RDA-Net: Residual Dual Attention Network with ASPP for Multi-Scale Low-Dose CT Image Denoising

Cai Yang, Huiyuan Chen*, Yingqi Zhang, and Tonghang Shangguan
School of Artificial Intelligence and Software Engineering, Nanyang Normal University, Nanyang, Henan, 473061, China
E-mail: nyyc@nynu.edu.cn, chenhuiyuan@nynu.edu.cn, zhangyingqi@nynu.edu.cn, shangguantonghang@nynu.edu.cn
*Corresponding author

Keywords: Low-dose CT, image denoising, multi-scale features, attention mechanism, residual learning

Received: July 4, 2025

Although low-dose CT (LDCT) helps to minimize radiation risks, it typically introduces noticeable noise artifacts that degrade the image clarity and hinder accurate clinical assessments. To address this problem, we introduce a new multi-level attention-based deep learning framework, RDA-Net, for efficiently suppressing noise in LDCT images and enhancing the recovery of structural details. The network, which has an encoder–decoder architecture, incorporates a composite convolutional module (Conv3×3→Conv5×5) to expand the receptive field. Temperature-enhance channel and special attention modules, which combine lightweight channel attention with temperature-regulated spatial attention to facilitate fine-grained feature modeling and noise suppression, are symmetrically embedded within the main network structure. This design effectively highlights critical regions while suppressing redundant features. To better recover fine structural details, the bottleneck layer incorporates an atrous spatial pyramid pooling module that captures contextual features through dilated convolutions at multiple scales. In addition, the network incorporates a three-level skip residual connection strategy to preserve shallow features, and also employs input-level residual learning to enhance training stability and reconstruction accuracy. Validation using the open-access LDCT dataset from the 2016 AAPM-Mayo Low Dose CT Challenge indicates that RDA-Net achieves values of 33.0235, 0.9100, and 9.1173 for the peak signal-to-noise ratio, structural similarity index measure, and root mean square error, respectively. RDA-Net demonstrates significant improvements over existing methods such as EDCNN, RED-CNN, and CTformer, and yields the best overall performance. Finally, experimental results confirm the model's strong ability to suppress noise and preserve details, underscoring its applicability in real-world scenarios.

Povzetek: Študija predstavi RDA-Net za razšumljanje LDCT, ki združi sestavljene konvolucije, temperaturno okrepljeno kanalno in prostorsko pozornost, ASPP ter večnivojske povratne povezave za ohranitev podrobnosti in stabilno rekonstrukcijo.

1 Introduction

CT imaging plays a crucial role in non-invasive clinical diagnostics, yet the ionizing radiation that it uses may present health concerns, especially for patients who need repeated scans such as pediatric monitoring and tumor follow-up examinations, where radiation exposure is of greater concern. To address this problem, low-dose CT (LDCT) technology has emerged as a solution that reduces radiation exposure. However, reducing the radiation dose often compromises the signal-to-noise ratio, resulting in increased noise and diminished detail and thereby affecting diagnostic reliability. Consequently, a key challenge in current research is how to effectively suppress image noise to improve image quality while maintaining low radiation doses.

Traditional denoising approaches, such as BM3D, Non-Local Means (NLM) filtering, sparse representation, and dictionary learning methods, primarily rely on low-rank or non-local similarity assumptions about the images. While these methods are effective under relatively simple noise conditions, they perform poorly when handling complex structures or non-uniform noise distributions that are typical of medical images. Table 1 summarizes the characteristics of these traditional methods.

The evolution of deep learning in recent years has significantly advanced convolutional neural network (CNN)-based image denoising. With strong feature representation and structure learning abilities, CNNs have opened new avenues for LDCT image quality enhancement.

Table 1: Characteristics of traditional denoising models

Model	Advantages	Disadvantages
Methods	Not dependent on projection data.	Difficult to efficiently remove image noise and artifacts.
BM3D	Effectively removes Gaussian noise; preserves fine details; no need for training data.	Weak in handling complex CT structures; poor generalization to non-Gaussian noise; sensitive to parameter settings.
Non-Local Means (NLM)	Exploits non-local self-similarity; theory is clear and interpretable.	Prone to blurring; limited denoising performance on complex structures; high computational cost.
Sparse Representation and Dictionary Learning	Learns sparse representation of image patches; capable of reconstructing structural details.	High training cost; low practicality; unstable results for LDCT image restoration.

RED-CNN is one of the earliest deep models that was applied to LDCT image denoising. By incorporating a symmetric encoder–decoder structure and a residual learning mechanism, RED-CNN improves reconstruction performance while preserving image details, effectively suppressing noise and retaining structural information. DnCNN employs deeper convolutional layers and residual mapping, enhancing adaptability to varying noise levels and achieving promising results in natural image denoising. To better preserve edges and fine details in medical images, EDCNN introduced a learnable edge-enhancement module and a composite loss function, improving the retention of image textures and structures while suppressing noise.

Although these methods enhance reconstruction while preserving details, they still face two major limitations: (1) convolution layers have a limited receptive field, which hampers the capture of multi-scale contextual information; and (2) these methods do not adequately model salient regions owing to insufficient suppression of redundant features, which restricts further improvements in their reconstruction performance.

Inspired by the recent success of transformer-based models in vision tasks, researchers are increasingly exploring the use of transformers in the context of medical imaging. CTformer is one of the earliest models that applied the transformer framework to LDCT denoising, utilizing self-attention to capture long-range dependencies and strengthen the modeling of global structural features. However, pure transformer-based models often struggle with modeling fine details and suffer from high computational complexity, which limits their applicability in real-world medical imaging scenarios.

Despite the progress made by existing approaches in LDCT denoising, challenges remain with respect to balancing noise suppression with detail preservation and achieving a trade-off between modeling capability and computational efficiency. These challenges are especially critical under ultra-low-dose imaging conditions, where it is extremely difficult to recover tissue edges, textures, and structural details from highly degraded data. Table 2 summarizes the characteristics of existing deep learning-based denoising models.

To provide a clear comparison of existing CT denoising methods, Table 3 summarizes both traditional and deep learning-based approaches along with their reported quantitative performance on the AAPM-Mayo LDCT dataset. This table highlights the improvements achieved by RDA-Net.

Table 2: Characteristics of deep learning-based denoising models.

Model	Advantages	Disadvantages
RED-CNN	Employs symmetric structure and residual learning to enhance training stability; outperforms traditional methods in denoising; simple and easy to implement.	Limited capability for recovering low-frequency structures; lacks global attention modeling; insufficient contextual understanding.
DnCNN	Utilizes deep networks with residual learning; handles varying noise levels; is widely applicable.	Lacks consideration of structural details in medical images; struggles to restore fine structures and smooth gradients.
EDCNN	Introduces edge-aware modules (such as Sobel) and composite loss functions to enhance detail preservation and edge structure recovery.	Poor performance on complex structures; limited ability to handle large-scale contexts and long-range dependencies.

CTformer	Uses a transformer for global attention modeling; captures long-range dependencies; excels at preserving detailed features.	High computational cost; large memory requirements; insufficient modeling of local structural details under low-dose conditions.

Table 3: Comparison of related work

Method	PSNR	SSIM	RMSE
BM3D	30.1201	0.8870	10.2367
NLM	29.8502	0.8805	10.4543
EDCNN	32.4052	0.9072	9.8475
RED-CNN	32.5899	0.9069	9.5776
CTformer	32.7999	0.9077	9.3699
RDA-Net (our method)	33.0235	0.9100	9.1173

To overcome the limitations of previous models, we propose a new denoising framework for LDCT images, which we call RDA-Net (Residual-enhanced Dual Attention Network). This model, which is built on a standard encoder-decoder framework [28], incorporates a multi-scale convolutional design featuring parallel 3×3 and 5×5 convolutions, while maintaining residual connections for skip links within the encoder. This strategy effectively enlarges the receptive field, improves feature representation, and better preserves image details. A lightweight yet effective dual-channel attention module, the Temperature-Enhanced Channel and Spatial Attention (TECA) module, is introduced at key stages in the network to suppress redundant responses and emphasize structural regions and thereby improve both detail preservation and noise suppression. Additionally, adaptive pooling is combined with an atrous spatial pyramid pooling (ASPP) module in the bottleneck layer, which enhances the model's ability to extract contextual cues across scales[31].

