Circuits, Systems, and Signal Processing (2024) 43:2533–2559
https://doi.org/10.1007/s00034-023-02575-0

Low-Dose CT Image Denoising with a Residual Multi-scale Feature Fusion Convolutional Neural Network and Enhanced Perceptual Loss

Farzan Niknejad Mazandarani[1] · Paul Babyn[2] · Javad Alirezaie[1,3]

Received: 28 June 2023 / Revised: 16 November 2023 / Accepted: 18 November 2023 /
Published online: 29 December 2023

Abstract

Computed tomography (CT) stands as a pivotal medical imaging technique, delivering timely and reliable clinical evaluations. Yet, its dependence on ionizing radiation raises health concerns. One mitigation strategy involves using reduced radiation for low-dose CT (LDCT) imaging; however, this often results in noise artifacts that undermine diagnostic precision. To address this issue, a distinctive CT image denoising technique has been introduced that utilizes deep neural networks to suppress image noise. This advanced CT image denoising network employs an attention mechanism for the feature extraction stage, facilitating the adaptive fusion of multi-scale local characteristics and channel-wide dependencies. Furthermore, a novel residual block has been incorporated, crafted to generate features with superior representational abilities, factoring in diverse spatial scales and eliminating redundant features. A unique loss function is also developed to optimize network parameters, focusing on preserving structural information by capturing high-frequency components and perceptually important details. Experimental results demonstrate the effectiveness of the proposed network in enhancing the quality of LDCT images.

✉ Javad Alirezaie
javad@torontomu.ca

Farzan Niknejad Mazandarani
fniknejad@torontomu.ca

Paul Babyn
paul.babyn@saskatoonhealthregion.ca

[1] Department of Electrical and Computer Engineering, Toronto Metropolitan University, 350 Victoria Street, Toronto, ON M5B2K3, Canada

[2] Department of Medical Imaging, University of Saskatchewan and Saskatoon Health Region, Saskatoon, SK, Canada

[3] Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada

Birkhäuser

Keywords CT image denoising · Deep learning · U-Net · Attention mechanism · Long range channel interdependence · Perceptual loss

1 Introduction

Computed tomography (CT) is a crucial medical imaging technique for visualizing internal body structures through cross-sectional images, offering several advantages, including high spatial resolution and fast image acquisition. These attributes enable clinicians to visualize internal anatomical structures with exquisite detail, facilitating accurate diagnoses and timely interventions. However, the use of CT imaging comes with a trade-off as it relies on X-rays, which can alter cellular and molecular processes. Hence, the CT imaging always involves the additional health risk specially when substantial radiation dose is employed [5, 9]. Given these risks, several clinical studies [13, 34] have been conducted based on the principle of as low as reasonably achievable (ALARA) to estimate effective dose.

Hence, different dose reduction strategies, such as fixed tube current (technique charts), tube current (mA) modulation and automatic exposure control (AEC) have been proposed [32]. However, these strategies lead to a decrease in the signal-to-noise ratio of the acquired CT images and reduce their visual qualities. In order to alleviate this issue, researchers have advanced various image denoising methods specialized for CT imaging. While researchers have made strides in improving the signal-to-noise ratio in LDCT images through various denoising methods, another pivotal development in this arena has been the integration of computer-aided diagnosis (CAD) systems [2]. These systems have become an integral part of medical image processing, particularly in the early detection of various diseases like leukemia and other cancers.

There are three categories of CT image denoising methods: image projection domain filtration, iterative reconstruction, and post-processing-based schemes. Image projection domain filtration methods initiate by assessing the raw sinogram data, which inherently has elevated noise due to the reduced X-ray photon counts in low-dose images. Subsequently, this data often undergoes a transformation, such as a Fourier or wavelet transform [25], enabling noise and signal separation in the frequency domain. Filters, like the Wiener filter, are then strategically applied to attenuate frequencies predominantly associated with noise, preserving those linked with genuine anatomical information [46]. Some advanced filtering techniques also employ adaptive filters, adjusting based on local noise and signal properties [42]. Once filtering concludes, an inverse transformation returns the sinogram to its spatial domain, but now with diminished noise. This refined sinogram subsequently undergoes standard CT reconstruction processes to produce the final image. As further developments in this field, the authors [30] propose a bilateral filtering method integrated with the CT noise model. In [21], the authors introduce a closed-form statistical model of sinogram data to address the statistical bias issue.

Unlike filtering-based methods where raw data are initially filtered out, resulting in the introduction of diffused noise, iterative reconstruction methods start with an initial image estimation that is continuously refined. During each iteration, a projected sinogram from the current image estimate is compared with the acquired noisy sinogram. The resulting discrepancy, or residual, informs the subsequent image adjustments. To avoid overfitting and amplify true anatomical details, various regularization techniques are employed. Total variation (TV) [4, 24, 40], dictionary learning [16, 35, 49] and non-local means [52, 53] are some current popular image priors that have been implemented in clinical scanners. These regularizations may encompass smoothness constraints, edge preservation, or statistical noise models, which characterize the inherent noise in low-dose imaging. The iterative process persists until well-defined convergence criteria are met, ensuring a balance between noise reduction and image fidelity. The incorporation of machine learning strategies in some advanced iterative methods further enhances their ability to discriminate between genuine anatomical features and noise, making them a pivotal tool in the pursuit of improved image quality in low-dose CT (LDCT) images [50]. However, these algorithms require designing a set of handcrafted regularization terms, and their complexity leads to longer reconstruction times.

Unlike the above-mentioned categories of CT image denoising, techniques based on post-processing, e.g., wavelet filtering [27, 36, 39], non-local means [19, 28], and total variation-based schemes [8, 37, 55] aim at reducing the noise in LDCT images after image reconstruction, and therefore, they do not need to process the raw data. This makes the use of such algorithms suitable for embedding into CT imaging systems. However, due to the numerous assumptions used in the development of these techniques, such as estimating the distribution of the noise, the images obtained may suffer from over-smoothing.

Recent trends in deep learning-based methods have led to a proliferation of studies in computer vision and medical imaging. The essence of these methodologies lies in their ability to model intricate nonlinear relationships within the data, a task that often proves challenging for traditional algorithms. The continuous evolution and fine-tuning of neural networks, alongside expanding LDCT datasets, signify that deep learning-based denoising techniques are not merely an incremental advancement, but a paradigm shift, cementing their place at the forefront of modern LDCT image processing. Building on this, past research has introduced a variety of supervised [6, 7, 11, 17, 20, 29, 43, 47, 51, 56] and unsupervised [18, 57] deep learning approaches to LDCT denoising. While unsupervised LDCT denoising methods offer practical advantages, such as not requiring paired samples, their clinical application remains limited due to less capable denoising currently.

Deep convolutional neural networks (DCNNs) facilitate the design of low-dose CT image denoising methods and generate images with high quality. For example, the authors [7] propose the residual encoder-decoder convolutional neural network (RED-CNN) that employs convolutional layers and skip connections in order to form an autoencoder-based CT image denoising model. To take the advantage of using the larger receptive field for the feature extraction process, the method of denoising convolutional neural network (DnCNN) [56] uses a deeper CNN that employs 17 convolutional layers. Further, in [11], the dilated residual learning network, referred to as the DRL, uses dilated convolutional layers and pre-defined Sobel operations to increase the receptive field and extract intricate details better at the image boundaries.

The utilization of multi-scale feature extraction has recently demonstrated excellent performance in obtaining useful features from CT images. The network proposed in [47] extracts the multi-scale features from CT images using a set of convolution operations with various kernel sizes. Further, the authors [17] modify the RED-CNN by incorporating the multi-scale convolutional layers into the network architecture. In addition, the study [43] introduces a domain-adaptive denoising network (DADN) that leverages a multi-scale noise estimation model and a U-Net network to diminish CT image noise. It is worth noting that down-sampling operations used in U-Net architecture lead to loss of important image details that adversely impact the image restoration performance. In order to address this issue, the proposed approach in [51] incorporates the attention mechanisms within the bottleneck layer of the U-Net network.

The attention mechanism has emerged as a powerful technique for extracting beneficial features from the input images. There exist two forms of attention mechanisms, namely, channel attention (CA) and spatial attention (SA), which suppress the redundant information from the feature map obtained by the neural network. In [29], the authors develop a denoising model that utilizes CA-based feature extraction layers, resulting in the improved visual quality of CT images compared to traditional CNNs. The authors in [54] utilize the foundational principles of CA by incorporating the SqueezeNet architecture to augment the capabilities of the Extreme Learning Machine (SNELM) for enhanced COVID-19 recognition. SqueezeNet offers a level of accuracy that rivals that of AlexNet [3], yet it achieves this with a fraction of the parameters. A crucial element of SqueezeNet's efficiency lies in its fire module, which ingeniously combines both 1 × 1 and 3 × 3 convolutional kernels.

Moreover, the attention-guided network (Dual-AGNet) [6] embeds three dimensional SA module (SAM) in dual projection and reconstruction domain networks for capturing the rich and representable sets of CT image features. Additionally, this framework has been trained using a combination of Structural Similarity Index (SSIM) and perceptual losses in order to maintain the structural details and prevent over-smoothing. The main drawback of these networks is that they only consider channel-wise interactions without extracting the spatial information of the various scales' feature map to refine the feature space. The authors [20] integrate the residual attention module into a Wasserstein distance generative adversarial network [12], referred to as WGAN-RAM, in which the generator effectively uses both CA and SA mechanisms to extract useful features. In [48], the novel Multiple-Input Deep Convolutional Attention Network (MIDCAN) is introduced for COVID-19 diagnosis, combining principles of CA and SA with multi-input strategies. MIDCAN is crafted to process two primary inputs-CCT and CXR images-simultaneously. Each image set is processed through its respective convolutional block attention module (CBAM), after which the extracted features are concatenated. Furthermore, the authors in [31], present a network based on the ResNet structure [14] for removing noise from CT images by enriched feature information obtained from CA and SA. However, CA and SA employed by these networks can only extract local information and are insufficient in extracting long-range channel interdependence.

