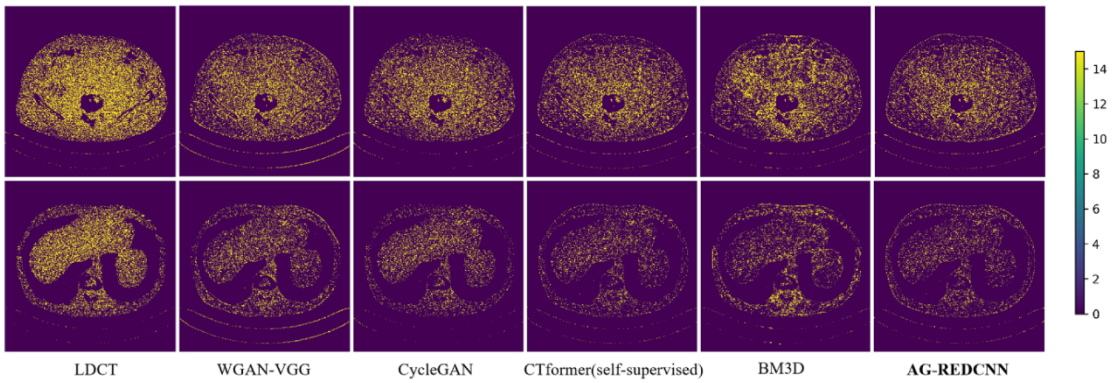

Fig 5. Comperation of Absolute Error Map

As shown, the error maps for WGAN-VGG and CycleGAN contain extensive bright regions, indicating significant residual noise. CTformer shows noticeable errors near the spine, while BM3D's map displays irregular bright patches, suggesting structural distortion from over-smoothing. In contrast, the proposed method's error map exhibits the fewest bright areas, demonstrating its superior pixel-level accuracy and structural fidelity. This visual analysis aligns with the previous quantitative and qualitative results, further verifying the effectiveness of our approach.

4.7. Ablation Study

Table 2. Comparison of Effects of different number of AG

Methods	**PSNR**	**RMSE**	**SSIM**
LDCT	40.8827	0.0056	0.9431
AG-REDCNN	43.7051	0.0040	0.9743
AG-REDCNN (2 AG blocks)	39.3448	0.0066	0.9346
AG-REDCNN (3 AG blocks)	38.4276	0.0074	0.9574

An ablation study was conducted to evaluate the effect of the number of Attention Gate modules. Models with 1, 2, and 3 AGs were compared. Both quantitative metrics and visual results demonstrate that the model with a single AG achieved the best performance, with optimal PSNR (43.7051 dB), RMSE (0.0040), and SSIM (0.9743) scores.

Increasing the AG count to 2 or 3 led to a noticeable decline in all metrics and introduced over-smoothing in visual results, causing the loss of fine structural details. This indicates that simply stacking more AG modules is detrimental. The analysis suggests that within the current self-supervised framework using GIF-generated pseudo-labels, stacking multiple AGs may excessively suppress high-frequency information during the multi-stage decoding process, destabilizing feature learning and resulting in blurred outputs. Therefore, embedding a single AG in the third skip connection provides the optimal balance between noise suppression and structural preservation.