Extensive evaluations using several public LDCT datasets reveal that RDA-Net accurately restores multi-scale textures and structural features, outperforming state-of-the-art models in the peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and root mean square error (RMSE). RDA-Net's superior performance compared to multiple cutting-edge methods indicates a more effective balance between structural fidelity and noise suppression.

2 Denoising model

Deep learning approaches for LDCT image denoising focus on learning an effective image mapping function. The networks aim to restore a high-quality image from noisy LDCT input, closely approximating the visual quality of standard-dose CT images. The denoising model can be formulated as follows. Let M signify the observed low-dose image and N the ideal normal-dose (reference) image. Formula (1) defines the relationship between M and N:

$$M = \eta(N) \tag{1}$$

where η represents the process in which the normal-dose CT (NDCT) image undergoes nonlinear degradation to generate an LDCT image M corresponding to a given dose threshold. The denoising task is to find a function ϕ that satisfies formula (2):

$$arg\min_{\phi} \left|\left|\phi(M) - N\right|\right|_2^2 \tag{2}$$

where ϕ denotes the set of function parameters optimized during training to minimize the objective function.

3 The RDA-Net model

3.1 Overall architecture

RDA-Net is purpose-built for denoising tasks involving LDCT imagery. It has an encoder–decoder architecture that incorporates both multi-scale context modeling and a dual-attention enhancement strategy. The overall architecture comprises three primary modules, as illustrated in Figure 1:

1. the downsampling path, which progressively extracts hierarchical features and compresses spatial information;
2. the bottleneck layer, which incorporates the ASPP module and derives contextual features across multiple spatial dimensions;
3. the upsampling path, which restores spatial resolution while fusing shallow detail features.

The overall architecture incorporates three types of residual connections and symmetrical attention modules to ensure robust denoising performance while enhancing the preservation of fine image structures.

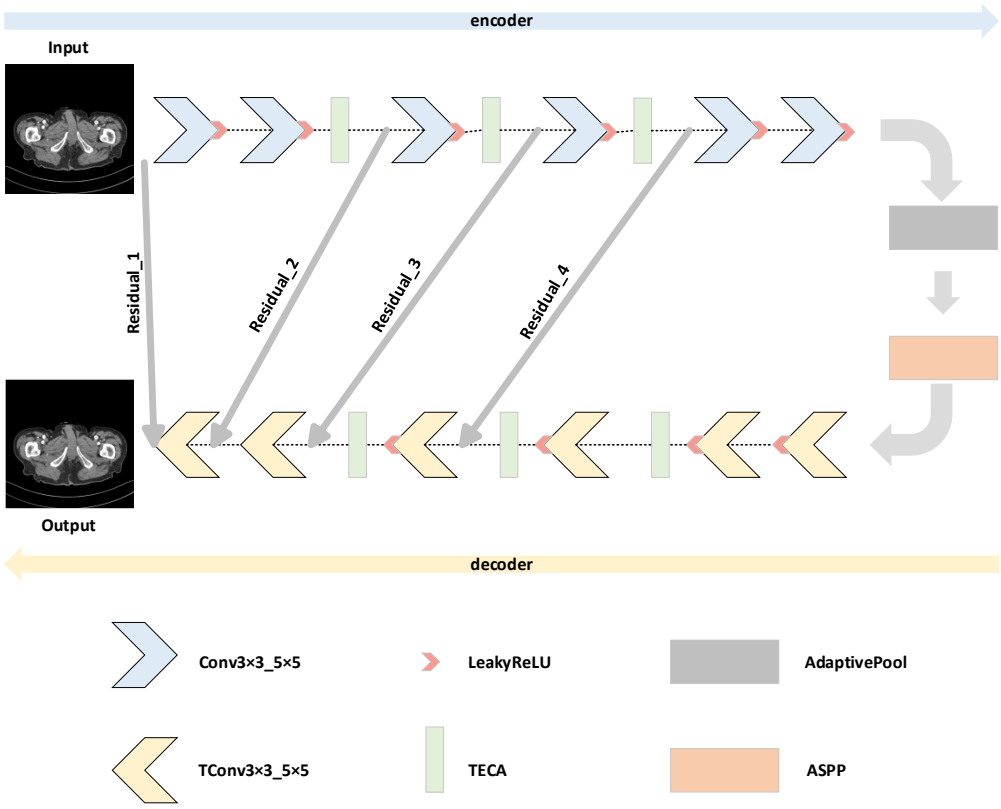

Figure 1: Overall Architecture of RDA-Net

3.2 The TECA module

3.2.1 The TECA architecture

Inspired by the CBAM module, we propose a lightweight yet more adaptive dual-attention mechanism, the temperature-enhanced channel and spatial attention (TECA) module. This module has a dual-branch structure that collaboratively optimizes feature responses along the channel and spatial dimensions. Figure 2 depicts the structural design of the TECA module. The spatial attention branch uses a learnable temperature parameter to enhance the model's adaptability and expressive capacity. This mechanism increases the network's sensitivity to detailed regions and mitigates the problem of low response that is typically caused by the standard sigmoid activation. The channel attention path employs an efficient channel attention (ECA) module to substitute the original CBAM design, which achieves a notable reduction in the number of parameters and the computational load without compromising effectiveness, thereby making the TECA module particularly suitable for resource-constrained scenarios.

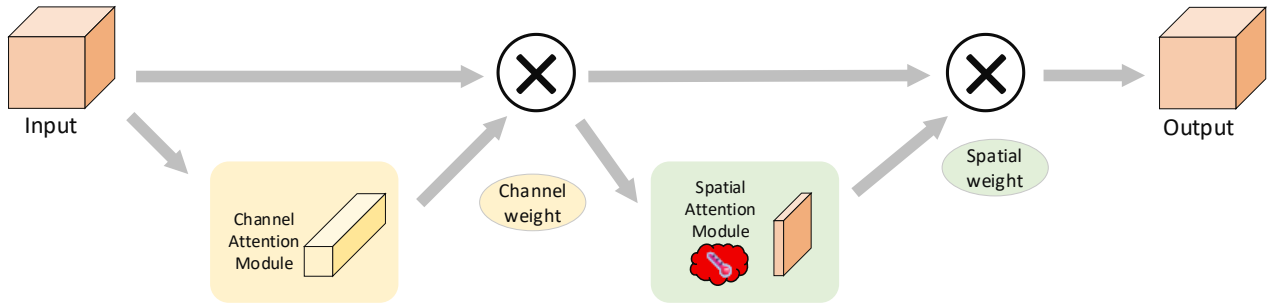

Figure 2: Structure of the TECA module

3.2.2 Channel attention

The original CBAM uses a multi-layer perceptron (MLP) to learn channel-wise dependencies, which improves performance but requires a considerable number of parameters and costly computations. To mitigate this problem, TECA incorporates the ECA mechanism, which

leverages 1D local convolutions in place of fully connected layers for efficient channel modeling.
Global average pooling is initially performed on the input features to generate a channel-wise descriptor. This descriptor is then processed by a lightweight 1D convolution with kernel size *k* applied along the channel dimension to capture local cross-channel interactions, followed by a sigmoid activation to produce channel-wise weights. The ECA module avoids dimensionality reduction and restoration steps, thereby preserving the integrity of feature representations while significantly reducing the parameter overhead. As illustrated in Figure 3, this step explicitly employs a 1D convolution—rather than a pointwise convolution—across the channels to model efficient local channel dependencies.