To address these issues, we propose the Multi-Scale Residual Attention Network (MRAN), which learns powerful low-level and high-level feature representations useful for preserving contextual and structural information of the CT images for denoising. Specifically, we develop a novel residual block referred to as the Adaptive multi-scale Feature Fusion Module (AMFFM) that employs a multi-scale pyramid convolution structure and channel attention weight operations to capture local attention at different scales. This allows the model to focus on specific or nearby sets of pixels or features within each channel, thereby capturing fine-grained details. On the other hand, global attention is achieved through the softmax operation at the end stage of the AMFFM. This results in recalibrating the attention weights across all channels, facilitating an understanding of the broader context within the CT image, which allows richer semantic information to be grasped, and establishes long-range channel dependencies. By integrating these local and global attention mechanisms, the AMFFM offers a more comprehensive representation of the CT image's features.

To further enhance the capability of the MRAN we encapsulate the AMFFM unit into the Spatial Residual Information Module (SRIM) which eases the training process of a deep network by skip connection for efficient information flow. Specifically, this helps in mitigating the gradient vanishing problem, allowing the network to capture high-level semantic information. The SRIM also leverages a SA mechanism to focus on important spatial locations in the feature map, directing the network's attention to regions that are more informative for the task at hand. Further, MRAN employs the U-Net architecture with skip connections between various layers, which each of which representing unique level of information. This leads to fusing information from various hierarchical levels of abstractions. It is noteworthy that although the downsampling operations of U-Net may lead to the omission of essential details, this challenge is adeptly and adaptively managed in our MRAN by employing the attention mechanism and leveraging residual connections.

By integrating AMFFM and SRIM, MRAN effectively captures and emphasizes a multiple range of features at different scales and levels of abstraction, making it suitable for complicated task of CT image denoising. While the model architecture sets the stage, traditional pre-trained VGG networks, commonly used in training of the CT image denoising models, have presented marked limitations in balancing the SSIM and PSNR of reconstructed images. Addressing this aspect of training, a PSNR-enhanced loss function is introduced. This innovative loss approach meticulously integrates edge detection with random-weighted convolution layers, markedly boosting the network's prowess in capturing high-frequency information.

The main contributions of this paper are summarized as follows:

- The Adaptive multi-scale feature fusion Module (AMFFM) captures image details by increasing the receptive field of the network using group convolution operations with kernels of different sizes. This block applies attention in two local and global stages to process inherent rich semantic information and enhance the long-range contextual channel-wise interactions of the pixels and consequently providing a superior CT image denoising performance.
- The Spatial Residual Information Module (SRIM) is introduced to ensure the MRAN network's training with larger depth and lower computational complexity using the residual learning technique and SA module.
- A novel PSNR-enhanced perceptual loss network is introduced to increase PSNR while simultaneously extracting perceptual features. This loss is used in conjunction with the mean absolute error (MAE) loss to reduce the difference between

ground truth and estimated images by focusing on high-frequency information extraction.
- The MRAN network is proposed based on the U-Net structure for removing noise at different special resolutions, which emphasizes various types of information at different scales and levels.

The remainder of the paper is organized as follows: First, Sect. 2 provides a detailed explanation of our approach. Section 3 showcases the effectiveness of our denoising framework and hybrid loss function, while also conducting ablation studies on the AMFFM and SRIM. Finally, conclusions on the proposed MRAN and the experimental results are presentedsec1 in Sect. 4.

2 Methodology

2.1 Denoising Model

The process of suppressing noise in CT images can be viewed as a mapping between low-dose CT images and their corresponding normal-dose versions. A relationship between LDCT and NDCT can be expressed as:

$$x_o = g(x_i) \tag{1}$$

where $x_i \in \mathbb{R}^{H\times W\times 1}$ and $x_o \in \mathbb{R}^{H\times W\times 1}$ represent the LDCT and normal-dose CT (NDCT) images, respectively, and g is the degradation function associated with high quantum noise, random noise and other factors, such as round-off errors. The denoising problem can be formulated as:

$$\hat{x}_o = \underset{f}{\operatorname{argmin}}_{\mathrm{L}}(f(x_i), x_o) \tag{2}$$

where the loss function denoted by L is computed by measuring the distance between the estimated high-dose image ($\hat{x}_o$) and the ground truth image (x_o) in each iteration over all training data. The noise in the LDCT image is intricate and equally allocated over the whole image in the reconstruction process. A deep CNN can be utilized to learn a function f for suppressing the LDCT noise and enhancing their visual qualities. The quality of image estimated by the model is significantly influenced by the design of the loss function. A commonly used loss function for training CNNs to perform low-dose CT image denoising is per-pixel loss, which seeks to minimize the distance between each pixel in the estimated image and the ground truth image. However, this approach tends to over-smooth the estimated image, leading to suboptimal results. To address this issue, in this paper, we design a hybrid loss function of per-pixel loss and PSNR-enhanced perceptual loss to constrain the generation outcome of our model and improve the visual quality of the estimated images.

2.2 Attention Mechanism Processing

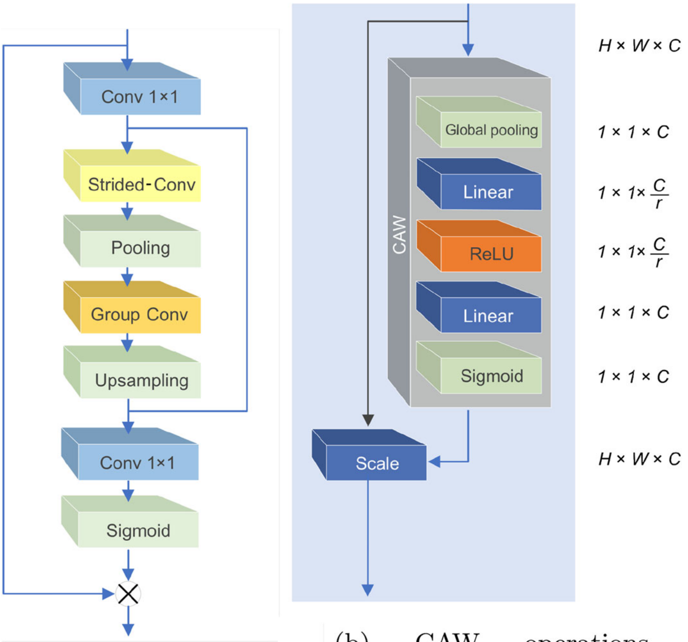
Squeeze-and-Excitation (SE) channel attention module [15] is an effective method for capturing informative features of CT images. The SE module consists of two steps including, Squeeze and Excitation. In the Squeeze step, global average pooling (GAP) is applied to the input features to generate channel descriptors, which summarize the feature maps in a channel-wise manner. As seen in Fig. 1a, this process reduces the spatial information of input feature map, x, from $H \times W \times C$ size to $1 \times 1 \times C$ by

$$x_{\text{gap}} = \text{GAP}(x) = \frac{1}{H \times W} \Sigma_{m=1}^{H} \Sigma_{n=1}^{W} x(m, n) \tag{3}$$

where H, W and C are height, width and the number of channels of feature map (x), respectively.

In the Excitation step, two linear layers along with Sigmoid operation are used to process the channel descriptors and estimate channel-wise attention vectors. Mathematically, the channel attention weights are processed with CAW function as,

$$w_{se} = \text{CAW}(x) = \sigma(W_2 \delta(W_1(x_{\text{gap}})) \tag{4}$$



(a) ESA module

(b) CAW operations within SE module

Fig. 1 The architecture of spatial attention (ESA) and Squeeze-and-Excitation channel attention (SE) modules

where $W_0 \in \mathbb{R}^{C \times \frac{C}{r}}$ and $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and r is the reduction ratio. The functions of the Sigmoid and RELU are denoted by σ and α, respectively.

The attention vectors are then used to re-calibrate the input features. Specifically, the re-calibration is carried out by multiplying the input features with the attention vectors, which emphasizes the informative channels and suppresses redundant ones as,

$$x_{\mathrm{se}} = x \circ w_{\mathrm{se}} \tag{5}$$

where $\circ$ is the element-wise multiplication operation.

In deep residual deep networks, SA plays a vital role by generating a per-pixel attention map that accentuates informative features by taking inter-spatial interaction into consideration. This mechanism is particularly significant in the context of very deep networks, as it enables such networks to prioritize information-rich spatial features. This approach is complementary to the CA mechanism.

The Enhanced Spatial Attention (ESA) block introduced in [22] employs both strided convolution and a large window-size max-pooling layer to achieve a wide receptive field, as shown in Fig. 1b. For performing a such operation in a lightweight framework, a 1 × 1 convolution is utilized for reducing the channel dimension in the ESA structure.

