https://doi.org/10.31449/inf.v49i6.7226 Informatica **49** (2025) 147–158 **147**

ICU-Net: A U-shaped Low-Dose CT Image Denoising Network Based on Codec Structure

Tonghang Shangguan, Cai Yang*, Yingqi Zhang
School of Artificial Intelligence and Software Engineering, Nanyang Normal University, Nanyang, Henan, 473061, China
E-mail: nyyc@163.com
*Corresponding author

Keywords: low-dose CT, normal-dose CT, improved ConvNext, hybrid attention, blended loss function

Received: September 27, 2024

Aims to address the issue of low-dose CT images (LDCT) introducing a considerable amount of noise due to radiation reduction, which subsequently results in a reduction in image quality and an impact on the validity of medical evaluations, a codec-based denoising model ICU-Net for LDCT images is proposed. The model utilizes an improved ConvNext block (ICB) for feature learning to extract feature data at different scales. Channel and spatial hybrid attention mechanisms (ECA) are introduced to suppress noise and artifacts. Furthermore, a blended loss function is implemented to counteract image over smoothing, which results in a denoised image that is more closely aligned with a normal-dose CT(NDCT) image. Experimental results show that the ICU-Net effectively suppresses the noise and artifacts in LDCT images. In comparison with the current denoising methods, the algorithm performs well and retains more texture details. The algorithm achieved PSNR, SSIM and RMSE values of 25.1285, 0.7217, and 43.0125 respectively, achieving the best results among the models compared.

Povzetek: Predlagan je model ICU-Net za odstranjevanje šuma iz nizkodoznih CT slik, ki temelji na izboljšanem ConvNext bloku in hibridnem mehanizmu pozornosti.

1 Introduction

Nowadays, computed tomography is extensively applied in clinical diagnostic procedures, treatment planning and disease monitoring. However, high doses of radiation are generated during the scanning process which can pose a risk to human health, so it's crucial to lower the radiation levels[1]. LDCT imaging technology has been driven by the need to reduce radiation exposure to patients. However, this technique inevitably introduces more noise and artifacts, which seriously affect the image quality and reduce the accuracy of clinical diagnosis[2]. The presence of noise and artifacts in LDCT images can be highly intricate, encompassing phenomena such as streak artifacts resulting from inhomogeneous photon distribution. To effectively address these issues, it is essential to employ robust denoising techniques. In the denoising process, it is crucial to preserve the structure and texture details of the image as much as possible, and avoid introducing new artifacts that can further reduce image quality and affect diagnostic accuracy, which is crucial for clinical diagnosis. It is a technical challenge to denoise without losing important diagnostic information, so that the denoised LDCT images are closer to NDCT images without affecting the accuracy of clinical diagnosis.

A number of algorithms have been put forth by researchers with the objective of enhancing the caliber of LDCT image[3]. The current traditional LDCT denoising methods are divided into sinogram domain filtering[4] and iterative reconstruction[5] algorithms that rely on the raw projection data, and image domain post-processing algorithms that work directly on theCT reconstructed image[6]. Since the original projection domain data are usually difficult to obtain, this creates difficulties for related research. Image post-processing methods, on the other hand, process the reconstructed CT images directly, which provides great convenience for researchers. Common methods include BM3D[7], non-local mean filtering[8], fast dictionary learning[9], and sparse projection method[10]. While traditional post-processing methods can improve image quality to a certain extent, it is often difficult to balance noise reduction and image detail retention. And it is also prone to the introduction of new noise, which can lead to image smoothing and other problems[11].

In light of the accelerated advancement of deep learning technology, a growing number of researchers are exploring its potential applications in the domain of LDCT image denoising[12]. This also provides a new solution idea for related research. And it has been proved that CNN is really effective in image denoising[13]. For example, CHEN et al. proposed RED-CNN for LDCT image denoising by combining CNN with residual learning, effectively removing artifacts and noise[14]. Building on this foundation, ZHANG et al. augmented the network's capabilities with null convolution, which led to a further improvement in the denoising effect[15].

LIANG et al. proposed the EDCNN model that combines an edge enhancement module and a composite loss function, using a trainable sobel operator to preserve the edges of CT image, which removes noise while retaining more detailed information[16]. Since the concept of Generative Adversarial Network was introduced, it has achieved remarkable results in the area of image noise mitigation[17]. YANG et al. proposed a WGAN-VGG[18]. The model employs an adversarial training method and combines VGG-19-based[19] perceptual loss and Wasserstein loss[20], thus significantly enhancing the quality of denoised images.

With the advent of the Transformer, the researchers found it to be equally good at LDCT image denoising[21]. The transformer model preserves global long-range dependencies above and below the data through a global self-attention mechanism, and performs well in a range of computer vision tasks. CTformer applies Transformer to LDCT image noise reduction for the first time[22]. TED-Net used the convolution-free Transformer with an encoder-decoder structure to maintain the consistency of the information[23]. Eformer combined Transformer's global modelling capabilities with residual learning methods to propose a residual Transformer and uses the edge enhancement module to generate CT images closer to normal dose[24]. In the context of CT image denoising, the Transformer assists the model in better understanding the global structure of the image while also focusing on local details[25]. This dual focus effectively reduces noise and preserves the detailed information within the image.

Despite the multitude of models that have been proposed for the denoising of LDCT images, the question of how to retain more detailed information while simultaneously removing noise remains a significant technical challenge[26]. Deep learning models, particularly those that prioritise global features, tend to overly smooth the image while removing noise, thereby impeding the ability to distinguish between noise and fine texture information[27]. This results in the loss of fine structure, edge and texture information. This is due to the fact that the model learns to categorise both noise and high-frequency details as components to be removed during the training process[28]. This is a common problem with current denoising algorithms. Features of above denoising models are summarized and presented in Table 1.

Table1: Features of above denoising models.