3.2.3 Spatial attention

In CBAM, spatial attention is generated by concatenating feature maps obtained from max pooling and average pooling, then using a convolutional layer to produce a spatial attention map. However, due to the saturation characteristic of the sigmoid activation function, the resulting attention maps can become overly smooth, making it difficult to effectively capture fine-grained spatial features. To address this limitation, the TECA module introduces a trainable temperature parameter τ to modulate the strength of the sigmoid response by smoothing the sigmoid function's output. Equation (3) is the modified activation definition that is used in the spatial attention branch:

$$\gamma_{spatial} = \sigma\left(\frac{Conv(AvgMax(x))}{\tau}\right) \tag{3}$$

The initial value of τ is set to 0.5, and it is adaptively optimized during training through gradient descent. Our implementation realizes the *Conv*(*AvgMax*(*x*)) operation by using parallel 3×3 and 5×5 convolutions on the concatenated average- and max-pooled feature maps, which improves the model's ability to capture both local details and broader contextual cues. This design introduces 71 additional parameters (19 from the 3×3 convolution, 51 from the 5×5 convolution, and 1 from τ). After scaling by the learnable parameter τ, the output is activated via the function σ to produce spatial attention weights γ_spatial, as depicted in Figure 4. This mechanism allows dynamic adjustment of the attention response range under varying input distributions, significantly improving the spatial attention module's ability to localize key regions.
Table 4 compares the CBAM and TECA modules.

Table 4: Comparison of CBAM and TECA

Component	**CBAM**	**TECA**	**Improvements by TECA**
Channel Attention	Fully connected (FC) + MLP	ECA + 1D convolution	Parameter-free; model size reduced by >98%
Spatial Attention	Fixed sigmoid activation	Temperature-modulated sigmoid	Adaptive attention scaling
Parameter Count	$\frac{C^2}{r} + 49$	$(k+1) + 71$	Reduced by at least 2 orders of magnitude
Computational Complexity	$O(\frac{HWC^2}{r})$	$O(HW + kC)$	Reduced by 30–50%

Note: *C* = number of channels; *H*, *W* = spatial dimensions; *I* = reduction ratio; *k* = ECA kernel size.

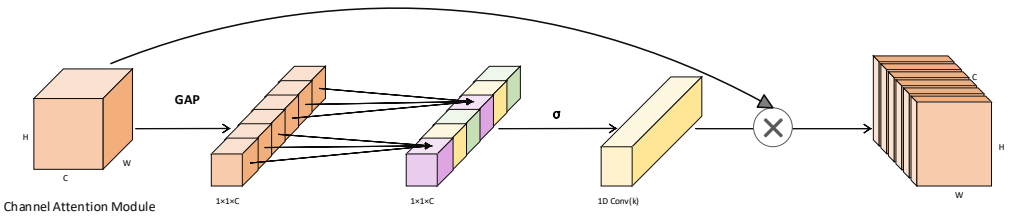

Figure 3: Channel attention structure.

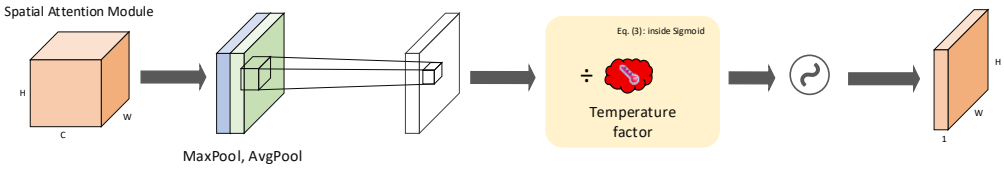

Figure 4: Spatial attention structure

3.3 Composite loss design

To enhance both visual fidelity and perceptual precision, we formulate a hybrid loss function by combining pixel-wise errors with perceptual discrepancies in the feature space. This loss function provides more comprehensive guidance for the model during training. The total loss, shown in equation (4), integrates the MSE and perceptual loss for balanced training objectives:

$$\mathcal{L}_{total} = \mathcal{L}_{MSE} + \beta \cdot \mathcal{L}_{perceptual}, \quad (4)$$

where β is a weighting hyperparameter that balances the contribution of perceptual loss. In our experiments, we set $\beta = 0.1$, which maintains pixel-level accuracy while encouraging the network to learn higher-level semantic and structural features.

By combining the two losses, the network is guided to preserve fine pixel details while also improving structural consistency and visual contrast. We verified the effectiveness and stability of this hybrid loss through practical experiments in CT image denoising.

3.3.1 Pixel-wise mean squared error loss

MSE loss quantifies pixel-wise discrepancies between predicted outputs and corresponding high-dose CT images, thereby contributing to the preservation of overall brightness and texture uniformity. Equation (5) gives the MSE loss:

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \quad (5)$$

where x_i and y_i represent the i-th pixel values in the predicted and reference images, respectively, and N is the total pixel count.

3.3.2 Perceptual loss function

A perceptual loss based on feature space is incorporated in the model to enhance its sensitivity to structural patterns and high-level semantic content. This loss is computed as the MSE between intermediate feature representations extracted from a pre-trained VGG network, as shown in equation (6):

$$\mathcal{L}_{perceptual} = \left|\left|\mu(x^*) - \mu(y)\right|\right|_2^2, \quad (6)$$

where $\mu(\cdot)$ represents features extracted from the early layers of the VGG network, x^* is the output image, and y is its high-dose counterpart. To better suit single-channel grayscale medical images, we develop a custom VGG16-based feature extractor adapted for one-channel input. The input data are normalized to ensure that their statistical properties match those of real CT image distributions with respect to the mean and variance.

While standard pretrained VGG16s (e.g., ImageNet) are trained on RGB natural images, their early features are not optimized for single-channel medical images and may emphasize irrelevant color patterns. Our custom single-channel VGG16 provides features that are better aligned with CT image characteristics. Compared to standard pretrained VGG16s, custom VGG16s can improve denoising performance by capturing more accurate structural details in LDCT images, which confirms the advantage of our tailored feature extractor for perceptual loss in medical imaging.

3.4 Atrous spacial pyramid pooling

In CNNs, the receptive field size plays a crucial role in enabling the model to extract contextual cues from image data. Conventional convolution and pooling methods typically reduce the resolution of feature maps when attempting to expand the receptive field, which negatively impacts the preservation of fine-grained structures. To address this limitation, we integrate an ASPP module into the bottleneck section of the RDA-Net framework, as illustrated in Figure 5. This design enhances the model's ability to recognize structural patterns and noise across multiple spatial scales. The ASPP structure consists of four parallel atrous convolution paths with dilation rates of 1, 6, 12, and 18. We selected these rates based on a combination of empirical testing and established practice in prior literature: The dilation rate 1 captures fine-grained, local details; the rates 6 and 12 provide mid-range context, allowing the network to integrate structural information at moderate scales; and the rate 18 captures long-range dependencies, enabling global contextual awareness without reducing feature map resolution.

The outputs from all branches are concatenated along the channel dimension and further refined with a 1×1 convolution to produce enhanced multi-scale contextual features, which are critical for effectively denoising LDCT images.