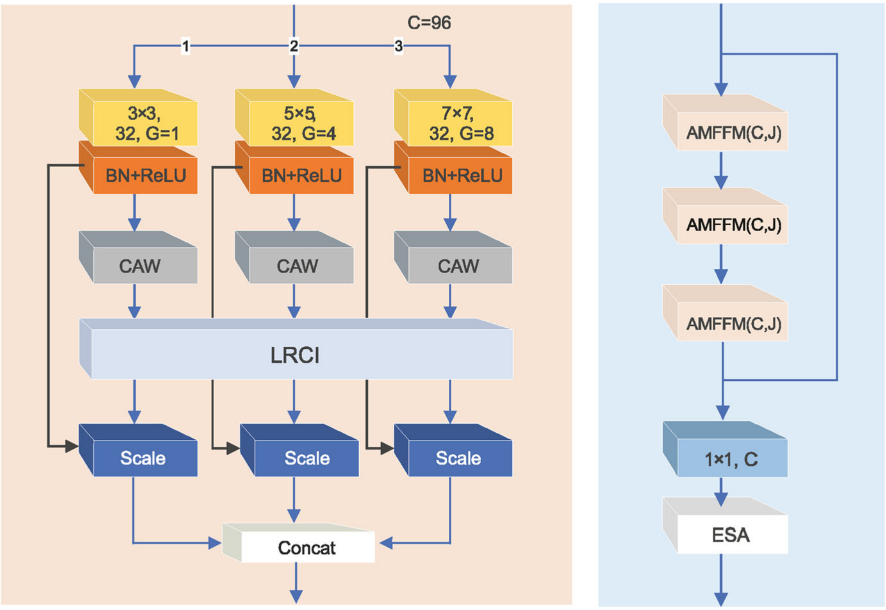
2.3 Multi-scale Feature Processing

In order to combine the merits of multi-scale feature extraction and CA mechanism, we design a novel feature extraction module referred to as an Adaptive multi-scale Feature Fusion Module (AMFFM) which comprises three main parts. Initially, it extracts multi-scale contextual features at a granular level to expand the receptive fields in a low-complexity manner. Then, it exploits the CA mechanism for capturing detailed CT features in each spatial scale that increases the representational ability of feature representations for CT image denoising. Finally, this module obtains the holistic interdependencies between multi-scale spatial information.

Figure 2a illustrates the overview of AMFFM with three different kernel-size feature extractors. Generally, the J convolutional layers with different kernel sizes can be implemented in parallel to obtain spatial features. Let $x \in \mathbb{R}^{H \times W \times C}$ be the input feature map. The proposed AMFFM extracts multi-scale features as,

$$y_j = \mathrm{Conv}(k_j \times k_j, C'_j = \frac{C}{N}, G = G_j)(x), \; j = 1, 2, ..., J \tag{6}$$

where $y_j \in \mathbb{R}^{H \times W \times C/J}$ is the obtained features from the j-th convolution layer with a kernel size of $k_j = 2j + 1$ and channel size of $C_j = C/J$. For example, the AMFFM shown in Fig. 2a consists of three convolution layers of 3 × 3, 5 × 5 and 7 × 7 kernel sizes, each of which contains 32 channels, given that the preceding layer has 96 channels. It should be noted that C must be divisible by J. Each convolution layer is followed by a batch normalization (BN) layer and a ReLU activation function. The BN layer is employed to accelerate the convergence and prevent issues such as



(a) AMFFM$(C = 96, J = 3)$ (b) SRIM(C, J)

Fig. 2 The architectures of proposed multi-scale attention and residual feature extraction modules

gradient vanishing or exploding which are commonly observed in the training of deep networks.

The computation cost of different kernel size convolution operations is associated with the increase in the number of network parameters. We employ group convolution operations in the proposed module's architecture for performing the multi-scale feature extraction process without increasing the number of parameters. Further, a novel rule is employed for determining the group size, G_j, without introducing additional parameters as,

$$G_j = \begin{cases} 2^{\frac{k_j-1}{2}}, & k_j > 3 \\ 0, & k_j = 3 \end{cases} \tag{7}$$

Each scale features of y_j is then input to the WSE module to effectively capture j-th CA weights (w_{sej}) as,

$$w_{\text{sej}} = \text{CAW}(y_j) \tag{8}$$

This enables the AMFFM to capture information from multiple scales, and result in improving the local attention process for high-level feature map. To enhance the interactivity of multi-scale channels, the CA weights are subjected to a Softmax operation, which normalizes the contribution of each of them within the long-range channel interdependence (LRCI) block. This enables the adaptive selection of spatial scales and facilitates global multi-scale channel-wise interaction. This process can be formulated as,

$$W_j = \mathrm{Softmax}(w_{\mathrm{se}j}) = \frac{\exp(w_{\mathrm{se}j})}{\Sigma_{j=1}^{J} \exp(w_{\mathrm{se}j})} \tag{9}$$

where W_j is the j-th attention vector after local and global interaction. Finally, the information-rich features are constructed by multiplication of weight vectors of W_j with corresponding features of y_j and concatenation process as,

$$\begin{aligned} y_{\mathrm{se}j} &= W_j \circ y_j \\ y_o &= \mathrm{Cat}([y_{\mathrm{se}1}, y_{\mathrm{se}2}, \ldots, y_{\mathrm{se}J}]) \end{aligned} \tag{10}$$

where $y_{\mathrm{se}j}$ and y_o are the re-calibrated features of j-th convolution layer and the output of the AMFFM module, respectively. The proposed AMFFM block obtains information in different scales and produces fine-grained features with the multi-scale attention mechanism, which can be used in different medical computer vision tasks.

The cascade of three AMFFM units is now utilized in a residual module, referred to as Spatial Residual Information Module (SRIM), in order to generate a reacher set of features. Specifically, the architecture SRIM is shown in Fig. 2b. To strengthen the residual features, inspired by [26], a 1 × 1 convolution layer and an ESA block are placed sequentially following stacked AMFFMs. The ESA block enhances the network's capability of emphasizing the important spatial features and extracting more representative ones. In Particular, this process can be described as follows with F_i and F_o as the input and output tensors:

$$\begin{aligned} F_{\mathrm{ext}1} &= h_{\mathrm{amf}}(F_i) \\ F_{\mathrm{ext}2} &= h_{\mathrm{amf}}(F_{\mathrm{ext}1}) \\ F_{\mathrm{ext}3} &= h_{\mathrm{amf}}(F_{\mathrm{ext}2}) \\ F_{\mathrm{res}} &= F_i + F_{\mathrm{ext}3} \\ F_o &= h_{\mathrm{ConvESA}}(F_{\mathrm{res}}) \end{aligned} \tag{11}$$

where h_{amf} represents the AMFFM parameters and h_{ConvESA} represents the 1 × 1 convolution and ESA operations parameters.

2.4 Network Overall Architecture

Many of the state-of-the-art image denoising schemes [1] utilize the U-Net structure for their overall network architectures. U-Net architectures are able to extract the structural information from the input noisy images, which is useful in predicting the noise. In view of this, we employ the U-Net as the backbone architecture of our CT image denoising network. The proposed efficient Multi-Scale Residual Attention Network (MRAN) that employs the SRIM is formed by three main modules, namely, Shallow Feature Extraction (SFE), Deep Feature Extraction (DFE), and Image Reconstruction

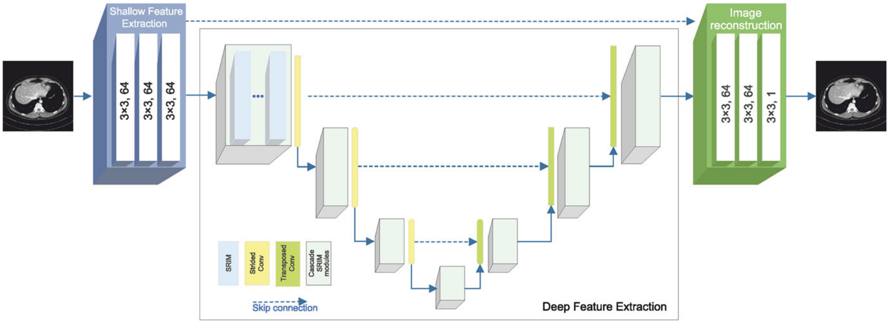


Fig. 3 The architecture of the proposed CT image denoising network

(IR), as depicted in Fig. 3. The SFE is composed of three stacked 3 × 3 convolution layers that are utilized to extract the coarse features from the input image. This process can be represented as,

$$F_0 = h_{\mathrm{sfe}}(x_i) \tag{12}$$

where h_{sfe} represents the SFE parameters and F_0 represents the features of the shallow layers.

The DFE architecture consists of an encoder and a decoder in seven stages. To set the feature resolutions, we utilize three 3 x 3 strided convolution layers with a stride value of 2 after the encoder stages and three transposed convolution layers with a stride value of 2 before the decoder stages. This U-Net architecture utilizes mirror skip connections between the peer encoder and decoder layers to facilitate the flow of information. Each stage of the DFE contains multiple stacked SRIMs, each of which hierarchically processes informative features. A detailed overview of the hyperparameters used in the proposed DFE is provided in Table 1. This process can be formulated as,

$$\begin{aligned} F_i &= S_i(h_i(F_{i-1})), & i &= 1, 2, 3 \\ F_i &= h_i(F_{i-1}), & i &= 4 \\ F_i &= h_i(T_{i-4}(F_{i-1} + F_{8-i})), & i &= 5, 6, 7 \end{aligned} \tag{13}$$

where h_i is the parameters of ith stage of DFE, S_i and T_i represent its corresponding strided and transposed convolution layer parameters, respectively. Finally, the reconstruction process is implemented with three 3 × 3 convolution layers as,

$$\hat{x}_o = h_{\mathrm{ir}}(F_0 + F_7) \tag{14}$$

where h_{ir} represents the image reconstruction parameters.