Models	Advantages	Disadvantages
Traditional methods	Not dependent on projection data.	Difficult to remove image noise and artifacts efficiently.
RED-CNN	The network structure is simple and easy to understand, less computationally demanding than other complex structures and easy to train.	Excessive smoothing brought about by the use of the MSE loss function leads to loss of detail in the resulting image.
EDCNN	Trainable Sobel operators help preserve the edges of the image structure, and the composite loss function avoids image over smoothing.	The ResNet-50 used in the multi-scale perceptual loss function was trained on a natural image dataset and is less suitable for CT images.
WGAN-VGG	Generate higher quality images that are perceptually closer to the real image.	Computationally expensive and difficult to train.
Transformer-based methods	Better at capturing global information and long-range feature interactions.	High computational complexity and memory requirements.

To tackle the problems of the above modelling approaches, the paper presents a denoising model called ICU-Net, which is based on the codec structure of U-Net[29]. To mitigate the effects of noise and artifacts on the image, the model incorporates an improved ConvNext[30] block (ICB) and a hybrid attention module[31](ECA). ICB uses a finer network structure to extract more noise and detail information, while the attention mechanism is used to focus on important edges and texture regions. Additionally, a blended loss function is employed to mitigate the problem of excessive image smoothing, which helps to retain the detailed information of the image[32]. The model retains the important texture information of the image while reducing the noise. The model is shown to perform well in the LDCT image denoising task through comparison and ablation experiments.

2 Fundamental knowledge

2.1 Denoising model

Deep learning methods in performing the LDCT image denoising process aim to get results that are close to the NDCT image progression possibilities. Suppose $M \in R^{m \times n}$ is low-dose image and $N \in R^{m \times n}$ is the normal dose image. Their relationship is shown in equation (1):

$$M = \varphi(N) \quad (1)$$

where φ is meant to illustrate the process of image decay. The process of recovering from M to N can be understood as a function F obtained by training the network such that F is approximated as σ^{-1}, and satisfies the following conditions of equation (2):

$$g = \arg\min_{F} \|F(M) - N\|_2^2 \quad (2)$$

where g denotes the value of the independent variable that minimises $\arg\min_{F} \|F(X) - Y\|_2^2$.

2.2 Denoising methods

Image denoising consists of three parts: data pre-processing, feature extraction and denoising[33]. The initial acquired dataset usually contains unusable data, such as mismatched image format, pixels that are too high, brightness that is too high or too low, and so on. At this point, image format conversion, geometric adjustment and other preprocessing operations can be used to provide more usable data for the model[34]. After acquiring the data, the image texture detail features are extracted (by convolution pooling, etc.) and the noise is separated from the background information. Finally, the denoising model is selected for denoising. The denoising method based on the traditional model extracts less information, resulting in denoising effect is not ideal. While techniques based on deep learning excel in learning the core attributes of the data by automatically learning the sample features, mapping low dimensional shallow features have high dimensional deep features[35].

CNN is the basic network for deep learning, which can automatically extract high-dimensional and effective features from an image. The CNN-based algorithm for image denoising is designed to learn the transition from noisy to clear images by being trained on a large collection of noisy and corresponding clear images, it can master the intricate process of transforming noisy images into clear ones. During training, the CNN continuously adjusts the weights using a back-propagation algorithm to reduce the discrepancy in the context of the output image as it relates to the reference image. It also performs well in LDCT images denoising, learning to map noisy images into noise-free images, effectively removing noise while preserving image detail. RED-CNN is a combination of cnn and autoencoder, which significantly improves the noise reduction of LDCT images[36]. The ICU-Net model proposed in this paper is borrowed from the autoencoder structure[37]. The convolutional layer is used as an encoder to suppress image noise and artifacts, and the pooling layer is removed to prevent the loss of detail information; the deconvolutional layer is used as a decoder to reconstruct details[38].

3 Algorithmic thought

ICU-Net is a U-structured network based on codecs, and the structure is shown in Figure 1. It is made up of four key parts: convolutional layers, deconvolutional layers, improved ConvNext block (ICB) and hybrid attention module (ECA), and composite loss function. In the figure 1, W and H are the image sizes of the input network.

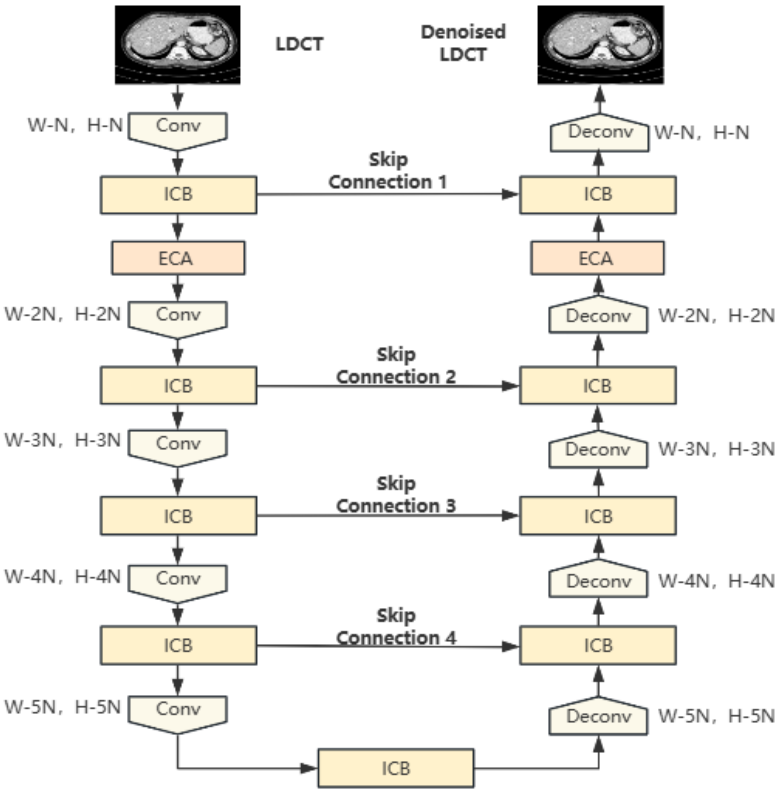

Figure 1: ICU-Net network framework structure.