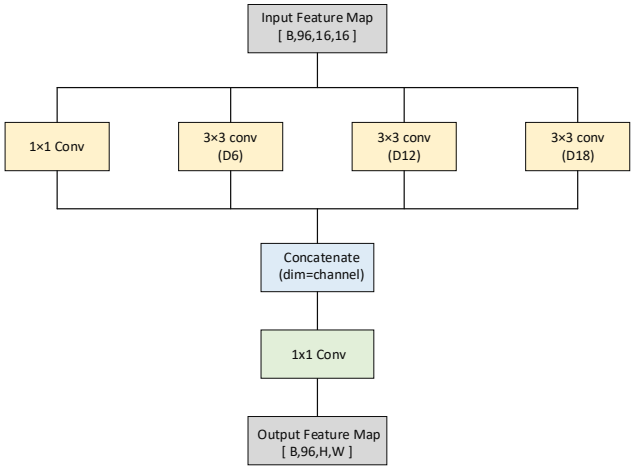

Figure 5: ASPP architecture diagram

3.5 Activation function and output range constraint

To better preserve image details while suppressing background noise, we adopted LeakyReLU as the activation function in RDA-Net. By introducing a non-zero slope on the negative axis, LeakyReLU mitigates the "dead neuron" issue commonly encountered when the traditional ReLU is used for LDCT images. All convolution and deconvolution outputs are activated using LeakyReLU(0.01), which enhances nonlinear representation capability and stabilizes the training process. Equation system (7) defines the leaky activation function:

$$\text{Leaky ReLU}(x) = \begin{cases} x & if\ x > 0 \\ \alpha x & if\ x \leq 0 \end{cases} \quad (7)$$

where α is a small constant, typically set to 0.01. During backpropagation, this function provides a non-zero gradient for inputs less than zero (unlike ReLU, which yields zero), thereby preventing jagged gradient flow and improving training stability.

To maintain the output image within the standardized grayscale interval [0, 1], the final network layer is followed by the application of a torch.clamp() function to constrain the pixel values. This clips each pixel value to the range [0,1], which ensures that the denoised image has a valid and consistent intensity distribution, avoiding outlier values and improving its clinical reliability.

3.6 Residual connections and multi-scale feature fusion

RDA-Net employs multiple residual links and a multi-scale feature fusion strategy across its feature extraction and reconstruction processes to effectively combine semantic and structural cues at various scales. During the encoding phase, residual preservation nodes (residual_1, residual_2, residual_3, and residual_4) are introduced at the first, second, third, and fourth layers, respectively. These nodes store essential intermediate feature representations. In the decoding phase, the preserved residual features are progressively added back to the corresponding deconvolution layers. This design allows the decoder to utilize intermediate features from the encoder during reconstruction, thereby enhancing both the accuracy and the consistency of feature recovery.

4 Experiments and results

4.1 Data sources and preprocessing procedures

The dataset that was utilized in this study is sourced from the 2016 AAPM-Mayo Low Dose CT Challenge, which provides abdominal CT scan data. This dataset includes a total of 2,378 slices with 3.0 mm thickness in both low-dose (quarter-dose) and full-dose formats across 10 anonymized patients. For testing, we used the data from patient L506, and the data for the other nine patients were allocated for training, maintaining a 9:1 training-to-testing ratio.

Each patient's DICOM images were first read and sorted according to the ImagePositionPatient field to reconstruct standardized 3D volume data. The raw pixel values were then converted to Hounsfield units (HUs) to unify the grayscale range. Abnormal pixel values were standardized to 0. The HU values were scaled to [0, 1] using the interval [-1024, 3072], which encompasses typical tissue density levels. After preprocessing, images were saved in .npy format and named using the convention patientID_sliceID_input/target, which facilitates streamlined data loading during network training. This preprocessing pipeline significantly improves the data consistency and reduces noise during model training, thereby providing standardized inputs for CT image denoising tasks.

4.2 Experimental setup and evaluation metrics

Training details

Framework: PyTorch on Ubuntu OS

Dataset split: The experiments used 50 abdominal CT cases from the AAPM-Mayo "Low Dose CT Image and Projection Data." The cases were divided with a 9:1 ratio between the training/validation and test collections. Specifically, 45 cases were used for the model development, among which 40 were employed for training and 5 for validation, and the remaining 5 cases were held out as the independent test set. This separation ensures fair evaluation and prevents data leakage across phases.

Input normalization: The original DICOM images were windowed to [−160, 240 HU] and then normalized to [−1, 1].

Patch extraction strategy: During training, random patches of size 64×64 were extracted from each slice (10 patches per slice). This increases data diversity, reduces

the GPU memory load, and accelerates convergence. During validation and testing, however, the full CT slices were used without patch extraction. This preserves spatial coherence in inference and avoids potential artifacts at patch boundaries.

Optimizer: Adam
Initial learning rate: 1×10−5
Batch Size: 16
Training epochs: 100
Checkpointing: Parameters were saved every 1,000 iterations, and performance was tested every 13,000 iterations.
Data pairing: Training and validation utilized paired CT images from low-dose (quarter_3mm) and normal-dose (full_3mm) scans.

Evaluation metrics

To thoroughly evaluate the model's denoising capabilities, three widely recognized image quality metrics were adopted:

PSNR: Quantifies the reconstruction error at the pixel level, with higher values implying better fidelity.
SSIM: Evaluates structural fidelity by comparing brightness, contrast, and textural patterns, with values close to 1 indicating more accurate preservation.
RMSE: Measures the global deviation between output and ground truth, with lower values indicating greater similarity.

Statistical analysis

To rigorously validate the superiority of our proposed RDA-Net over the comparison methods, we conducted paired t-tests on the PSNR, SSIM, and RMSE across the test dataset (Table 5). In addition, we report 95% confidence intervals for all metrics, thereby providing a quantitative assessment of variability and robustness. As shown in Table 6, RDA-Net achieved a PSNR of 33.0200 ± 0.14 dB, an SSIM of 0.9104 ± 0.0014, and an RMSE of 9.1160 ± 0.0142. Compared to RED-CNN and CTformer, the improvements are statistically significant (paired t-test, $p < 0.001$ for most metrics). These analyses confirm that the observed improvements are statistically significant and not due to chance, and thus demonstrate the robustness and reliability of our proposed method.

Table 5: Paired t-tests (vs RDA-Net).

Refer-ence	Compar-ison	t-stat	p-value
RDA-Net	RED-CNN	PSNR: 22.0454 SSIM: 11.2056 RMSE: -66.2808	PSNR: 2.5058e-05 SSIM: 3.6116e-04 RMSE: 3.1041e-07
RDA-Net	CTformer	PSNR: 3.8103 SSIM: 4.8107 RMSE: -22.6488	PSNR: 1.8935e-02 SSIM: 8.5809e-03 RMSE: 2.2508e-05

Table 6: Mean±CI.

Method	**PSNR mean ± 95% CI**	**SSIM mean ± 95% CI**	**RMSE mean ± 95% CI**
RED-CNN	32.5200 ± 0.1039	0.9028 ± 0.0010	9.3640 ± 0.0142
CTformer	32.7800 ± 0.1039	0.9068 ± 0.0010	9.3000 ± 0.0196
RDA-Net	33.0200 ± 0.1416	0.9104 ± 0.0014	9.1160 ± 0.0142

4.3 Comparative experiments

To critically evaluate RDA-Net's performance in LDCT image denoising, we conducted comparative studies with three leading baseline models: RED-CNN, EDCNN, and CTformer. The evaluations include both quantitative metrics and qualitative visual analysis to ensure a comprehensive comparison of noise removal and structure preservation.

As summarized in Table 7, the quantitative results demonstrate that RDA-Net consistently surpasses all three baseline methods in terms of the PSNR, SSIM, and RMSE, which underscores its strong performance in both denoising and maintaining structural integrity.

To further support the robustness of our method beyond the training setup, we also conducted compartive experiments on 1mm slice thickness data contained in the same dataset, which were not used for the model development. These thinner-slice images differ substantially from their 3 mm counterparts in terms of noise characteristics and spatial resolution, and therefore serve as a meaningful domain shift within the dataset. The consistent performance of RDA-Net on this unseen 1 mm subset indicates that it can generalize across varying reconstruction settings, even though our study intentionally focuses on the Mayo dataset.

Table 7: Comparisons of PSNR, SSIM and RMSE.