Table 1 Summary of the hyperparameters utilized in the proposed DFE module

Stages	AMMFM(C,J)	SRIM #	Output
1	(64,4)	1	64 × 64
2	(96,3)	1	32 × 32
3	(64,2)	2	16 × 16
4	(64,1)	3	8 × 8
5	(64,2)	2	16 × 16
6	(96,3)	1	32 × 32
7	(64,4)	1	64 × 64

2.5 Loss Functions

The selection of an appropriate loss function is crucial in the training of the neural network. To enhance the training process of the proposed network, a combination of Mean Absolute Error (MAE) and perceptual losses is proposed. While the MAE loss function is capable of removing noise on a per-pixel basis by computing the L1 norm difference between the predicted and ground truth images, it may introduce the over-smoothing artifact in the estimated image. The MAE loss function is defined as,

$$L_{\text{MAE}} = \frac{1}{N} \sum_{j=1}^{N} \|\hat{x}_{oj} - x_{oj}\| \tag{15}$$

On the other hand, perceptual loss is frequently employed in image restoration because of its ability to capture the details and contents of images. The perceptual loss function is formulated as follows:

$$L_{\text{Perceptual}} = \sum_{j=1}^{N} \frac{1}{h_j w_j d_j} \|\phi_j(\hat{x}_o) - \phi_j(x_o)\|^2 \tag{16}$$

where h, w, and d are the height, weight and depth of the features extracted by the feature extractor Q. In recent years, pre-trained VGG networks [38] have been widely used as feature extractors for CT image processing. For example, WGAN-VGG [44] employs perceptual features from the 16th layer of VGG19 network, while DRL [11] utilizes 2nd, 4th, 7th and 10th layers of VGG16 network. However, we empirically observed that although the perceptual loss extracts structural features and is effective in improving the SSIM, it leads to decrease in the Peak Signal-to-Noise Ratio (PSNR) of the reconstructed images. To further investigate, we calculate the average of the intermediate layers' features (AIF) of the VGG19 network based on Eq. 17, and visualize them for the given NDCT image in Fig. 4. As seen from Fig. 4b to f, the features produced by shallow layers of the VGG network mostly contain high-frequency information and well-represent the CT image content information, while the features generated by the deeper layers contain low-frequency information and lack detailed structures. Therefore, one can conclude that a shallow network could

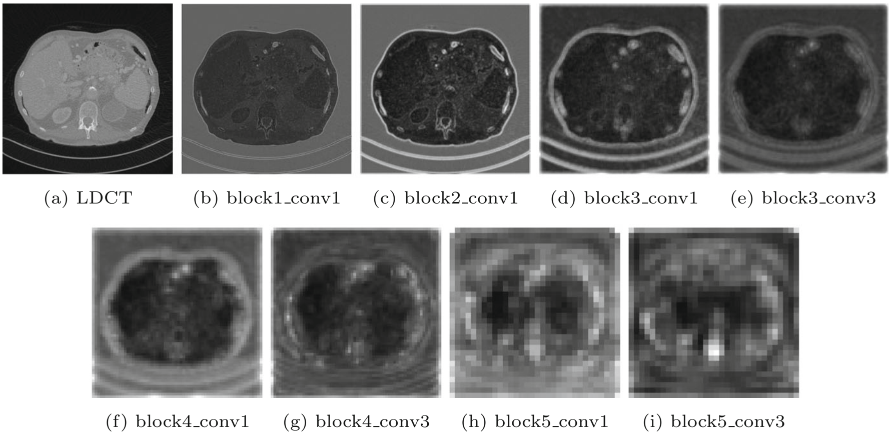


Fig. 4 The average of intermediate features of VGG19 network. For better comparison, we display intermediate features with different resolutions at the same size

capture substantial perceptual features.

$$\mathrm{AIF}_{h,w} = \sqrt{\sum_{i=1}^{D_i} \frac{1}{D_i} (\phi_i(x_o)_{h,w,d})^2} \tag{17}$$

Drawing from the aforementioned observations and recent research findings which indicate that a well-structured random-weighted network, without training, can provide superior perceptual performance [23, 41], we develop a novel loss for network training process that not only captures intricate CT details but also enhances the PSNR values of the predicted image. The set of operations used by our proposed loss is depicted in Fig. 5. As seen in Fig. 5, the proposed loss function referred to as PSNR-enhanced employs an edge detection layer, followed by two random-weighted convolution layers with a ReLU activation layer in between them. The network's capacity for capturing high-frequency information is enhanced by incorporating pre-defined Sobel kernels in the x, y, and diagonal (45° and 135°) coordinates, along with a Laplacian kernel (6). All these operations are embedded in the Edge detection layer. The extracted high-frequency features from an NDCT input image are illustrated in Fig. 6. It is seen from Fig. 6 that the proposed loss term indeed contributes to obtaining useful information whose employment in the backpropagation process of the network potentially results in superior performance. To evaluate the performance of the PSNR-enhanced network, we use the following equation to visualize the detailed image (DI) obtained by the perceptual loss network.

$$DI_{h,w} = \sqrt{\sum_{i=1}^{D_i} \frac{1}{D_i} (\phi_i(z)_{h,w,d} - \phi_i(z')_{h,w,d})^2} \tag{18}$$

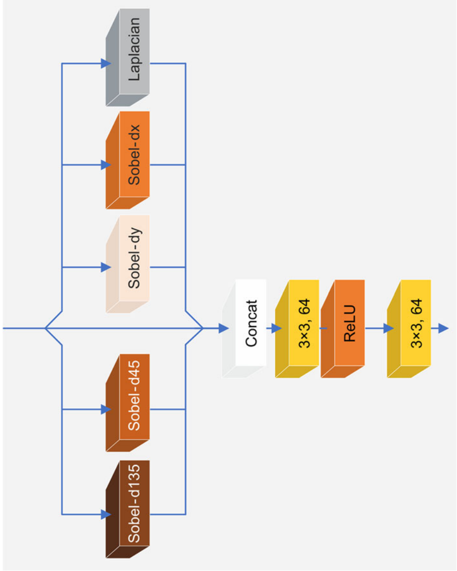
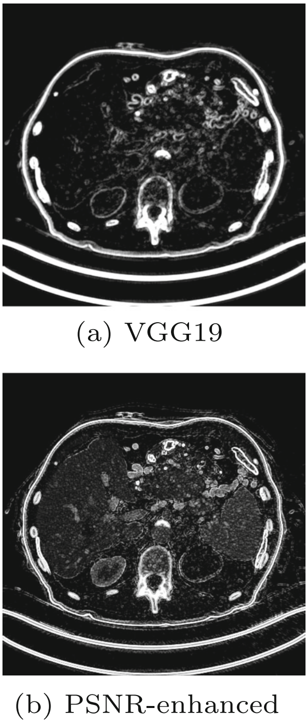
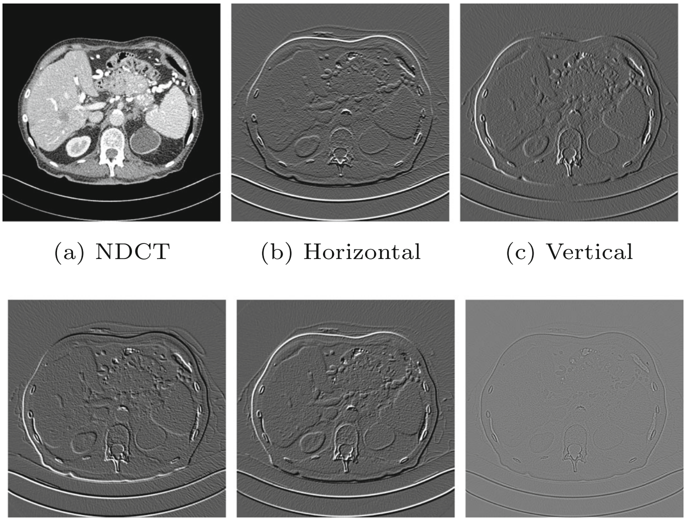


Fig. 5 The architecture of the proposed PSNR-enhanced perceptual loss network


(a) VGG19

(b) PSNR-enhanced

Fig. 6 Detailed images obtained from PSNR-enhanced and VGG19 perceptual loss networks



(a) NDCT (b) Horizontal (c) Vertical

(d) Diagonal (45°) (e) Diagonal (135°) (f) Laplacian

Fig. 7 High-frequency features extracted through Sobel and Laplacian operations in the edge detection layer of the PSNR-enhanced loss network

where z and z' denote the NDCT images from validation set and their respective degraded smoothed versions. Figure 7 presents sample detailed images obtained from both PSNR-enhanced and VGG19 perceptual networks, as seen from this Fig., it is evident that our PSNR-enhanced network outlines the detailed image with richer textures than those provided by the pre-trained VGG19 network. Therefore, it is concluded that the utilization of a shallow random-weight loss network, which receives high-frequency information as input, can effectively extract perceptual features.

The overall hybrid loss function for the proposed MRAN can be expressed as,

$$\text{Loss} = \gamma_1 L_{\text{MAE}}(x_o, \hat{x}_o) + \gamma_2 L_{\text{Perceptual}}(x_o, \hat{x}_o) \tag{19}$$

where γ_1 and γ_2 are the hyperparameters for the loss components. During the training process, the hyperparameters are determined by selecting the maximum loss value from each epoch and using it to update the hyperparameter values. The loss function with the highest loss value is assigned a higher scale compared to the other function.

3 Experiments

3.1 Datasets and Training Settings

In this work, the proposed denoising network is trained with three distinct datasets, namely, deceased piglet, phantom thoracic and clinical TCIA datasets.

The Piglet dataset [45] comprises 900 CT image pairs with a thickness of 0.625 mm, an X-ray current of 300 mAs for normal-dose images and 15 mAs for low-dose images, taken at a peak voltage of 100 KVp. The Thoracic dataset [10] consists of 407 pairs of CT images acquired from an anthropomorphic thoracic phantom. The current tube used for acquiring the normal-dose and low-dose CT images is 480 mAs and 60 mAs, respectively, with a peak voltage of 120 KVp and a slice thickness of 0.75 mm.