Because of the encoder-decoder structure, it is first downsampled using a series of convolutional layers as a stacked encoder, which is used to suppress image noise and artifacts. More feature information is extracted as the image size gets smaller. The ICB block is added after each convolution layer, which has multiple convolution blocks of different sizes for extracting different feature data at different scales, allowing the feature extraction strategy to be flexibly adapted according to the image content and different stages of the network. The ICB block ensures that the image size remains constant while simultaneously extracting additional LDCT images of the underlying and high-level feature information for feature representation learning. This process does not affect the subsequent convolution operation. After five convolutional operations followed by one ICB Increases, the network's aptitude for recognizing detailed elements within images. Subsequently, a series of deconvolutional layers are employed as a stacked decoder for up-sampling, whereby the image information features are extracted prior to expanding the picture to recover the structural details of the image. The decoder structure is one-to-one with the encoder and undergoes an ICB block after each deconvolution operation.

A convolution with a step size of 1 and a kernel size of 5 is used to ensure that the inputs and outputs of the network are identical. Therefore, N in the diagram is 4. In order to transfer the image data from the shallow layer to the deeper layer, a shortcut connection is implemented to connect the encoder with the matching decoder[39]. This connection allows for the transfer of image data without the loss of detailed information that would otherwise occur due to the deepening of the network. In addition, this ECA block is added to the network structure to enhance the network's focus on the details and noise of the image, which reduces the sensitivity to other low information regions. The blended loss function is employed to mitigate the oversmoothing phenomenon and refine the detail of noise-eliminated images, approximating it to the original normal dose image.

4 Algorithmic implementation

4.1 ICB block

Inspired by the literature on ConvNext, an improved ConvNext block (ICB) is presented. It modernises the structure of standard CNNs to be close to the design of the Visual Transformer, making the ConvNext block exhibits a greater degree of simplicity than the ViT block. It consists of a 7×7 deep convolutional and two 1×1 convolutional layers. Besides, a layer normalisation (LN) and a nonlinear gelu activation function are added before the convolutional layers. As CT images typically encompass minor lesions and contextual information, a 7×7 convolution kernel is not equipped to extract the detailed features needed for accurate analysis[40]. The structure of the ICB block is illustrated in Figure 2. In this module, 3×3, 5×5 and 7×7 deep convolutional layers are used to extract the CT image information features sequentially, and gradually improve the sensory field to improve the recognition of small lesions on CT images, so that the extracted details are richer.

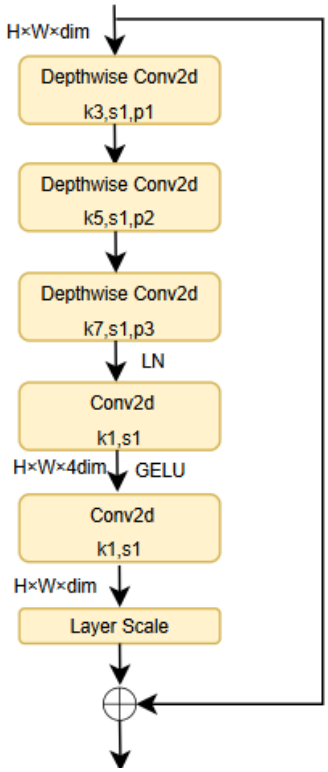

Figure 2: ICB structure.

4.2 Hybrid attention module

In classical denoising methods, the network treats the mapping of features from each channel of the image equally, which may lead to an over-focus on low-feature regions and an under-focus on high-information regions. The attention mechanism adequately extracts contextual information about the noise contaminated image according to the dependence of different feature channels and pixel information, thus suppressing the noise more effectively. Accordingly, the hybrid attention mechanism is introduced, which is designed to improve the focus on image feature information from the channel and spatial dimensions. The structure is illustrated in Figure 3.

The Channel Attention Mechanism is a technique that assigns a degree of importance to each channel feature by learning the weights associated with each feature[41]. It emphasizes more on the features with high weights reducing the attention to unimportant features based on this importance level. Initially, the spatial dimensions of the input feature map F, which is $H \times W \times C$ in size, are subjected to global max and average pooling, resulting in two feature maps of size$1 \times 1 \times C$. Then, the two results are fed into a shared MLP to learn. Finally, the MLP

output is subjected to an ADD operation, followed by a mapping process with a Sigmoid activation function, which ultimately results in a channel attention weight matrix M_C. The structure is shown in Figure 4, and the formula is shown in equation (3):

$$M_C(F) = \varphi\left(MLP(AvgPool(F)) + MLP(MaxPool(F))\right) \quad (3)$$

The Spatial Attention Mechanism identifies the locations of significant information within the feature map[42]. The first action is that performs global max and average pooling on the channel dimension of the input feature map F, with $H \times W \times C$ dimensions, to extract two feature maps of $H \times W \times 1$. Then, the two results are concatenated to obtain a feature map size of $H \times W \times 2$. Finally, a 3×3 convolution operation is executed on the spliced result to obtain a feature map size of $H \times W \times 1$, followed by a sigmoid function to get a spatial attention weight matrix M_S. The structure is shown in Figure 5 and the formula is shown in equation 4:

$$M_S(F) = \left(f^{3\times3}([AvgPool(F); MaxPool(F)])\right) \quad (4)$$

The input feature map F is computed by the channel attention and pixel attention modules to obtain a new feature map F'. The addition of hybrid attention module effectively enhances the network model's attention to image ROI, suppress irrelevant information, and enable the network model to fit better. The formula is shown in equation (5):