Network	PSNR(3mm)	SSIM(3mm)	RMSE(3mm)	PSNR(1mm)	SSIM(1mm)	RMSE(1mm)
LDCT	29.2489	0.8759	14.2416	24.4688	0.8246	24.6370
EDCNN	32.4052	0.9072	9.8475	25.5776	0.8366	19.2673
RED-CNN	32.5899	0.9069	9.5776	26.7513	0.8463	18.7128
CTformer	32.7999	0.9077	9.3699	27.5328	0.8416	17.2718
RDA-Net	33.0235	0.9100	9.1173	28.4181	0.8492	15.4928

For clearer visual assessment of the models' denoising effectiveness on real LDCT data, two sample slices were randomly selected from the test set, and regions of interest (ROIs) at identical locations (marked with red rectangles) were highlighted for an enlarged comparison, as shown in Figures 6–9. RDA-Net produces reconstructions that visually and quantitatively resemble NDCT scans, whereas the initial LDCT images display significant speckle noise and edge blurring. While RED-CNN and EDCNN effectively reduce overall noise, they retain noticeable residual noise in low-contrast regions. CTformer performs better in preserving fine details but still suffers from over-smoothing and blurring at tissue boundaries. In contrast, RDA-Net strikes a superior balance between noise suppression and structural preservation: It not only effectively suppresses background noise but also restores fine texture details with sharper edges and clearer depictions of anatomical boundaries. The parenchymal structures within the ROIs indicate improved anatomical fidelity, which is critical for radiologists when they assess small or low-contrast lesions, and therefore the qualitative findings suggest that RDA-Net outputs can potentially enhance diagnostic confidence. The transitions at the ROI boundaries are smooth and natural, with no visible artifacts.

In conclusion, RDA-Net delivers superior performance over both conventional CNN-based approaches and recent transformer-based models, excelling in objective scores and qualitative visual analysis, which validates the robustness of the model's architectural design.

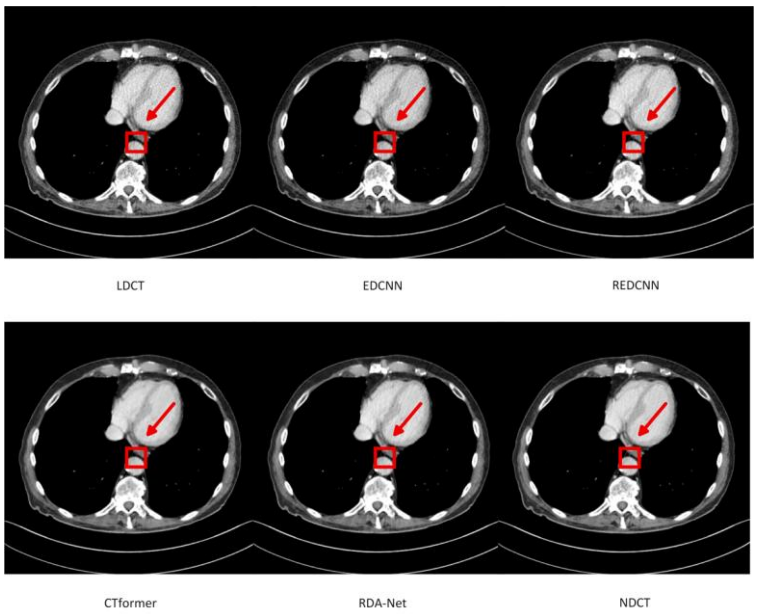

Figure 6: Comparison of slice 1 across methods.

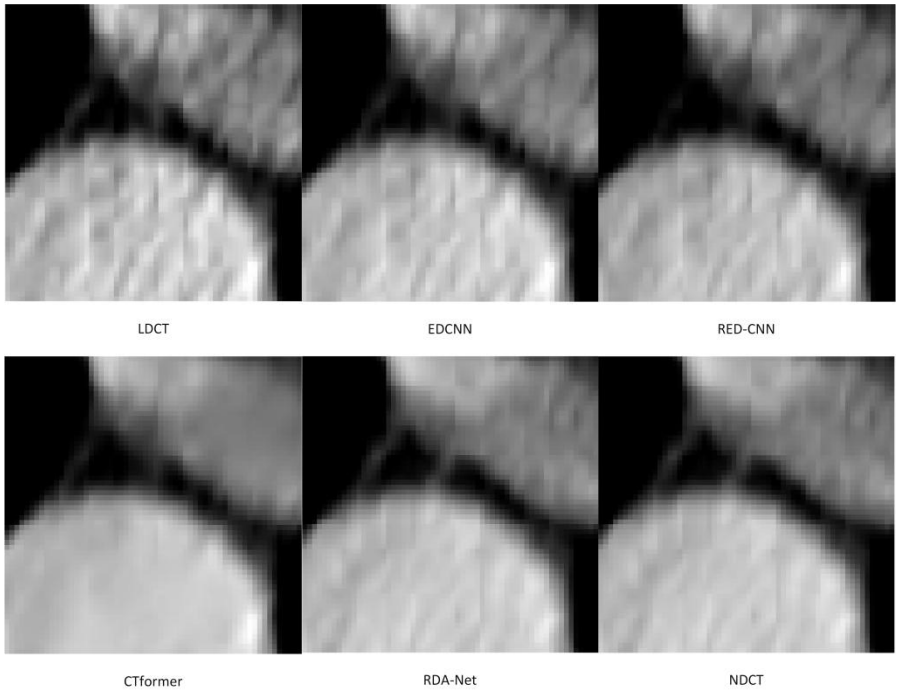

Figure 7: Enlarged ROI region from slice 1.

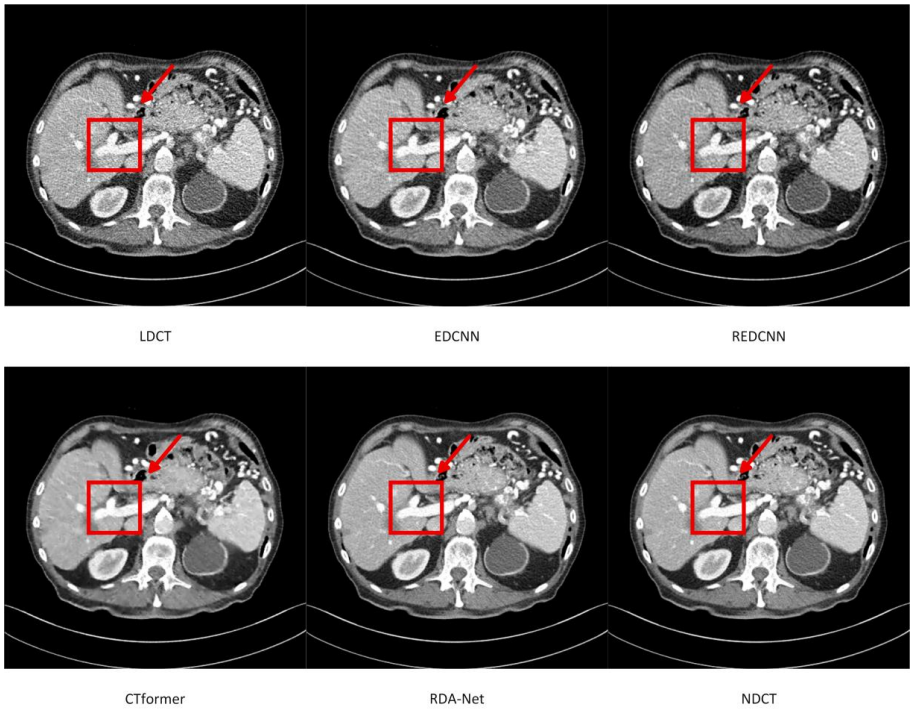

Figure 8: Comparison of slice 2 across methods.