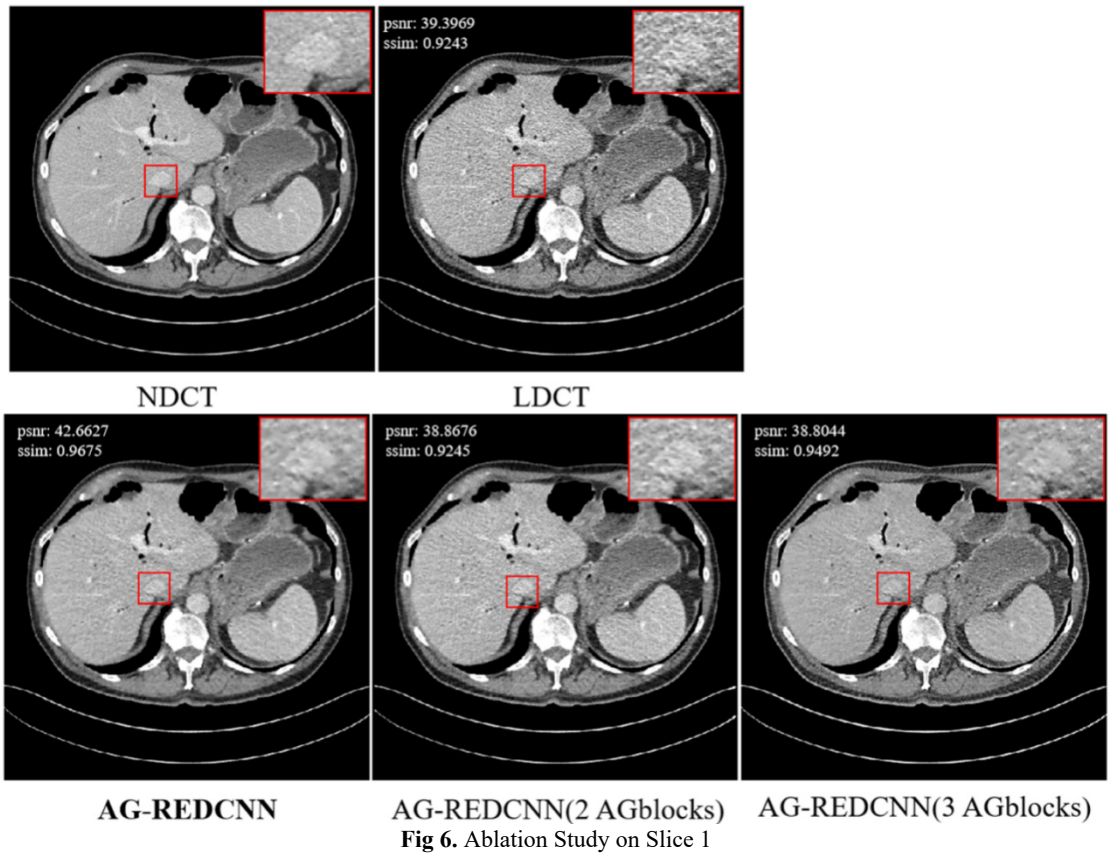

Fig 6. Ablation Study on Slice 1

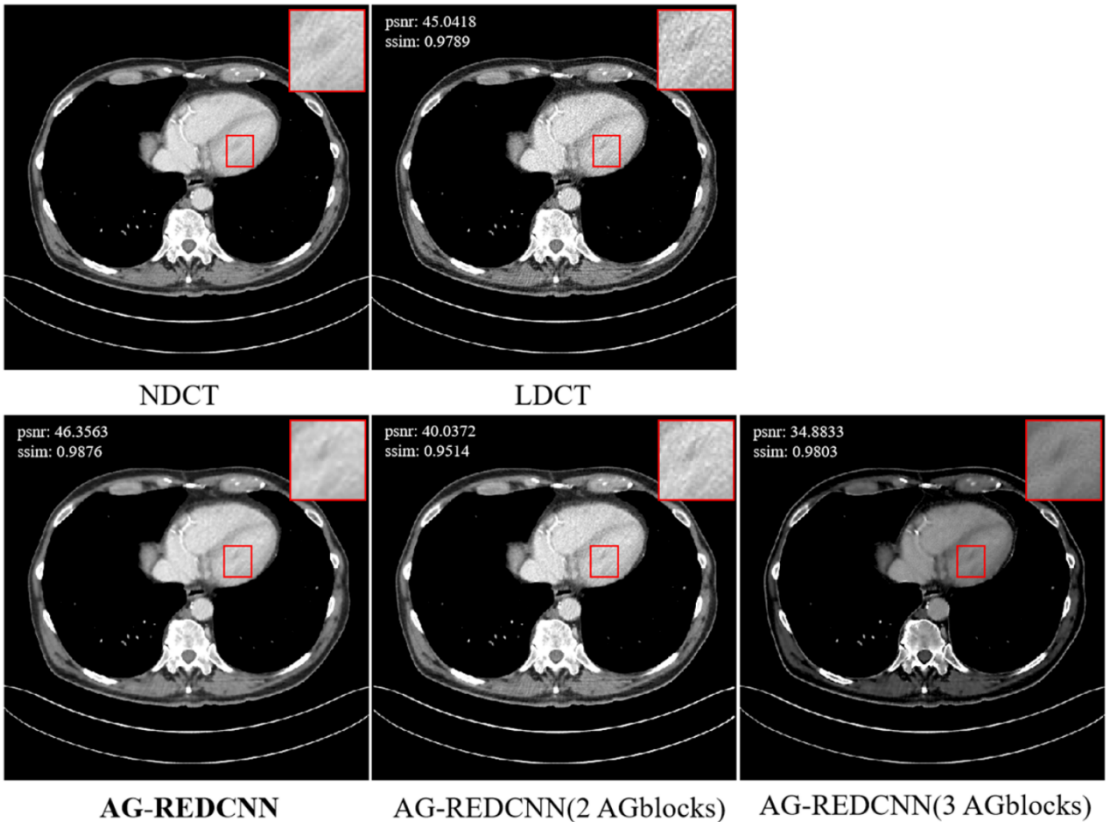

Fig 7. Ablation Study on Slice 2

4.8. Effectiveness Analysis of the Attention Gate Module

Table 3. Effectiveness of AG Module

Methods	**PSNR**	**RMSE**	**SSIM**
LDCT	40.8827	0.0056	0.9431
REDCNN(supervised)	**44.9239**	**0.0035**	**0.9797**
REDCNN (self-supervised)	43.6570	0.0040	0.9727
AG-REDCNN	43.7051	0.0040	0.9743