$$F' = M_S \otimes (M_C \otimes F) \quad (5)$$

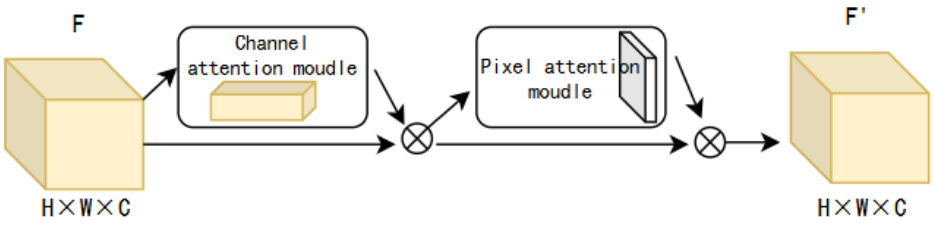

Figure 3: Hybrid Attention Module Structure

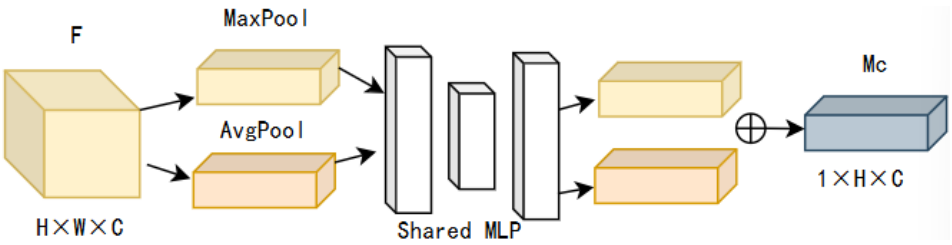

Figure 4: Channel Attention Structure.

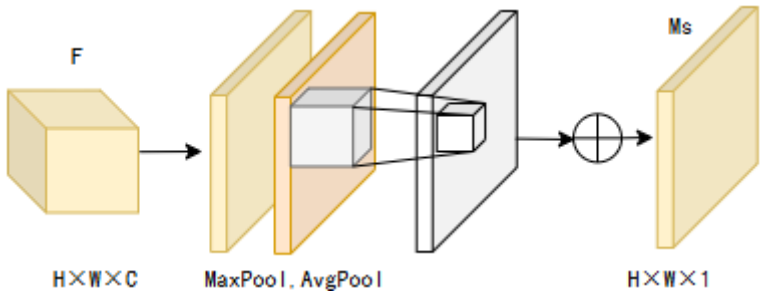

Figure 5: Spatial attention structure.

4.3 Blended loss function

A hybrid loss function comprising Huber loss and feature perception loss is employed as part of the optimisation network. Where the Huber loss is a variant of the L1 loss that provides better robustness in the handling of anomalies compared to the L2 loss, which can be expressed in equation (6):

$$L_{Huber} = \frac{1}{HWC}\sqrt{\|P_{Predict} - P_{GT}\|^2 + \varepsilon^2} \quad (6)$$

Where P_{GT} and $P_{Predict}$ are regularization constants, denote the normal dose image and the output image of the

network model, respectively. The symbol "ε" denotes the regularization constant.

The Huber loss enables the measurement of both the target and generated images at the pixel level. However, optimizing only this loss may ignore some of the detailed information in the image, which could result in an overly smooth image. In order to get better extract the spatial structure information during image transformation task, the Resnet-50 network serving as a feature extraction tool is introduced to obtain multi-scale perceptual loss[43]. Its functional formula is shown in equation (7):

$$L_{multi-p} = \frac{1}{BS}\sum_{i=1}^{B}\sum_{s=1}^{S}\left\|\phi_S(F(x_i,\theta)-\phi_S(y_i,\hat{\theta}))\right\|^2 \quad (7)$$

Where x_i is used as input image data, y_i is the target image data and B represents the count of images. F stands for the model with parameter $\theta.\emptyset$ denotes the model with Resnet-50 used to compute the perceptual loss. S is the four stages of Resnet-50. The overall representation of the composite loss function is shown in equation (8):

$$L_{Total} = L_{Huber} + \lambda_p L_{multi-p} \quad (8)$$

λ_p is the weight parameter.

5 Experiment

5.1 Datasets and evaluation indicators

The dataset used to train the network was taken from "Low Dose CT Image and Projection Data" on the TCIA website[44]. This dataset is publicly available on the web and is widely used in studies of noise reduction in LDCT images. The experiments in this paper use 50 abdominal cases, with a ratio of 9:1 within the training and test data collections. Loading the images from the DICOM format ensured spatial alignment by sorting the images according to their properties. By applying the Hounsfield unit transformation, the pixel values are normalized to reflect the true density of the different tissues. Finally, the image intensities by scaling the pixel values are normalized to the interval [0, 1] to remove the differences in feature distribution between different images.

To evaluate the denoising performance of CT images fairly and objectively, this paper adopts the PSNR SSIM and RMSE as the evaluation criteria for the quality of denoised CT images. A higher value of PSNR indicates a reduction in noise in the denoised LDCT image. The SSIM is a measure of the similarity between a denoised image and a normal dose image. A value of SSIM closer to 1 indicates a greater similarity between the two images. RMSE evaluates the denoising effect by calculating the difference between the denoised image and the original image, the smaller the RMSE value the better the denoising effect.

5.2 Model configuration

The ICU-Net is trained under the PyTorch environment, the patch size is 64×64, the batch size is 16, the number of training rounds is 100 and the initial learning rate is 1×10^{-5}. Because the number of datasets is sufficient, operations such as data expansion are no longer needed. The perceptual loss function weight λ is set to 0.01 and all the experiments in this paper are run on 34 GB NVIDIA Tesla V100 GPUs. Despite that the training process is conducted on CT image patches, the proposed network is nevertheless capable of handling images of any arbitrary size. Consequently, the test images can be fed directly into the ICU-Net without any form of decomposition.

5.3 Comparative experiment

The current mainstream LDCT image denoising methods including BM3D[6], Red-CNN[14] and EDCNN[16] are selected for comparison experiments. For the fairness of the experiment, two LDCT images were randomly selected from the final denoising result, and the comparison experiments after processing by each algorithm are shown in Figure 6 and 8, while fixed local ROIs are selected in Figure 6 and 8 for observing the details and texture information of the images after processing by each algorithm. The enlarged ROI area is shown in Figure 7 and Figure 9.