The clinical TCIA dataset [33] is composed of 299 patients utilizing different commercial CT scanners. For this study, we focus on a subset of 150 patients who are scanned using either the SOMATOM Definition AS+ or SOMATOM Definition Flash Siemens Healthineers CT scanners. Within this subset, there are 1782 non-contrast head CT images for acute cognitive or motor deficit, 16648 non-contrast chest scans for high-risk pulmonary nodule screening, and 7380 contrast-enhanced CT images of the abdomen for detecting metastatic liver lesions. The LDCT images are reconstructed using the filtered back projection technique after introducing Poisson noise to the standard clinical protocol-generated normal-dose projection image, obtained with a 330 mAs X-ray current tube, 120 KVp peak voltage, and 1.25 mm slice thickness. The LDCT images from head and abdomen regions are provided at 25% of the standard dose, while those from chest regions are provided at 10% of the normal-dose.

All datasets have images with spatial resolution of 512 × 512 pixels. The standard 80–10–10% proportion is used for training, validation, and testing. Additionally, each training dataset is divided into 64 × 64 overlapping patches to increase the number of training images and reduce the network's computational burden. We augment the training datasets with horizontal flip and random rotation operations. During the selection process, we disregard image patches that predominantly contain air. The predicted denoised images are evaluated by PSNR and SSIM metrics.

The proposed network in this study is trained with a total of 100 epochs and a mini-batch size of 16, using the ADAM optimizer with $\beta_1 = 0.01$ and $\beta_2 = 0.999$. The initial learning rate is set to 1×10^4 and is decreased by a factor of 10 after 75 epochs. The network is implemented using the TensorFlow package on a machine equipped with NVIDIA GeForce GTX 3090 GPU.

3.2 Ablation Study

In this section, we present a set of extensive experiments to demonstrate the efficacy of the different modules used in our method in which the LDCT dataset is utilized to train the networks. Firstly, we confirm the effectiveness of the AMFFM design by examining the impact of multi-directional long-range channel re-calibration on learning multi-scale feature representations. Next, we evaluate the rationale behind increasing the depth of the network in SRIM. Further, we investigate the effectiveness of SFE and IR in improving the denoising performance of the proposed RMAN. Finally, we assess the performance of the network trained with either PSNR-enhanced or VGG-based perceptual losses.

To confirm the effectiveness of the proposed AMFFM on the network performance, we conduct the ablation experiment on the basic model of our MRAN, where only one AMFFM in every seven stages of DFE is implemented. Specifically, we evaluate the impact of kernel size and group size relations, multi-scale feature extraction, and long-range channel interdependence on the performance of the task of LDCT denoising.

Firstly, we set the J to four and made appropriate adjustments to the G_j and k_j. The output presented in Table 2 demonstrate that the denoising performance achieves the highest value when we choose G_j as 1, 4, 8, and 16 for 3×3, 5×5, 7×7, and 9×9 kernel size convolution layers, which also validates Eq. 7. The results in Table 3, where we keep the G_js constant, demonstrate that employing multi-scale feature extraction yields superior PSNR and SSIM values compared to using fixed kernel sizes.

To further enhance our network's performance, we evaluate the impact of using different kernel sizes of AMFFM at each stage of the denoising process. Table 4 shows that the network significantly enhances its ability to remove noise when different branch sizes of AMFFM are embedded at varying depths within the network. Specifically, the network reaches the highest PSNR and SSIM values when J is set to 4, 3, 2, 1, 2, 3, and 4 branches in the seven DFE stages, respectively. The size of the feature map in each stage where multi-scale convolution is applied is a crucial factor in achieving this outcome. In particular, in the network's bottleneck where the feature resolution is 8×8, using a small kernel size convolution is more effective for processing and extracting noise features.

Next, we aim at comprehensively evaluating the performance of our proposed SRIM architecture for LDCT denoising. Specifically, we conduct experiments to investigate the impact of our design choices, including the number of AMFFM structures, residual connections, and spatial attention modules. In this regard, we depict the PSNR curves

Table 2 Effects of Group Size (G_j) changes on PSNR and SSIM in AMFFM

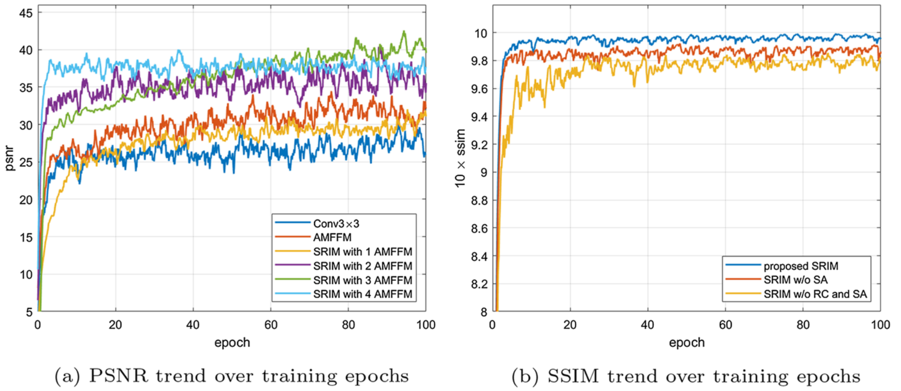
Kernel_size	Group_size	PSNR	SSIM
3,5,7,9	16,16,16,16	36.42	0.8312
3,5,7,9	4,8,8,16	36.51	0.8340
3,5,7,9	1,4,8,16	**37.04**	**0.8418**

Table 3 Effects of Kernel Size (k_j) changes on PSNR and SSIM in AMFFM

Kernel_size	Group_size	PSNR	SSIM
3,5,7,9	1,4,8,16	**37.04**	**0.8418**
5,5,7,7	1,4,8,16	36.86	0.8411
5,5,5,5	1,4,8,16	36.89	0.8416
7,7,7,7	1,4,8,16	36.90	0.8417

Table 4 Effects of multi-scale feature extraction at differnt stages of DFE on the denoising performance of MRAN

AMFFM(C,J) [stages: 1,2,3,4,5,6,7]	PSNR	SSIM	Params
[(64,1),(64,1),(64,1),(96,1),(64,1),(64,1),(64,1)]	36.84	0.8461	7.14×10^5
[(64,2),(64,2),(64,2),(128,2),(64,2),(64,2),(64,2)]	36.91	0.8479	7.60×10^5
[(96,3),(32,2),(96,3),(32,2),(96,3),(32,2),(96,3)]	37.22	0.8513	7.89×10^5
[(64,4),(96,3),(64,2),(64,1),(64,2),(96,3),(64,4)]	**37.49**	**0.8550**	7.32×10^5
[(64,4),(64,4),(64,4),(128,4),(64,4),(64,4),(64,4)]	37.11	0.8534	7.18×10^5



(a) PSNR trend over training epochs (b) SSIM trend over training epochs

Fig. 8 The study of AMFFM

of the denoising network with varying SRIM architectures. As shown in Fig. 8a, it can be found that increasing the number of stacked AMFFM structures in SRIM leads to improving the denoising performance. We also observe that even though the PSNR of SRIM with four AMFFM units is superior to that of SRIM with three AMFFM structures in early training epochs, the slope of the green curve is larger and finally surpasses the former (blue curve). Furthermore, we find that using standard convolution or only one AMFFM unit in the RMAN architecture leads to the inferior denoising performance.

To evaluate the impact of residual connections and spatial attention modules, we depict the SSIM trends (Fig. 8b) comparing the performance of the proposed SRIM to versions without these components. Our results indicate that the proposed SRIM can capture and utilize fused feature information effectively, as confirmed by the superior SSIM performance compared to the other versions.

We now validate the effectiveness of the SFE and IR modules in LDCT noise suppression by exploring the impact of the number of stacked standard convolution layers in each module. We first fix the number of SFE layers at 3 and increase the number of IR layers from 0 to 5. We then use 3 IR layers and increases the number of SFE layers from 0 to 5. In both of these experiments, increasing the number of convolution layers leads to improving denoising performance. As seen in Fig. 9, the performance improvement plateaus after 3 layers, indicating that 3 layers is the optimal number of layers for both modules.

Finally, we train the proposed RMAN scheme using per-pixel losses (the mean squared error (MSE) and mean absolute error (MAE)), and perceptual losses based on the VGG network and PSNR-enhanced network. Our results from Fig. 10 demonstrate that these two types of losses have different initial values, with per-pixel losses starting below 0.2, and perceptual losses starting above 0.2. This observation confirms the need for hyperparameter tuning to train the network effectively with a hybrid loss function. Furthermore, we observe that both MAE and MSE losses perform well in denoising task, while the former exhibited rapid convergence. When comparing the decrease in the PSNR-enhanced training loss (red curve) with that of the VGG training loss (blue

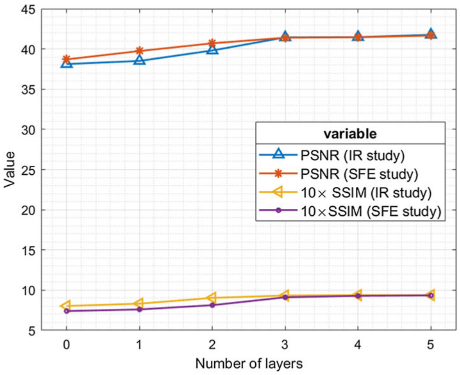


Fig. 9 The study of the SFE and IR modules

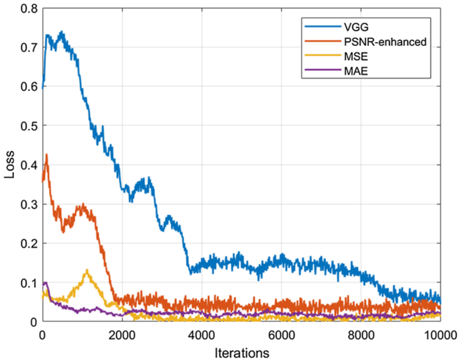


Fig. 10 Changing trend of different loss functions used for training of MRAN

curve), we find that the former decreases more rapidly, indicating its greater influence on the network's convergence speed and noise reduction.