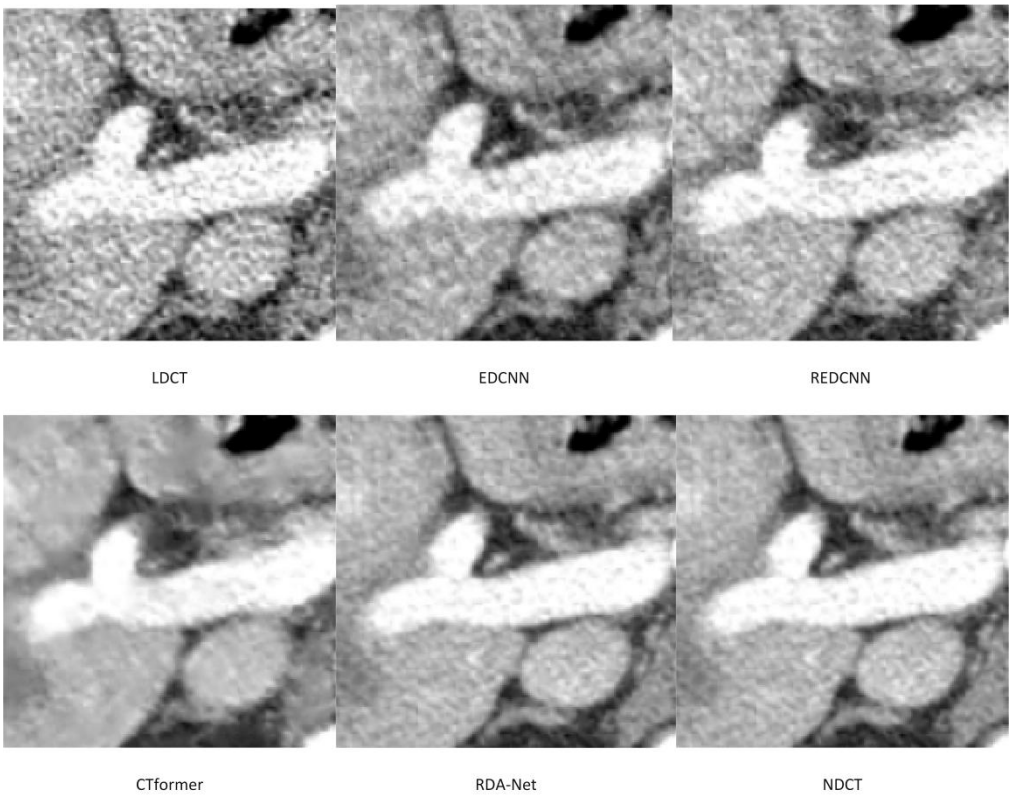

Figure 9: Enlarged ROI region from slice 2.

5 Ablation experiment

To identify the individual impact of each component of RDA-Net, we performed a series of ablation studies that focused on the network's overall structure, the TECA module configuration, and the composite loss function. The influence of each part was assessed by systematically removing or altering specific modules. The detailed configurations are presented in Tables 8–10.

Table 8: Ablation study performance comparison of the RDA-Net modules.

Number	Activation Function LeakyReLU	Hybrid Loss Function	ASPP	TECA	PSNR ↑	SSIM ↑	RMSE ↓	FLOPs(G) ↓
A0	×	×	×	×	31.5799	0.8012	10.7389	205.54
A1	√	×	×	×	31.6588	0.8635	10.5640	205.55
A2	√	√	×	×	31.8135	0.8639	10.3932	205.67
A3	√	√	√	×	32.5849	0.9018	9.5293	205.75
A4	√	√	√	√	33.0235	0.9100	9.1173	205.79

Table 8 indicates that RDA-Net achieves top scores across all evaluation indices. Incorporating the LeakyReLU activation function effectively addresses the issue of neuron inactivity during denoising. The hybrid loss function enables better retention of image details while suppressing noise. The ASPP component improves the network's capacity to extract structural patterns across different spatial scales. The significant performance improvement observed after integrating the TECA module indicates its effectiveness in leveraging contextual information for adaptive feature refinement and thereby improving the overall effectiveness and practicality of the model. It is also worth noting that adding these modules results in a modest increase in floating point operations per second (FLOPs). To further quantify the computational burden, we measured the training and inference times on a standard GPU. The total training time for 100 epochs on the AAPM-Mayo dataset was approximately 4 hours, while the average inference time per CT slice (512×512) was 45 milliseconds. These

benchmarks indicate that the additional modules introduce only a minor computational overhead. Considering the substantial improvements in the PSNR, SSIM, and RMSE, the increase in computation is acceptable, and the model's high efficiency makes it suitable for practical use.

To further confirm the effectiveness of the TECA structure and the temperature parameter, we conducted ablation experiments (reported in Table 9) that analyze the influence of channel attention (ECA), spatial attention, and the learnable temperature factor. The experimental outcomes reveal that the baseline model B0 without any attention mechanisms exhibits the weakest performance across the PSNR, SSIM, and RMSE metrics. Introducing channel attention (B1) significantly improves the reconstruction quality, which suggests that modeling local channel dependencies is beneficial for feature extraction. Adding spatial attention (B2 and B3) further improves structural recovery, with B3, which combines both channel and spatial attention, achieving better results and validating their complementary effects. Finally, model B4, which integrates the temperature factor within the spatial attention module, further optimizes the attention distribution, achieving a PSNR of 33.0235 and an SSIM of 0.9100, which demonstrates the full TECA design's superior efficacy in detailed image recovery.

In addition to the ablation experiments on the spatial and channel attention components (Table 9), we also investigated the influence of the temperature parameter τ introduced in the spatial attention mechanism. This parameter controls the sharpness of the attention distribution, thereby balancing a focus on salient structures and broader contextual information.

To assess the impact of the temperature parameter, we conducted a sensitivity study by varying τ in the range [0.5,1.0,2.0,5.0]. Table 10 shows the experimental results, which indicate that extremely small values (e.g., τ=0.5) lead to overly sharp attention maps, which reduce generalization and cause slight structural distortion, and that—conversely—very large values (e.g., τ=5.0) result in excessively smooth attention maps, which weaken detail preservation. The best performance was consistently observed when τ was set to 1.0, which achieves an effective trade-off between structural fidelity and noise suppression.

These findings confirm that while RDA-Net is robust to moderate variations of τ, an appropriately chosen value is crucial for maximizing the effectiveness of the TECA module.

Table 9: Performance impact of the TECA module design.

Number	Channel attention	Spatial attention	PSNR ↑	SSIM ↑	RMSE ↓
B0	×	×	31.8135	0.8639	10.3932
B1	√	×	32.9368	0.9090	9.2104
B2	×	√	32.9556	0.9087	9.1877
B3	√	√	33.0028	0.9093	9.1373
B4	√	√ (temperature factor)	33.0235	0.9100	9.1173

Table 10: Sensitivity study of the temperature parameter τ.

τ	PSNR mean ± 95% CI		SSIM mean ± 95% CI		RMSE mean ± 95% CI
0.5	32.867±0.210		0.903±0.005		9.412±0.128
1.0	33.023±0.185		0.910±0.004		9.117±0.110
2.0	32.745±0.198		0.905±0.004		9.285±0.116
5.0	32.102±0.256		0.896±0.006		9.678±0.142

As shown in Table 11, the combination of pixel-level loss and perceptual loss computed by our custom VGG16 model effectively captures both low-level precision and high-level perceptual features. Compared with the standard ImageNet-pretrained VGG16, the custom VGG16 better adapts to our dataset, resulting in improved reconstruction fidelity and achieving the highest scores on the PSNR, SSIM, and RMSE evaluations.

Table 11: Loss function ablation experiment configuration.

Number	$\mathcal{L}_{MSE}$	$\mathcal{L}_{perceptual}$	PSNR ↑	SSIM ↑	RMSE ↓
C1	√	×	32.9932	0.9093	9.1493
C2	×	√	32.8776	0.9090	9.1377
C3	√	√ (standard pretrained VGG)	32.9987	0.9097	9.1267
C4	√	√ (custom VGG16)	33.0235	0.9100	9.1173

In summary, the experiments demonstrate that LeakyReLU effectively prevents gradient vanishing, ASPP enhances the multi-scale representation capabilities of the model, and TECA, through its joint channel and spatial attention, significantly improves the restoration of structural details. Moreover, the use of a temperature factor in spatial attention alleviates attention saturation and enhances accuracy. The hybrid loss function (MSE +Perceptual) not only boosts the PSNR and SSIM but also significantly reduces the RMSE, which shows strong support for optimizing the subjective image quality.