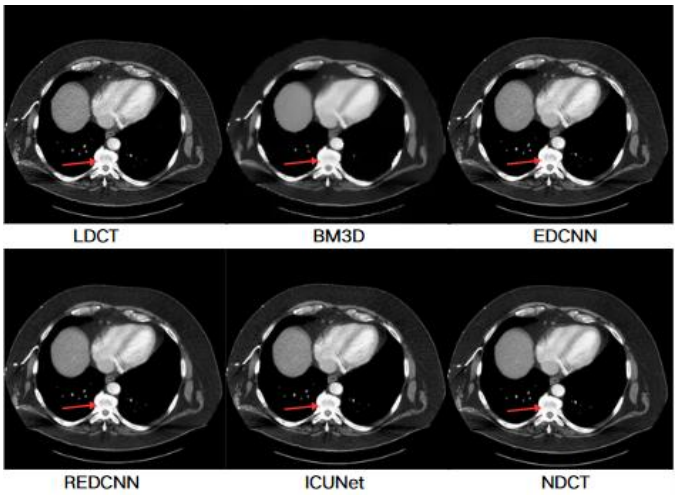

Figure 6: Abdominal transverse CT image slice 1.

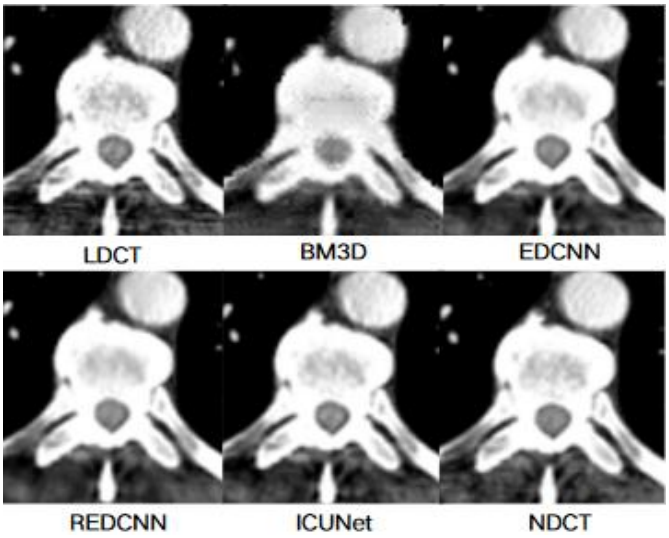

Figure 7: Amplified portion of ROI region in Slice 1.

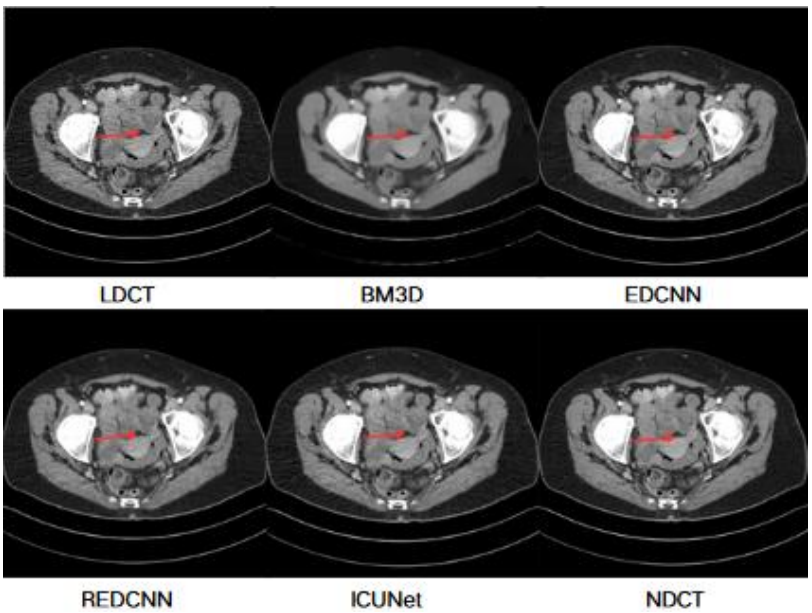

Figure 8: Abdominal transverse CT image slice 2.

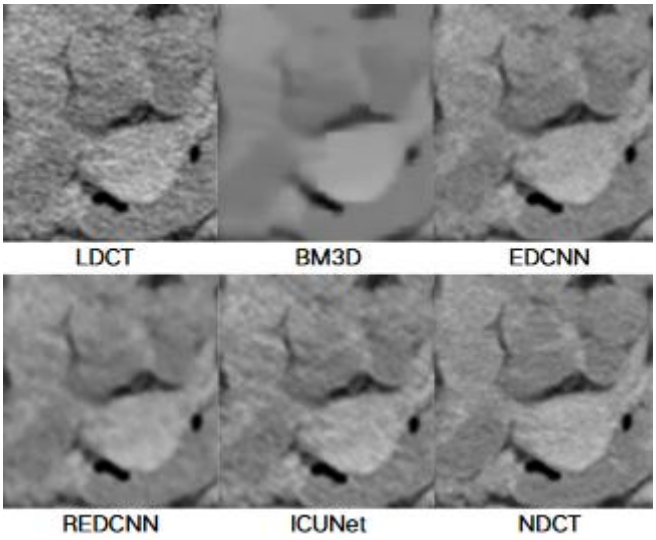

Figure 9: Amplified portion of ROI region in Slice 2.

As shown in Figure 6 to Figure 9, the outcomes of the four denoising techniques exhibit varying degrees of enhancement in comparison to the original LDCT images. From the figure, it can be seen that BM3D has limited denoising, and the denoised LDCT image loses a lot of detail information, which is worse than other methods in this paper. Red-CNN eliminates the majority of noise from LDCT, but lacks in complete texture preservation of the image due to the obvious over smoothing that occurs in the image using the MSE loss function. EDCNN improves on Red-CNN for image smoothing by using a composite function of MSE and multiscale perceptual loss. In comparison to alternative algorithms, the denoising network is capable of effectively removing noise and artifacts while simultaneously restoring greater levels of texture detail. This enables the preservation of edge information pertaining to the structure, thereby enhancing the image quality and approximating the properties of a routine dose CT image.