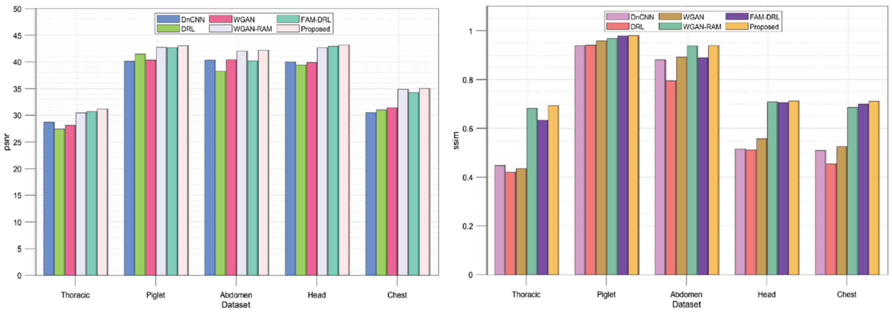
3.3 Denoising Result Comparison

In order to investigate the validity of the proposed methodology, we now perform comparative quantitative and qualitative analyses with the results obtained from other state-of-the-art techniques, namely, DnCNN, DRL, WGAN-VGG, WGAN-RAM, and FAM-DRL.

We analyze the denoising performance of the above-mentioned models for diverse datasets and calculate the PSNR and SSIM for test images. These models are trained and tested on each dataset, and the average quantitative results are shown in Table 5. Notably, for the Piglet dataset, the PSNR values range from 40 to 43, while the SSIM values range from 0.93 to 0.98 across different models. The average PSNR trend across experiments is illustrated in Fig. 11a. For the Thoracic dataset, the proposed model achieves a higher PSNR value (31.17) compared to the other models. These PSNR values are consistent across the other datasets. Conversely, the FAM-DRL and WGAN-RAM models exhibit a significant variance in PSNR results compared to the proposed network across different datasets. Specifically, the FAM-DRL model yields

Table 5 Average PSNR and SSIM results for different denoising methods applied to test images from different datasets (80%, 10%, and 10% for training, validation, and test partitions, respectively)

Network					TCIA dataset					
	Thoracic		Piglet		Abdomen		Head		Chest	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
DnCNN	28.73	0.4473	40.12	0.9387	40.32	0.8808	39.98	0.5142	30.51	0.5102
DRL	27.43	0.4204	41.45	0.9406	38.29	0.7954	39.39	0.5114	31.01	0.4533
WGAN-VGG	28.12	0.4352	40.34	0.9582	40.44	0.8919	39.88	0.5570	31.44	0.5254
WGAN-RAM	30.44	0.6821	42.72	0.9689	42.01	0.9374	42.67	0.7090	34.87	0.6860
FAM-DRL	30.67	0.6336	42.67	0.9789	40.17	0.8890	42.92	0.7055	34.25	0.6996
Proposed	**31.17**	**0.6923**	**43.04**	**0.9800**	**42.16**	**0.9392**	**43.18**	**0.7116**	**35.06**	**0.7105**



(a) Average PSNR results

(b) Average SSIM results

Fig. 11 Quantitative results on test images of different datasets utilizing different denoising methods

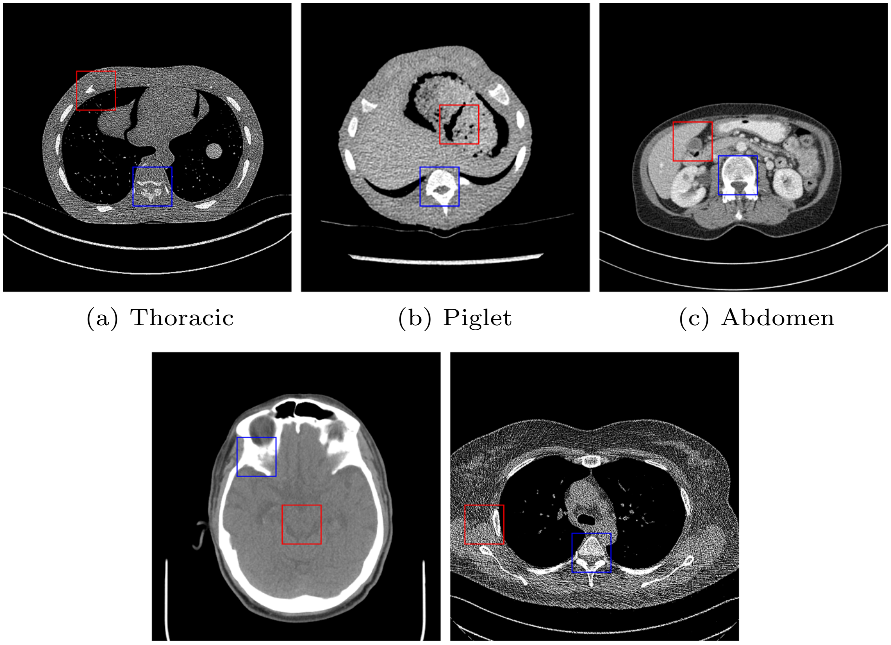
superior results in the head TCIA dataset, whereas the WGAN-RAM model achieves higher PSNR values in the abdomen and chest TCIA dataset.

Figure 11b shows the trend of the SSIM, which suggests that the SSIM of the other models is slightly lower across different datasets than the proposed model. In the Thoracic dataset, the FAM-DRL model stands out as the second best-performing method thanks to its objective of minimizing the structural dissimilarity between the LDCT and NDCT image pairs. Moreover, the WGAN-RAM model achieves comparable results due to its utilization of multi-scale convolution layers. However, when comparing different datasets, no clear pattern emerges to indicate which model outperforms the proposed one. This discrepancy can be attributed to the variability of the structural information and level of noise in different datasets.

In a quantitative evaluation on the Thoracic dataset, our proposed method yields an SSIM value of 0.6923, surpassing WGAN-RAM's 0.6821. This shows its heightened capability in preserving the intricate structural and contextual features of thoracic CT images. Upon analysis of the Piglet dataset, our model marks an SSIM of 0.9800, slightly exceeding FAM-DRL's 0.9789, indicating its adeptness at capturing detailed piglet anatomy. In the context of the Abdomen dataset, a close correspondence is observed between our method and WGAN-RAM, with SSIM values of 0.9392 and 0.9374, respectively. This accentuates the efficacy of both methodologies in safeguarding essential structural and contextual attributes of abdominal CT images.

Figure 12 illustrates five representative LDCT slices of test images from various datasets including, Thoracic, Piglet and TCIA (chest, abdomen, and head) datasets. Blue and red regions of interest (ROIs) highlight structural details and anatomical parts, especially where the deviations between denoising results are pronounced. The blue and red LDCT regions are shown in Figs. 13, 14, 15, 16, 17a, and corresponding NDCT images are presented in Figs. 13, 14, 15, 16, 17h. Further, we show the visual results of various algorithms in Figs. 13, 14, 15, 16, 17b–g.

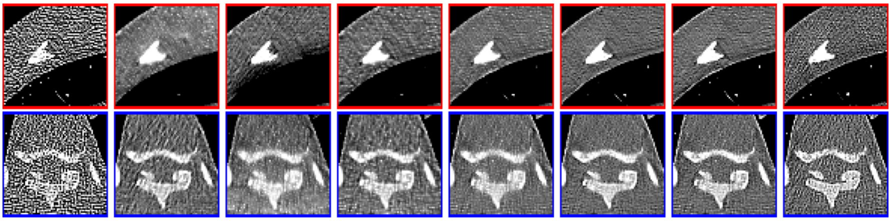
To assess the impact of reducing training dataset size and enlarging the testing dataset, we have divided the dataset into three: 70%, 10%, and 20% for training, validation, and testing, respectively. This division allows us to assess the generalizability of our model with a larger number of testing samples. The outcomes of this experiment, presented in Table 6, show that an increased number of testing samples does not



(a) Thoracic (b) Piglet (c) Abdomen

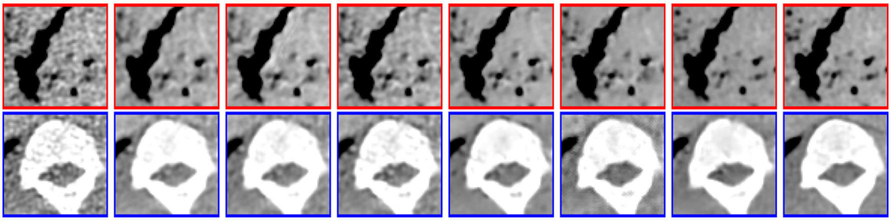
(d) Head (e) Chest

Fig. 12 Sample LDCT images from test images of different datasets with two blue and red ROIs (Color figure online)



(a) LDCT (b) DnCNN (c) DRL (d) GAN-VGG (e) GAN-RAM (f) FAM-DRL (g) MRAN (h) NDCT

Fig. 13 Selected ROIs from LDCT sample image (Fig. 12a) and its corresponding NDCT image in the Thoracic dataset, along with the results of various denoising methods