6 Discussion

Our proposed RDA-Net introduces structural optimizations and performance improvements to existing CT image denoising models. Experimental findings reveal that RDA-Net delivers superior performance compared to conventional techniques such as BM3D and NLM, as well as advanced deep learning models, including EDCNN, RED-CNN, and CTformer. Quantitatively, RDA-Net achieves a PSNR of 33.0235 dB, an SSIM of 0.9100, and an RMSE of 9.1173, representing improvements of approximately 0.22 dB in PSNR, 0.0023 in SSIM, and 0.25 in RMSE over the strongest earlier method. These gains can be attributed to several key components: The ASPP module enhances multi-scale contextual feature extraction, particularly improving sensitivity to fine structural details in edges and lesion areas; the TECA module dynamically balances spatial and channel attention, allowing selective enhancement of meaningful features while suppressing noise; and the hybrid loss function (MSE + perceptual loss with a customized single-channel VGG16) further encourages preservation of both pixel-level accuracy and high-level semantic structures. Additionally, multi-stage residual skip connections ensure that critical intermediate features are retained and restored throughout the network, reducing information loss during denoising.

Despite the excellent performance of RDA-Net in various respects, some limitations remain. First, the inclusion of multiple modules—although designed to be lightweight—still increases the overall parameter count compared to the baseline model. Subsequent research could investigate compression strategies such as pruning or knowledge distillation, with the aim of retaining model effectiveness while minimizing computational overhead. Second, the temperature factor in the TECA module is currently fixed, and adaptive adjustment of this factor based on image content could further improve attention robustness. Finally, the current model has been validated on only 2D CT images; future research should extend RDA-Net to 3D volumetric reconstruction and cross-modality tasks, improving its generalizability and practical utility in clinical applications. Such a 3D extension would face several challenges, including increased GPU memory requirements due to volumetric data, the design of effective 3D convolutional kernels to capture spatial context, and handling the anisotropic resolution that is often present in clinical CT scans. Addressing these issues will be crucial for realizing efficient and high-quality 3D LDCT denoising.

Beyond the technical performance, it is also important to consider the potential clinical implications of RDA-Net. In LDCT imaging, excessive noise can obscure subtle anatomical structures and reduce radiologists' confidence regarding the detection of small or low-contrast lesions. By effectively suppressing noise while preserving fine structural details, RDA-Net may contribute to improved lesion detectability and diagnostic accuracy. Although a formal reader study was not performed in this work, previous research has demonstrated that enhanced image quality from deep learning-based denoising often correlates with better diagnostic performance.

Finally, while commercial iterative reconstruction and vendor-provided deep learning reconstruction methods are widely used in current clinical scanners, RDA-Net offers a flexible and model-agnostic framework that can be adapted to various protocols and dose levels, beyond vendor-specific implementations. Future work will include direct comparisons with these commercial tools as well as radiologist-based evaluations to more rigorously establish clinical utility.

7 Conclusion

This study introduces a lightweight residual architecture named RDA-Net. Explicitly developed to address the noise reduction challenge inherent in LDCT imaging, the model targets both visual quality and structural detail recovery by integrating attention mechanisms and multi-scale feature fusion strategies. Our proposed network incorporates a TECA attention module, an ASPP

structure, multi-level residual connections, and LeakyReLU activation, which collectively enhance the network's capacity for structural detail perception and image restoration accuracy. Quantitative analysis based on the PSNR, SSIM, and RMSE indicates that the experimental outcomes validate the model's performance advantages. In addition to quantitative improvements, our findings also suggest potential clinical benefits in terms of improved lesion visibility and diagnostic confidence, which will be validated in future reader studies. RDA-Net's performance markedly surpasses conventional denoising approaches and notable deep learning baselines such as RED-CNN and EDCNN. These results confirm RDA-Net's comprehensive advantages in structural fidelity, noise suppression, and subjective visual quality. In the future, further research will focus on model compression, a 3D extension, and cross-modality adaptation, aiming to enhance the model's applicability in real-world clinical decision support systems.

Availability of data and materials

This paper proposes a new multi-level attention-based deep learning framework for efficiently suppressing noise in LDCT images and enhancing the recovery of structural details. The specific information of the paper can be exchanged with the author.

Conflict of interest

The authors confirm that the content of this article has no conflict of interest.

Acknowledgement

This research study is supported by the General Project of Education Science Planning in Henan Province (2023YB0174), the Graduate Education Reform Project in Henan Province (2023SJGLX300Y), the Graduate Education Reform and Quality Improvement Project of Nanyang Normal University (2023ZLGC06), and the Research Projects of Nanyang Normal University (2025STP009, 2025STP010).

References

[1] Thanh D, Surya P. A review on CT and X-ray images denoising methods. Informatica, 2019, 43(2): 151-159. https://doi.org/10.31449/inf. v43i2.2179

[2] Diwakar M, Kumar M. A review on CT image noise and its denoising. Biomedical Signal Processing and Control, 2018, 42: 73-88. https://doi.org/10.1016/j.bspc.2018.01.010

[3] Hendrick RE, Smith RA. Benefit-to-radiation-risk of low-dose computed tomography lung cancer screening. Cancer, 2024, 130(2): 216-223. https://doi.org/10.1002/cncr.34855

[4] Valente D, Gentileschi PM, Valenti A, et al. Cumulative dose from recurrent CT scans: Exploring the DNA damage response in human non-transformed cells. International Journal of Molecular Sciences, 2024, 25(13): 7064. https://doi.org/10.3390/ijms25137064

[5] Sadia R T, Chen J, Zhang J. CT image denoising methods for image quality improvement and radiation dose reduction. Journal of Applied Clinical Medical Physics, 2024, 25(2): e14270. https://doi.org/10.1002/acm2.14270

[6] Brenner D J, Hall E J. Computed tomography–An increasing source of radiation exposure. New England Journal of Medicine, 2007, 357(22): 2277-2284. https://doi.org/ 10.1056/NEJMra072149.

[7] Szczykutowicz T P, Toia G V, Dhanantwari A, et al. A review of deep learning CT reconstruction: Concepts, limitations, and promise in clinical practice. Current Radiology Reports. 2022, 10: 101–115. https://doi.org/10.1007/s40134-022-00399-5.

[8] Shen Mo, Sun Rongrong and Ye Wen. Sparse Representation-Based LDCT Image Quality Assessment Using the JND Model. IEEE Access, 2025, 13: 10422-10431, 2025. https://doi.org/10.1109/ACCESS.2025.3528882

[9] Danielyan A, Katkovnik V, Egiazarian K. BM3D frames and variational image deblurring. IEEE Transactions on Image Processing, 2012, 21(4): 1715-1728. https://doi.org/10.1109/TIP.2011.2176954.