In order to evaluate the experimental results more fairly, the evaluation metrics achieved by each algorithm for denoising effect on the test set are shown in Table 2.

Table 2: Comparison of PSNR, SSIM, RMSE, and Training costs of different algorithms.

Network	PSNR	SSIM	RMSE	Training costs (s)
LDCT	22.4535	0.6727	47.4759	/
BM3D	20.3125	0.5745	46.8562	500
Red-CNN	24.9269	0.7179	43.3968	32620
EDCNN	24.8712	0.7154	43.3565	31238
ICU-Net	**25.1285**	**0.7217**	**43.0125**	**32730**

Since LDCT is an unprocessed low-dose CT image, there is no data on training cost. As evidenced by the data presented in the table, ICU-Net gains the highest scores for all the PSNR SSIM and RMSE metrics in comparison to other algorithms, thereby demonstrating its superior performance. Since the ICU-Net model is added with ICB and ECA modules, the training cost is higher. Although the training cost is the highest, this has a negligible effect on the model as the server arithmetic becomes more powerful.

5.4 Ablation experiment

The ICU-Net focuses on enhancing the performance of the algorithm by refining the network structure and optimizing the loss function. In order to confirm the efficacy of ICB block, ECA block and composite loss function, ablation experiments are conducted in this section. The model without ICB and ECA is taken as the baseline model M1, the model with only ICB block added in the baseline model is M2, the model with only ECA fast is M3, and the M4 model is the model ICU-Net. The results of the quantitative indicators of the experiment are shown in Table 3 and Table 4.

Table 3: Quantitative indicators for ablation experiments.

Network	PSNR	SSIM	RMSE	Training costs(s)
M1	23.1214	0.6735	43.9681	31850
M2(ICB)	24.9017	0.7113	43.1073	38243
M3(ECA)	23.1578	0.6929	42.9665	32330
ICU-Net	**25.1285**	**0.7217**	**43.0125**	**32850**

Table 4: Quantitative metrics for loss function ablation experiments.

Network	PSNR	SSIM	RMSE	Training costs(s)
ICU-Net+Huber loss	25.0649	0.7197	43.1073	31850
ICU-Net +multi-p loss	24.9269	0.7135	44.8528	38724
ICU-Net +Total loss	**25.1285**	**0.7207**	**43.1125**	**32330**

From Table 3, it can be seen that PSNR, RMSE and SSIM metrics are improved compared to LDCT images when using the ICB block and the model with ECA alone. And the metrics are the highest when the two are combined together. While obtaining better imaging quality, the cost of training has not increased dramatically. The efficacy of the network structure ICU-Net has been fully validated. As evidenced in Table 4, the application of the Huber function in isolation resulted in the lowest training cost, the utilization of the multi-p function led to the highest training cost. The combination of both functions yielded the most optimal results, with training costs situated between the aforementioned extremes, thereby corroborating the effectiveness of the composite loss function.

6 Discussion

This paper proposes an enhanced LDCT denoising algorithm, ICU-Net, which is based on autoencoder architecture. The proposed algorithm enhances the denoising capabilities of the network by optimizing the network framework and the loss function. Compared with mainstream CT image denoising algorithms, the ICU-Net has the optimal performance both in terms of visual effect and evaluation indexes, which proves the good performance of the algorithm.

The network deploys a series of convolutional layers in an encoder stack to incrementally reduce noise and artifacts from the lower to the upper layers, thus preserving the key information in the block. A series of deconvolutional layers is also used to form a stacked decoder for image reconstruction. The learning of feature representation in the network through the improved ConvNext block (ICB), assists the network in concentrating more on the overall image information, thereby enhancing its noise reduction capabilities. The fully convolutional nature of the network facilitates fine-tuning, which need not adjust the input patch size or interpolate absolute or relative positional deviations. This is advantageous for denoising tasks where the input image size varies. The fine-grained convolutional operations and deep network structure enable the identification and suppression of noise while preserving image detail and structure, resulting in higher scores on metrics such as PSNR, SSIM and RMSE.

A hybrid attention mechanism is employed to boost the network's focus on the intricacies and noise within the image and reduce its sensitivity to other regions of low information. With channel attention, the model can identify which channels contain more information about the image content and thus give more weight to these channels in the denoising process. Spatial attention, on the other hand, helps the model identify which regions of the image are more important, such as edges and textures, which require more protection during denoising. By improving feature selectivity and representation, the hybrid attention mechanism enables the model to identify and focus on key features in the image, helping the model to handle the denoising task more efficiently and thus achieve higher values in metrics such as PSNR, SSIM and RMSE. Finally, a composite loss function is used to avoid the problem of over-smoothing of the image, bringing the noise-reduced image more in line with the standard dose image.

7 Conclusion

Despite the notable success of the ICU-Net model in LDCT image denoising, there remain certain limitations and challenges to be addressed. The current model is mainly trained and tested for specific datasets, and the generalization ability of the model needs to be further validated and improved for LDCT images acquired with different devices and different scanning parameters. Further research could concentrate on expanding the data set to incorporate images from a range of devices and parameters, thus improving the model's generalisability and adaptability. Although ICU-Net performs well in denoising performance, its relatively high computational complexity may limit its feasibility in practical applications. Future research could explore more efficient network architectures or optimization strategies to reduce the demand for computational resources and increase processing speed. By conducting in-depth research in these areas, more efficient and reliable LDCT image denoising techniques are developed to better support clinical diagnosis and treatment.

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 Special Research Project on Wisdom Teaching in Undergraduate Colleges and Universities of Henan Province, the General Project of Educational Science Planning of Henan Province (Research on Talent Cultivation Mode of Software Engineering under the Integration of New Engineering and OBE Concepts 2023YB0174), Henan Province Undergraduate University Industry Education Integration Research Project, Postgraduate Education Reform

Project of Henan Province(2023SJGLX300Y), the Teaching Materials Project of New Format of New Engineering in Undergraduate Colleges and Universities of Henan Province and Project of Reform and Quality Improvement of Postgraduate Education in Nanyang Normal College (2023ZLGC06).