(a) LDCT (b) DnCNN (c) DRL (d) GAN-VGG (e) GAN-RAM (f) FAM-DRL (g) MRAN (h) NDCT

Fig. 14 Selected ROIs from LDCT sample image (Fig. 12b) and its corresponding NDCT image in the Piglet dataset, along with the results of various denoising methods

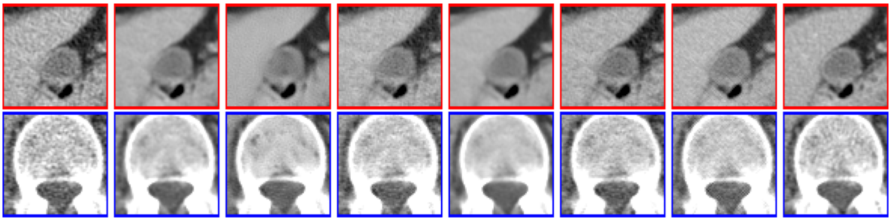

(a) LDCT (b) DnCNN (c) DRL (d) GAN-VGG (e) GAN-RAM (f) FAM-DRL (g) MRAN (h) NDCT

Fig. 15 Selected ROIs from LDCT sample image (Fig. 12c) and its corresponding NDCT image in the TCIA dataset (Abdomen), along with the results of various denoising methods

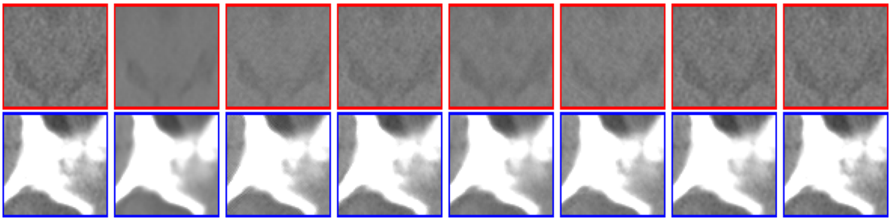

(a) LDCT (b) DnCNN (c) DRL (d) GAN-VGG (e) GAN-RAM (f) FAM-DRL (g) MRAN (h) NDCT

Fig. 16 Selected ROIs from LDCT sample image (Fig. 12d) and its corresponding NDCT image in the TCIA dataset (Head), along with the results of various denoising methods

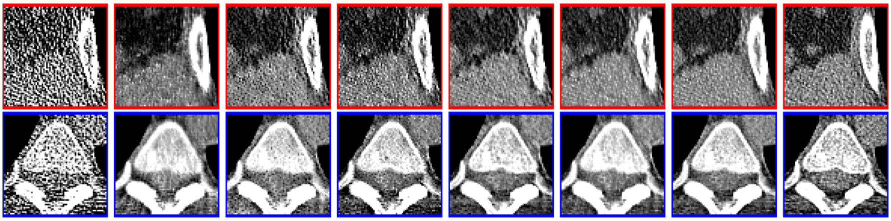

(a) LDCT (b) DnCNN (c) DRL (d) GAN-VGG (e) GAN-RAM (f) FAM-DRL (g) MRAN (h) NDCT

Fig. 17 Selected ROIs from LDCT sample image (Fig. 12e) and its corresponding NDCT image in the TCIA dataset (Chest), along with the results of various denoising methods

Table 6 Average PSNR and SSIM results for different denoising methods applied to test images from different datasets (70%, 10%, and 20% for training, validation, and test partitions, respectively)

Network					TCIA dataset					
	Thoracic		Piglet		Abdomen		Head		Chest	
	PSNR	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**	**PSNR**	**SSIM**
DnCNN	28.68	0.4461	40.01	0.9372	40.25	0.8762	39.94	0.5105	30.45	0.5070
DRL	27.40	0.4185	41.39	0.9398	38.22	0.7924	39.35	0.5072	30.95	0.4511
WGAN-VGG	28.10	0.4310	40.28	0.9573	40.33	0.8875	39.83	0.5545	31.38	0.5230
WGAN-RAM	30.42	0.6801	42.66	0.9675	41.94	0.9345	42.62	0.7079	34.84	0.6825
FAM-DRL	30.65	0.6301	42.62	0.9775	40.14	0.8865	42.85	0.7010	34.21	0.6971
Proposed	**31.14**	**0.6910**	**43.01**	**0.9792**	**42.13**	**0.9382**	**43.15**	**0.7092**	**35.02**	**0.7089**

significantly affect the performance of various networks. Additionally, the table indicates that, with the new testing sample partitioning, our proposed method surpasses other techniques in terms of PSNR and SSIM metrics.

The presence of noise and artifacts in LDCT images is considerably higher than in NDCT images across all datasets. Visual inspection of DnCNN results reveals over-smoothing issue, (as observed in Figs. 13, 14, 15, 16, 17b). This can be attributed to the use of the MSE loss function, which is known to produce the over-smoothing issue around the edges. Similarly, the WGAN-VGG model produces checkerboard artifacts (as observed in Figs. 13, 14, 15, 16, 17d), which is a known effect of using the VGG perceptual loss. Although the DRL model exhibits superior visual results compared to DnCNN and WGAN due to its hybrid loss function during training, it is still susceptible to similar issues, highlighting the importance of network structure as a key factor in denoising.

Figures 13, 14, 15, 16, 17e depict the results of the WGAN-RAM model, which incorporates a residual attention module into the WGAN architecture to preserve the textural details of the images. The model achieves this objective with minimal residual artifacts, resulting in a more natural visual effect. In fact, the proposed network leverages a combination of PSNR-enhanced and MAE loss functions to produce well-structured denoised images that closely approximate the quality of the NDCT images, (as demonstrated in Figs. 13, 14, 15, 16, 17g).

4 Conclusion

In the CT image denoising technique, it is highly desirable to decrease the amount of radiation used for acquiring images. However, this could lead to noise artifacts making the low-dose images unreliable for diagnosis. In this paper, an innovative deep learning-based CT image denoising approach has been developed that has emphasized multi-scale information in LDCT images. The novel AMFFM, a core component of MRAN, has utilized multi-scale pyramid convolutions combined with CAW operations. This combination has allowed the model to effectively discern and capture the intricate details present across different channels of the CT features. Furthermore, by integration of AMFFM with SRIM, the network has gained a more profound insight into the context and nuances of CT images. Additionally, the introduction of the Spatial Attention (SA) module has not only refined the training process but also significantly improved the accuracy of feature extraction from the images.

The proposed deep CT image denoising network effectively captures the importance of the features that contribute to enhancing the denoising performance by employing both channel-wise and spatial-wise interactions of the pixels. Moreover, the proposed deep CT image denoising network has been shown to outperform state-of-the-art CT image denoising networks in the literature on benchmark datasets. To produce high-quality CT denoised images that are more visually pleasing, we have designed a new loss function that employs the high-frequency components of the output images to calculate the loss value. The results of various ablation experiments have shown the effectiveness of the different techniques utilized in our CT image denoising network design. This is reflected in the improved PSNR and SSIM values achieved during various model comparisons. Visual examination further underscores the network's proficiency in generating meticulously structured, high-quality denoised images that closely approximate the quality of NDCT images.

Funding The funding was provided by NSERC Discovery grant (RGPIN-2020-04441).

Data Availability The datasets of this study are available from https://opg.optica.org/oe/fulltext.cfm?uri=oe-18-14-15244&id=203597#articleDatasets, https://github.com/xinario/SAGAN and https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=52758026.

References

1. A. Abdelhamed, M. Afifi, R. Timofte, M.S. Brown, Ntire 2020 challenge on real image denoising: Dataset, methods and results, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 496–497 (2020)
2. A.A. Abdulla, Efficient computer-aided diagnosis technique for leukaemia cancer detection. IET Image Proc. **14**(17), 4435–4440 (2020)
3. M.Z. Alom, T.M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M.S. Nasrin, B.C. Van Esesn, A.A.S. Awwal, V.K. Asari, The history began from alexnet: a comprehensive survey on deep learning approaches. arXiv preprint arXiv:1803.01164 (2018)
4. Y. Behroozi, M. Yazdi, A.Z. Asli, Hyperspectral image denoising based on superpixel segmentation low-rank matrix approximation and total variation. Circuits Syst. Signal Process. **41**(6), 3372–3396 (2022)
5. D.J. Brenner, C.D. Elliston, E.J. Hall, W.E. Berdon et al., Estimated risks of radiation-induced fatal cancer from pediatric ct. Am. J. Roentgenol. **176**(2), 289–296 (2001)
6. L. Chao, P. Zhang, Y. Wang, Z. Wang, W. Xu, Q. Li, Dual-domain attention-guided convolutional neural network for low-dose cone-beam computed tomography reconstruction. Knowl. Based Syst. (2022). https://doi.org/10.1016/j.knosys.2022.109295
7. H. Chen, Y. Zhang, M.K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, G. Wang, Low-dose ct with a residual encoder-decoder convolutional neural network. IEEE Trans. Med. Imaging **36**(12), 2524–2535 (2017)
8. W. Chen, Y. Shao, Y. Wang, Q. Zhang, Y. Liu, L. Yao, Y. Chen, G. Yang, Z. Gui, A novel total variation model for low-dose ct image denoising. IEEE Access **6**, 78892–78903 (2018)
9. A. Ferrero, N. Takahashi, T.J. Vrtiska, A.E. Krambeck, J.C. Lieske, C.H. McCollough, Understanding, justifying, and optimizing radiation exposure for ct imaging in nephrourology. Nat. Rev. Urol. **16**(4), 231–244 (2019)
10. M.A. Gavrielides, L.M. Kinnard, K.J. Myers, J. Peregoy, W.F. Pritchard, R. Zeng, J. Esparza, J. Karanian, N. Petrick, A resource for the assessment of lung nodule size estimation methods: database of thoracic ct scans of an anthropomorphic phantom. Opt. Express **18**(14), 15244–15255 (2010)
11. M. Gholizadeh-Ansari, J. Alirezaie, P. Babyn, Deep learning for low-dose ct denoising using perceptual loss and edge detection layer. J. Digit. Imaging **33**, 504–515 (2020)
12. I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of wasserstein gans. Adv. Neural Inf. Process. Syst. 30 (2017)
13. D. Hart, M. Hillier, B. Wall, Doses to patients from medical X-ray examinations in the UK. 2000 review (2002)
14. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)
15. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks
16. Y. Huo, D. Wang, Y. Qi, P. Lian, A new gaussian kernel filtering algorithm involving the sparse criterion. Circuits Syst. Signal Process. **42**(1), 522–539 (2023)
17. L. Jia, A. Huang, X. He, Z. Li, J. Liang, A residual multi-scale feature extraction network with hybrid loss for low-dose computed tomography denoising. Available at SSRN 4327683
18. W. Kim, J. Lee, J.-H. Choi, An unsupervised two-step training framework for low-dose computed tomography denoising. Med. Phys. (2023). https://doi.org/10.1002/mp.16628