Building upon the previous experiments, a comparison was conducted among the supervised RED-CNN, the self-supervised RED-CNN, and the proposed AG-REDCNN to further validate the role of the AG module. The results are summarized in Table 3. As expected, the supervised RED-CNN achieved the highest scores (PSNR: 44.9239, RMSE: 0.0035, SSIM: 0.9797) but depends on paired normal-dose CT images. Under the same self-supervised framework, AG-REDCNN outperformed the self-supervised RED-CNN without AG, achieving a PSNR of 43.7051 and SSIM of 0.9743, compared to 43.6570 and 0.9727, respectively. This demonstrates that the AG module provides a consistent performance gain. These results confirm that the AG module effectively enhances feature representation and structural recovery within the self-supervised learning framework.

5. Conclusion

This study proposes a self-supervised method for Low-Dose CT denoising. It uses guided filtering to generate edge-preserving pseudo-labels from LDCT images, eliminating the need for paired normal-dose data. An Attention Gate-enhanced network is designed to focus on key anatomical structures. Experiments show the method outperforms several advanced techniques in noise suppression and detail preservation, offering a clinically practical solution.

References

[1] Borsdorf, Anja, Rainer Raupach, Thomas Flohr, and Joachim Hornegger. "Wavelet based noise reduction in CT-images using correlation analysis." IEEE transactions on medical imaging 27, no. 12 (2008): 1685-1703.

[2] Li, Zeheng, Junzhou Huang, Lifeng Yu, Yujie Chi, and Mingwu Jin. "Low-dose CT image denoising using cycle-consistent adversarial networks." In 2019 IEEE nuclear science symposium and medical imaging conference (NSS/MIC), pp. 1-3. IEEE, 2019.

[3] Zhang, Lei, Jianshe Xiong, and Yuezhong Zhou. "Edge-enhanced dense network based on attention for low-dose ct denoising." In 2023 8th international conference on image, vision and computing (ICIVC), pp. 449-454. IEEE, 2023.

[4] Huang, Jiaxin, Kecheng Chen, Jiayu Sun, Xiaorong Pu, and Yazhou Ren. "Cross domain low-dose CT image denoising with semantic information alignment." In 2022 IEEE international conference on image processing (ICIP), pp. 4228-4232. IEEE, 2022.

[5] Trung, Nguyen Thanh, Dinh-Hoan Trinh, Nguyen Linh Trung, Tran Thi Thuy Quynh, and Manh-Ha Luu. "Dilated residual convolutional neural networks for low-dose CT image denoising." In 2020 IEEE Asia pacific conference on circuits and systems (APCCAS), pp. 189-192. IEEE, 2020.

[6] Koh, Sungho, Nam Kyung Lee, Suk Kim, Seung Baek Hong, Dong Uk Kim, and Sung Yong Han. "The efficacy of low-dose CT with deep learning image reconstruction in the surveillance of incidentally detected pancreatic cystic lesions." Abdominal Radiology 48, no. 8 (2023): 2585-2595.

[7] Ma, Yaoyao, Jing Wang, Chao Xu, Yuling Huang, Minghang Chu, Zhiwei Fan, Yishen Xu, and Di Wu. "CDAF-Net: a contextual contrast detail attention feature fusion network for

low-dose CT denoising." IEEE Journal of Biomedical and Health Informatics 29, no. 3 (2024): 2048-2060.

[8] Kim, Seongjun, Byeongjoon Kim, and Jongduk Baek. "Dose-aware denoising diffusion model for low-dose CT." Physics in Medicine & Biology 70, no. 14 (2025): 145019.

[9] Wang, Shubin, Yi Liu, Pengcheng Zhang, Ping Chen, Zhiyuan Li, Rongbiao Yan, Shu Li, Ruifeng Hou, and Zhiguo Gui. "Compound feature attention network with edge enhancement for low-dose CT denoising." Journal of X-Ray Science and Technology 31, no. 5 (2023): 915-933.

[10] Wang, Zhenchuan, Minghui Liu, Xuan Cheng, Jinqi Zhu, Xiaomin Wang, Haigang Gong, Ming Liu, and Lifeng Xu. "Self-adaption and texture generation: A hybrid loss function for low-dose CT denoising." Journal of Applied Clinical Medical Physics 24, no. 9 (2023): e14113.

[11] Choi, Kihwan, and Sungwon Kim. "Statistical image restoration for low-dose CT using convolutional neural networks." In 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp. 1303-1306. IEEE, 2020.