References

[1] Balogh, Z. A., Huszar, T., & Kis, B. J. (2020). Iterative ring artifact removal method for helical computed tomography scans. *Journal of computer assisted tomography*, vol. 44, pp. 796-805. https://www.scihub.ee/10.1097/RCT.0000000000001070

[2] Thanh, D., & Surya, P. (2019). A review on CT and X-ray images denoising methods. *Informatica*, vol. 43, no. 2. https://doi.org/10.31449/inf.v43i2.2179

[3] McLeavy, C. M., Chunara, M. H., Gravell, R. J., Rauf, A., Cushnie, A., Talbot, C. S. & Hawkins, R. M. (2021). The future of CT: deep learning reconstruction. *Clinical radiology*, vol. 76, pp. 407-415. https://doi.org/10.1016/j.crad.2021.01.010

[4] Yu M, Han M, Baek J. (2022). A convolutional neural network based super resolution technique of CT image utilizing both sinogram domain and image domain data. *Medical Imaging 2022: Image Processing*, vol. 12032, pp. 564-569. https://doi.org/10.1117/12.2611972

[5] Trotta, L. D., Matenine, D., Martini, M., Lemaréchal, Y., Francus, P., & Després, P. (2022). On the use of voxel-driven backprojection and iterative reconstruction for small ROI CT imaging. *In 7th International Conference on Image Formation in X-Ray Computed Tomography,* vol. 12304, pp. 647-653. https://doi.org/10.1117/12.2647012

[6] Koetzier, L. R., Mastrodicasa, D., Szczykutowicz, T. P., van der Werf, N. R., Wang, A. S., Sandfort, V., ... & Willemink, M. J. (2023). Deep learning image reconstruction for CT: technical principles and clinical prospects. *Radiology*, vol. 306, no. 3. https://ieeexplore.ieee.org/document/9675309

[7] Salehjahromi, M., Zhang, Y., & Yu, H. (2018). Iterative spectral CT reconstruction based on low rank and average-image-incorporated BM3D. *Physics in Medicine & Biology*, vol. 63, no. 15. https://iopscience.iop.org/article/10.1088/1361-6560/aad356/meta

[8] Kim, B. G., Kang, S. H., Park, C. R., Jeong, H. W., & Lee, Y. (2020). Noise level and similarity analysis for computed tomographic thoracic image with fast non-local means denoising algorithm. *Applied Sciences*, vol. 10, pp. 7455-7469. https://doi.org/10.3390/app10217455

[9] Deeba, F., Kun, S., Dharejo, F. A., & Zhou, Y. (2019). Sparse representation based computed tomography images reconstruction by coupled dictionary learning algorithm. *IET Image Process*, vol. 14, pp. 2365–2375. https://doi.org/10.1049/iet-ipr.2019.1312

[10] Erichson, N. B., Zheng, P., Manohar, K., Brunton, S. L., Kutz, J. N., & Aravkin, A. Y. (2020). Sparse principal component analysis via variable projection. *SIAM Journal on Applied Mathematics*, vol. 80, pp. 977-1002. https://ieeexplore.ieee.org/abstract/document/7947200

[11] Izadi, S., Sutton, D., & Hamarneh, G. (2023). Image denoising in the deep learning era. *Artificial Intelligence Review*, vol. 56, pp. 5929-5974. https://link.springer.com/article/10.1007/s10462-022-10305-2

[12] Gholizadeh-Ansari, M., Alirezaie, J., & Babyn, P. (2020). Deep learning for low-dose CT denoising using perceptual loss and edge detection layer. *Journal of digital imaging*, vol. 33, pp. 504-515. https://link.springer.com/article/10.1007/s10278-019-00274-4

[13] Kim, B., Divel, S. E., Pelc, N. J., & Baek, J. (2021). CNN-based CT denoising with an accurate image domain noise insertion technique. *In Medical Imaging 2021: Physics of Medical Imaging*, vol. 11595, pp. 1074-1079. https://doi.org/10.1117/12.2580997

[14] Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., ... & Wang, G. (2017). Low-dose CT with a residual encoder-decoder convolutional neural network. *arXiv e-prints*. https://ieeexplore.ieee.org/abstract/document/7947200

[15] Yungang, Z. H. A. N. G., Jianfeng, Y. A. N. G., & Benshun, Y. I. (2019). Improved Residual Encoder-Decoder Network for Low-Dose CT Image Denoising. *Journal of Shanghai Jiaotong University*, vol. 53, pp. 983-989. https://link.cnki.net/doi/10.16183/j.cnki.jsjtu.2019.08.014

[16] Liang, T., Jin, Y., Li, Y., & Wang, T. (2020, December). Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising. *arXiv e-prints*. https://ieeexplore.ieee.org/abstract/document/9320928

[17] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. *Communications of the ACM*, vol. 63, pp. 139-144. https://dl.acm.org/doi/abs/10.1145/3422622

[18] Ma, Y., Wei, B., Feng, P., He, P., Guo, X., & Wang, G. (2020). Low-dose CT image denoising using a

generative adversarial network with a hybrid loss function for noise learning. *IEEE Access*, vol. 8, pp. 67519-67529. https://ieeexplore.ieee.org/abstract/document/9058648

[19] Sudha, V., & Ganeshbabu, T. R. (2021). A Convolutional Neural Network Classifier VGG-19 Architecture for Lesion Detection and Grading in Diabetic Retinopathy Based on Deep Learning. *Computers, Materials & Continua*, vol. 66, pp. 827-842. https://doi.org/10.32604/cmc.2020.012008

[20] Heitz, E., Vanhoey, K., Chambon, T., & Belcour, L. (2021). A sliced wasserstein loss for neural texture synthesis. *arXiv e-prints*. https://doi.org/10.48550/arXiv.2006.07229