19. Z. Li, L. Yu, J.D. Trzasko, D.S. Lake, D.J. Blezek, J.G. Fletcher, C.H. McCollough, A. Manduca, Adaptive nonlocal means filtering based on local noise level for ct denoising. Med. Phys. **41**(1), 011908 (2014)
20. M. Li, Q. Du, L. Duan, X. Yang, J. Zheng, H. Jiang, M. Li, Incorporation of residual attention modules into two neural networks for low-dose ct denoising. Med. Phys. **48**, 2973–2990 (2021). https://doi.org/10.1002/mp.14856
21. K. Li, J.R. Chen, M. Feng, Construction of a nearly unbiased statistical estimator of sinogram to address ct number bias issues in low-dose photon counting ct. IEEE Trans. Med. Imaging (2023). https://doi.org/10.1109/TMI.2023.3240840
22. J. Liu, W. Zhang, Y. Tang, J. Tang, G. Wu, Residual feature aggregation network for image super-resolution
23. Y. Liu, H. Chen, Y. Chen, W. Yin, C. Shen, Generic perceptual loss for modeling structured output dependencies
24. Y. Liu, J. Ma, Y. Fan, Z. Liang, Adaptive-weighted total variation minimization for sparse data toward low-dose x-ray computed tomography image reconstruction. Phys. Med. Biol. **57**(23), 7923 (2012)
25. Y. Liu, Z. Gui, Q. Zhang, Noise reduction for low-dose x-ray ct based on fuzzy logical in stationary wavelet domain. Optik-Int. J. Light Electron Opt. **124**(18), 3348–3352 (2013)
26. J. Liu, J. Tang, G. Wu, Residual feature distillation network for lightweight image super-resolution, in *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16*, pp. 41–55 (2020). Springer
27. P. Luo, X. Qu, X. Qing, J. Gu, Ct image denoising using double density dual tree complex wavelet with modified thresholding, in *2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)*, pp. 287–290 (2018). IEEE
28. J. Ma, J. Huang, Q. Feng, H. Zhang, H. Lu, Z. Liang, W. Chen, Low-dose computed tomography image restoration using previous normal-dose scan. Med. Phys. **38**(10), 5713–5731 (2011)
29. L. Ma, H. Xue, G. Yang, Z. Zhang, C. Li, Y. Yao, Y. Teng, Scrdn: Residual dense network with self-calibrated convolutions for low dose ct image denoising. Nucl. Inst. Methods Phys. Res. **1045**, 167625 (2023)
30. A. Manduca, L. Yu, J.D. Trzasko, N. Khaylova, J.M. Kofler, C.M. McCollough, J.G. Fletcher, Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct. Med. Phys. **36**(11), 4911–4919 (2009)
31. L. Marcos, J. Alirezaie, P. Babyn, Low dose ct denoising by resnet with fused attention modules and integrated loss functions. Front. Signal Process. **1**, 1–11 (2022). https://doi.org/10.3389/frsip.2021.812193
32. C.H. McCollough, A.N. Primak, N. Braun, J. Kofler, L. Yu, J. Christner, Strategies for reducing radiation dose in CT. Radiol. Clin. (2009). https://doi.org/10.1016/j.rcl.2008.10.006
33. T.R. Moen, B. Chen, D.R. Holmes III., X. Duan, Z. Yu, L. Yu, S. Leng, J.G. Fletcher, C.H. McCollough, Low-dose ct image and projection dataset. Med. Phys. (2021). https://doi.org/10.7937/9NPB-2637
34. P.A. Oakley, D.E. Harrison, Death of the alara radiation protection principle as used in the medical sector. Dose-Response **18**(2), 1559325820921641 (2020)
35. Y. Pathak, K. Arya, S. Tiwari, Low-dose ct image reconstruction using gain intervention-based dictionary learning. Mod. Phys. Lett. B **32**(14), 1850148 (2018)
36. K. Rao, M. Bansal, G. Kaur, An effective ct medical image enhancement system based on dt-cwt and adaptable morphology. Circuits Syst. Signal Process. **42**(2), 1034–1062 (2023)
37. D.S. Rigie, A.A. Sanchez, P.J. La Rivière, Assessment of vectorial total variation penalties on realistic dual-energy ct data. Phys. Med. Biol. **62**(8), 3284 (2017)
38. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in *3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings*, pp. 1–14 (2015)
39. M. Su, J. Zheng, Y. Yang, Q. Wu, A new multipath mitigation method based on adaptive thresholding wavelet denoising and double reference shift strategy. GPS Sol. **22**, 1–12 (2018)
40. Z. Tian, X. Jia, K. Yuan, T. Pan, S.B. Jiang, Low-dose ct reconstruction via edge-preserving total variation regularization. Phys. Med. Biol. **56**(18), 5949 (2011)
41. P. Wang, Y. Li, A. Research, N. Vasconcelos, S. Diego, Rethinking and improving the robustness of image style transfer
42. J. Wang, H. Lu, T. Li, Z. Liang, Sinogram noise reduction for low-dose ct by statistics-based nonlinear filters, in *Medical Imaging 2005: Image Processing*, vol. 5747, pp. 2058–2066 (2005). SPIE

43. J. Wang, Y. Tang, Z. Wu, B.M.W. Tsui, W. Chen, X. Yang, J. Zheng, M. Li, Domain-adaptive denoising network for low-dose ct via noise estimation and transfer learning. Med. Phys. **50**, 74–88 (2023). https://doi.org/10.1002/mp.15952
44. Q. Yang, P. Yan, S. Member, Y. Zhang, Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. IEEE Trans. Med. Imaging **37**, 1348–1357 (2018)
45. X. Yi, P. Babyn, Sharpness-aware low-dose ct denoising using conditional generative adversarial network. J. Digit. Imaging **31**, 655–669 (2018)
46. Y. Zhang, J. Zhang, H. Lu, Statistical sinogram smoothing for low-dose ct with segmentation-based adaptive filtering. IEEE Trans. Nucl. Sci. **57**(5), 2587–2598 (2010)
47. J. Zhang, H.L. Zhou, Y. Niu, J.C. Lv, J. Chen, Y. Cheng, Cnn and multi-feature extraction based denoising of ct images. Biomed. Signal Process. Control **67**, 102545 (2021). https://doi.org/10.1016/j.bspc.2021.102545
48. Y.-D. Zhang, Z. Zhang, X. Zhang, S.-H. Wang, Midcan: a multiple input deep convolutional attention network for covid-19 diagnosis based on chest ct and chest X-ray. Pattern Recogn. Lett. **150**, 8–16 (2021)
49. J. Zhang, J. Lv, Y. Cheng, A novel denoising method for medical ct images based on moving decomposition framework. Circuits Syst. Signal Process. **41**(12), 6885–6905 (2022)
50. M. Zhang, S. Gu, Y. Shi, The use of deep learning methods in low-dose computed tomography image reconstruction: a systematic review. Compl. Intel. Syst. **8**(6), 5545–5561 (2022)
51. J. Zhang, Y. Niu, Z. Shangguan, W. Gong, Y. Cheng, A novel denoising method for ct images based on u-net and multi-attention. Comput. Biol. Med. **152**, 106387 (2023). https://doi.org/10.1016/j.compbiomed.2022.106387
52. P. Zhang, Y. Liu, Z. Gui, Y. Chen, L. Jia, A region-adaptive non-local denoising algorithm for low-dose computed tomography images. Math. Biosci. Eng. **20**(2), 2831–2846 (2023)
53. T. Zhang, D. Wu, X. Mo, The rank residual constraint model with weighted schatten p-norm minimization for image denoising. Circuits Syst. Signal Process. pp. 1–19 (2023)
54. Y. Zhang, M.A. Khan, Z. Zhu, S. Wang, Snelm: Squeezenet-guided elm for covid-19 recognition. Comput. Syst. Sci. Eng. **46**(1), 13 (2023)
55. W. Zhao, H. Lu, Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation. IEEE Trans. Instrum. Meas. **66**(9), 2283–2294 (2017)
56. T. Zhao, M. McNitt-Gray, D. Ruan, A convolutional neural network for ultra-low-dose ct denoising and emphysema screening. Med. Phys. **46**(9), 3941–3950 (2019)
57. F. Zhao, M. Liu, Z. Gao, X. Jiang, R. Wang, L. Zhang, Dual-scale similarity-guided cycle generative adversarial network for unsupervised low-dose ct denoising. Comput. Biol. Med. **161**, 107029 (2023)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.