[10] Buades A, Coll B, Morel J-M. A review of image denoising algorithms, with a new one. Multiscale Modeling and Simulation, 2005, 4(2): 490-530. https://doi.org/10.1137/040616024. hal-00271141

[11] Erichson N B, Zheng P, Manohar K, et al. Sparse principal component analysis via variable projection. SIAM Journal on Applied Mathematics, 2020, 80(2): 977-1002. https://doi.org/10.1137/18M1211350

[12] Deeba F, Kun S, Ali Dharejo F, et al. Sparse representation based computed tomography images reconstruction by coupled dictionary learning algorithm. IET Image Processing, 2020, 14(11): 2365-2375. https://doi.org/10.1049/iet-ipr.2019.1312

[13] Faist D, Gnesin S, Medici S, et al. Lung lesion detectability on images obtained from decimated and CNN-based denoised [18F]-FDG PET/CT scan: An observer-based study for lung-cancer screening. European Journal of Nuclear Medicine and Molecular Imaging, 2025, (prepublish): 1-11. https://doi.org/10.1007/s00259-025-07259-2

[14] Wilson M, Chang S, Koons K E, et al. Task-specific deep learning-based denoising for UHR cardiac PCD-CT adaptive to imaging conditions and patient characteristics: Impact on image quality and clinical diagnosis and quantitative assessment. Proceedings of SPIE—the International Society for Optical Engineering, 2025, 13405. https://doi.org/10.1117/12.3047283

[15] Chen H, Zhang Y, Kalra M K, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. IEEE Transactions on Medical Imaging, 2017, 36(12): 2524-2535.

https://doi.org/10.1109/TMI.2017.2715284.
[16] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing, 2017, 26(7): 3142-3155. https://doi.org/10.1109/TIP.2017.2662206.
[17] Liang T, Jin Y, Li Y, et al. EDCNN: Edge enhancement-based densely connected network with compound loss for low-dose CT Denoising. arXiv, 2011.00139. https://doi.org/10.1109/ICSP48669.2020.9320928
[18] Trung N T, Trinh D-H, Trung N L, et al. Low-dose CT image denoising using deep convolutional neural networks with extended receptive fields. Signal, Image and Video Processing, 2022, 16(7): 1-9. https://doi.org/10.1007/s11760-022-02157-8
[19] Guernine A, Kimour M T. Optimized training for convolutional neural network using enhanced grey wolf optimization algorithm. Informatica, 2021, 45(5): 731-739. https://doi.org/10.31449/inf.v45i5.3497
[20] Carmo N, Moreira L and Backes A. Evaluating multiple combinations of models and encoders to enhance LDCT images. Signal, Image and Video Processing, 2025, 19(11): 1-11. https://doi.org/10.1007/s11760-025-04468-y
[21] Dayang W, Fenglei F, Wu Z, et al. CTformer: Convolution-free Token2Token dilated vision transformer for low-dose CT denoising. Physics in Medicine and Biology, 2023, 68(6): 1-17. https://doi.org/10.1088/1361-6560/acc000
[22] Liu Z, Park T. DMOIT: Denoised multi-omics integration approach based on transformer multi-head self-attention mechanism. Frontiers in Genetics, 2024(15): 1488683-1488683. https://doi.org/10.3389/fgene.2024.1488683
[23] Shen Z, Qin F, Gi R, et al. IDTransformer: Infrared image denoising method based on convolutional transposed self-attention. Alexandria Engineering Journal, 2025, 110: 310-321. https://doi.org/10.1016/j.aej.2024.09.101
[24] Zhu Yuting, He Qiang et al. Self-supervised noise2noise method utilizing corrupted images with a modular network for LDCT denoising. Pattern Recognition, 2025, 161:1-10. https://doi.org/10.1016/j.patcog.2024.111285
[25] Dong V, Maidment T D, Borges L, et al. Assessment of patch-based mammogram denoising methods using virtual clinical trials and deep learning: Trade-off between denoising strength and preservation of structural details. Proceedings of SPIE Medical Imaging 2022: Physics of Medical Imaging, 2022:22-31. https://doi.org/10.1117/12.2612900
[26] Du Y, Liu Y, Wu H, et al. Combination of edge enhancement and cold diffusion model for low dose CT image denoising. Biomedizinische Tech (Berlin), 2024, 70(2): 157-169. https://doi.org/10.1515/bmt-2024-0362
[27] Cui X, Guo Y, Hao W, et al. Low-dose CT image denoising based on edge prior and high-frequency sensitive feature fusion network. Signal, Image and Video Processing, 2023, 17(7): 3387-3396. https://doi.org/10.1007/s11760-023-02560-9
[28] Ferdi A, Benierbah S, Nakib A. Residual encoder-decoder based architecture for medical image denoising. Multimedia Tools and Applications, 2025, 84: 21625–21642. https://doi.org/10.1007/s11042-024-20175-1
[29] Chai X, Wei T, Chen Z, et al. LDN-RC: A lightweight denoising network with residual connection to improve adversarial robustness. Applied Intelligence, 2022, 53(5): 5224-5239. https://doi.org/10.1007/s10489-022-03847-z
[30] Boucherit I, Kheddar H. Reinforced residual encoder–decoder network for image denoising via deeper encoding and balanced skip connections. Big Data and Cognitive Computing, 2025, 9(4): 82. https://doi.org/10.3390/bdcc9040082
[31] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9), 1904-1916. https://doi.org/10.1109/TPAMI.2015.2389824.
[32] Eulig E, Ommer B, Kachelrieß M. Benchmarking deep learning-based low-dose CT image denoising algorithms. Medical Physics, 2024, 51(12): 8776-8788. https://doi.org/10.1002/mp.17379
[33] Gohla G, Estler A, Zerweck L, et al. Deep learning-based denoising enables high-quality, fully diagnostic neuroradiological trauma CT at 25% radiation dose. Academic Radiology, 2025, 32(1): 373-390. https://doi.org/10.1016/j.acra.2024.08.018
[34] Zhang J, Gong W, Ye L, et al. A review of deep learning methods for denoising of medical low-dose CT images. Computers in Biology and Medicine, 2024, 171: 108112. https://doi.org/10.1016/j.compbiomed.2024.108112
[35] Mohammed B, Anouar A, Nadjia B. Grad-CAM guided preprocessing and convolutional neural network for efficient mammogram images classification. Informatica, 2023, 47(10): 129-140. https://doi.org/10.31449/inf.v47i10.4821
[36] Poranki V K R, Rao B S. DRG-Net: Diabetic retinopathy grading network using graph learning with extreme gradient boosting classifier. Informatica, 2024, 48(2): 129-140. https://doi.org/10.31449/inf.v48i2.5078
[37] Satti P, Sharma N, Garg B. Min-max average pooling-based filter for impulse noise removal. IEEE Signal Processing Letters, 2020, 27: 1475-1479. https://doi.org/10.1109/LSP.2020.3016868
[38] Niknejad F M, Luella M, et al. Gradient-based optimization algorithm for hybrid loss function in low-dose CT denoising. 44th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2022, 3834-3838. https://doi.org/10.1109/EMBC48229.2022.9871380
[39] Ma Y, Wei B, Feng P, et al. Low-dose CT image denoising using a generative adversarial network with a hybrid loss function for noise learning. IEEE Access, 2020, 8: 67519-67529. https://doi.org/10.1109/ACCESS.2020.2986388
[40] Echine K, Darouichi A. 3D AGRes-UNet: An

advanced deep learning model for brain tumor segmentation from multimodal MRI. Multimedia Tools and Applications, 2025, (prepublish): 1-20. https://doi.org/10.1007/s11042-025-20828-9

[41] Horst K K, Zhou Z, Hull N C, et al. Radiation dose reduction in pediatric computed tomography (CT) using deep convolutional neural network denoising. Clinical Radiology, 2025, 80: 106705. https://doi.org/10.1016/j.crad.2024.09.011

[42] Gao L, Jin X, Zhang Y, et al. MSCNet: Multi-scale connected network for image denoising. Signal, Image and Video Processing, 2025, 19(6): 427-427. https://doi.org/10.1007/s11760-025-03981-4

[43] Liu S, Jaki B M M, Huang Y, et al. Multiscale hierarchy denoising method for heritage building point cloud model noise removal .npj Heritage Science, 2025, 13: 199. https://doi.org/10.1038/s40494-025-01639-5

[44] Hein D, Stevens G, Wang A, et al. PFCM: Poisson flow consistency models for low-dose CT image denoising. IEEE Transactions on Medical Imaging, 2025, 44(7): 2989-3001. https://doi.org/10.1109/TMI.2025.3558019

[45] Li X, Shang K, Butala M D, et al. Cross-modal enhanced sparse CT imaging via null-space denoising diffusion with random medical measurement embedding. Alexandria Engineering Journal, 2025,126: 565-577. https://doi.org/10.1016/j.aej.2025.04.021