[21] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. (2017). Attention is all you need. *In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*, pp. 6000–6010. https://dl.acm.org/doi/10.5555/3295222.3295349

[22] Wang, D., Fan, F., Wu, Z., Liu, R., Wang, F., & Yu, H. (2023). CTformer: convolution-free Token2Token dilated vision transformer for low-dose CT denoising. *Physics in Medicine & Biology*, vol. 68, pp. 1-17. https://iopscience.iop.org/article/10.1088/1361-6560/acc000/meta

[23] Wang, Y., Liu, Q., & Lei, Y. (2024). Ted-net: Dispersal attention for perceiving interaction region in indirectly-contact hoi detection. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, pp. 5603-5615. https://dl.acm.org/doi/10.1109/TCSVT.2024.3358952

[24] Luthra, A., Sulakhe, H., Mittal, T., Iyer, A., & Yadav, S. (2021). Eformer: Edge enhancement-based transformer for medical image denoising. *arXiv e-prints*. https://doi.org/10.48550/arXiv.2109.08044

[25] Xue, T., & Ma, P. (2023). TC-net: transformer combined with cnn for image denoising. *Applied Intelligence*, vol. 53, pp. 6753-6762. https://link.springer.com/article/10.1007/s10489-022-03785-w

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

[27] Alzubaidi, L., Zhang, J., Humaidi, A.J. *et al.* (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of Big Data 8*, vol. 53, pp. 1-74. https://doi.org/10.1186/s40537-021-00444-8

[28] Ilesanmi, A. E., & Ilesanmi, T. O. (2021). Methods for image denoising using convolutional neural network: a review. *Complex & Intelligent Systems*, vol. 7, pp. 2179-2198. https://doi.org/10.1007/s40747-021-00428-4

[29] Ehab, W., & Li, Y. (2023). Performance Analysis of UNet and Variants for Medical Image Segmentation. *arXiv e-prints.* https://arxiv.org/abs/2309.13013

[30] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. *arXiv e-prints*. https://doi.org/10.48550/arXiv.2201.03545

[31] Kyro, G. W., Brent, R. I., & Batista, V. S. (2023). Hac-net: A hybrid attention-based convolutional neural network for highly accurate protein–ligand binding affinity prediction. *Journal of Chemical Information and Modeling*, vol. 63, pp. 1947-1960. https://pubs.acs.org/doi/abs/10.1021/acs.jcim.3c00251

[32] Gokcesu, K., & Gokcesu, H. (2021). Generalized huber loss for robust learning and its efficient minimization for a robust statistic. arXiv e-prints. https://arxiv.org/abs/2108.12627

[33] Xu, G., & Aminu, M. J. (2022). An efficient procedure for removing salt and pepper noise in images. *Informatica*, vol. 46, no. 2. https://doi.org/10.1016/j.inffus.2019.09.003

[34] Sagheer, S. V. M., & George, S. N. (2020). A review on medical image denoising algorithms. *Biomedical signal processing and control*, vol. 61. https://doi.org/10.1016/j.bspc.2020.102036

[35] El-Shafai, W., El-Nabi, S. A., Ali, A. M., El-Rabaie, E. S. M., & Abd El-Samie, F. E. (2024). Traditional and deep-learning-based denoising methods for medical images. *Multimedia Tools and Applications*, vol. 83, pp. 52061-52088. https://link.springer.com/article/10.1007/s11042-023-14328-x

[36] Silva, J. F., Faraggi, V., Ramirez, C., Egana, A., & Pavez, E. (2024). Understanding Encoder-Decoder Structures in Machine Learning Using Information Measures. arXiv e-prints. https://doi.org/10.48550/arXiv.2405.20452

[37] Debnath, A., & Mondal, U. K. (2024). Lossless audio codec based on cnn, weighted tree and arithmetic encoding (LACCWA). *Multimedia Tools and Applications*, vol. 83, pp. 48737-48759. https://link.springer.com/article/10.1007/s11042-023-17393-4

[38] Chefrour, A., & Drissi, S. (2023). K-CAE: Image Classification Using Convolutional AutoEncoder Pre-Training and K-means Clustering. *Informatica*, vol. 47, no. 7. https://link.springer.com/chapter/10.1007/978-3-030-31756-0_5

[39] Geirhos, R., Jacobsen, J. H., Michaelis, C., Zemel, R., Brendel, W., Bethge, M., & Wichmann, F. A. (2020). Shortcut learning in deep neural networks. *Nature Machine Intelligence*, vol. 2, pp.

665-673.
https://www.nature.com/articles/s42256-020-00257-z

[40] Greffier, J., Villani, N., Defez, D., Dabli, D., & Si-Mohamed, S. (2023). Spectral CT imaging: technical principles of dual-energy CT and multi-energy photon-counting CT. *Diagnostic and Interventional Imaging,* vol. 104, pp. 167-177. https://doi.org/10.1016/j.diii.2022.11.003

[41] Pecoraro, R., Basile, V., & Bono, V. (2022). Local multi-head channel self-attention for facial expression recognition. *Information*, vol. 13, pp. 1-17. https://doi.org/10.3390/info13090419

[42] Punn, N. S., & Agarwal, S. (2022). RCA-IUnet: a residual cross-spatial attention-guided inception U-Net model for tumor segmentation in breast ultrasound imaging. *Machine Vision and Applications*, vol. 33, pp. 1-10. https://link.springer.com/article/10.1007/s00138-022-01280-3

[43] Koonce, B., & Koonce, B. (2021). ResNet 50. *Convolutional neural networks with swift for tensorflow*, pp. 63-72. https://link.springer.com/chapter/10.1007/978-1-4842-6168-2_6

[44] Moen, T. R., Chen, B., Holmes III, D. R., Duan, X., Yu, Z., Yu, L., ... & McCollough, C. H. (2021). Low-dose CT image and projection dataset. *Medical physics*, vol. 48, pp. 902-911. https://doi.org/10.1002/mp.14594