China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 11, No. 3, September 2024 https://doi.org/10.15878/j.instr.202400187

Article

Multi-scale Attention Dilated Residual Image Denoising Network Based on Skip Connection

Zhiting Du, Xianchun Zhou*, Mengnan Lv, Yuze Chen, Binxin Tang

School of Electronics Information Engineering, Nanjing University of Information Science & Technology, Nanjing 21004, China

* Corresponding author email: zhouxc2008@163.com

Abstract: In the field of image denoising, deep learning technology holds a dominance. However, the current network model tends to lose fine-grained information with the depth of the network. To address this issue, this paper proposes a Multi-scale Attention Dilated Residual Image Denoising Network (MADRNet) based on skip connection, which consists of Dense Interval Transmission Block(DTB), Sparse Residual Block(SRB), Dilated Residual Attention Reconstruction Block(DRAB) and Noise Extraction Block(NEB). The DTB enhances the classical dense layer by reducing information redundancy and extracting more accurate feature information. Meanwhile, SRB improves feature information exchange and model generalization through the use of sparse mechanism and skip connection strategy with different expansion factors. The NEB is primarily responsible for extracting and estimating noise. Its output, together with that of the sparse residual module, acts on the DRAB to effectively prevent loss of shallow feature information and improve denoising effect. Furthermore, the DRAB integrates an dilated residual block into an attention mechanism to extract hidden noise information while using residual learning technology to reconstruct clear images. We respectively examined the performance of MADRNet in gray image denoising, color image denoising and real image denoising. The experiment results demonstrate that proposed network outperforms some excellent image denoising network in terms of peak signal-to-noise ratio, structural similarity index measurement and denoising time. The proposed network effectively addresses issues associated with the loss of detail information.

Keywords: image denoising; deep learning; dilated residual block; sparse residual block

Citation: Zhiting Du, Xianchun Zhou, Mengnan Lv, Yuze Chen, and Binxin Tang. "Multi-scale Attention Dilated Residual Image Denoising Network Based on Skip Connection." Instrumentation 11, no. 3 (2024). https://doi.org/10.15878/j.instr.202400187.

0 Introduction

The rapid development of computer science and technology has led to the widespread application of image processing technology in various fields such as remote sensing, transportation, military, and artificial intelligence. However, due to limitations in sensor quality and environmental conditions, noise is often introduced during the image capture process[1]. The objective of image denoising is to restore the original detailed information from noisy images, which poses a significant challenge. In the denoising process, it is essential not only to enhance the overall visual quality of the image but also to preserve its fine edges and features. Therefore, within the domain of image processing, image denoising technology represents a critical area for research[2].

Currently, image denoising technology is primarily categorized into model-based denoising methods and image denoising methods based on deep learning. Model-based image denoising algorithms are further classified into spatial domain methods, variation domain methods, and sparsity methods[3]. The non-local means algorithm (NLM) explores the non-local average of all pixels in an image and preserves edge information in the transform domain using self-similar patches[4]. The block method of 3 dimension (BM3D)[5] searches for similar blocks in the image and then utilizes the NLM algorithm of 3D transformation to remove noise. Weighted Nuclear

Norm Minimization[6] (WNNM) enhances the denoising effect by leveraging prior knowledge and sparsity method. While these approaches have yielded promising results, most traditional methods encounter challenges such as slow processing speed and manual parameter adjustment, making them unsuitable for real-time data applications.

Currently, convolutional neural networks (CNNs) hold a dominance in the field of image denoising. Most deep learning methods generate a noise map and then reconstruct a clear image by subtracting this noise map, focusing on noise learning and image recovery. Zhang et al[7]. introduced a residual learning CNN model (DnCNN), which utilizes the residual learning method to extract Gaussian noise features. Experimental results demonstrate that DnCNN exhibits robustness and effective denoising capabilities. Additionally, Kai Zhang et al. proposes the fast and efficient CNN model[8] (FFDNet) which enhances the DnCNN model by incorporating an adjustable noise level image as input, showing competitive performance in real image denoising. Furthermore, Tian et al[9]. Propose Enhanced CNN for Image Denoising (ECNDNet) which employs extended convolution to increase receptive fields and obtain accurate noise features. While Convolutional Blind Denoising Network[10] (CBDNet) combines noise estimation subnetwork with non-blind denoising network to improve generalization ability of the FFDNet for removing complex noises.

With the increase of network depth, there will also be a corresponding increase in the number of parameters and denoising time, which may result in overfitting and hinder practical application. The connection between deep and shallow layers will gradually weaken, leading to loss of details and hindering model denoising. Gaussian noise features are very similar to images with complex backgrounds, which can generate difficulty in noise extraction.

To address this problem, numerous scholars have proposed models with robust generalization and expressive capabilities. Tian et al[11]. developed an effective Attention-Guided Dilated Network (ADNet), which incorporates a sparse mechanism comprising dilated convolution and ordinary convolution to enhance performance and efficiency. Zhong et al[12]. introduced the Dense Residual Feature Extraction Network (DRFENet), which enhances ADNet by introducing skip connection and interval transfer strategies, resulting in outstanding generalization capabilities.

A multi-scale attentional dilated residual network (MADRNet) with skip connections is proposed in this paper to address the issue of potential loss of detailed information during the processing of noise in deep learning image denoising technology. The contributions of this paper are as follows:

1. We propose an interval transmission strategy based on skip connection, which is applied to DTB and significantly enhances noise feature extraction and information exchange between the upper and lower layers of the model.

2. In SRB, we introduce a strategy that combines extended convolution and skip join to enlarge the acceptance field, better distinguish noise from complex background, and obtain more detailed information through jump connections.

3. A noise extraction block is introduced to pay full attention to global similarity information of images by expanding convolution, using average pooling to smooth and suppress noise for improved network performance. The use of skip connection strategy ensures effective transmission of context information and improves model efficiency.

4. Our proposed attention mechanism includes DRB, where we combine convolution with dilated convolution to expand the receptive field, make it focus more on local and global features of the image, address grid effect caused by dilated convolution, thus further improving denoising performance.

In Section 1, we introduce some related works, including the skip connections, dilated convolution, attention mechanism, and sparse mechanism. In Section 2, we expound the proposed model MADRNet in detail. In Section 3, we present the experimental data and results of MADRNet. In Section 4, we provide the summary and prospect.

1 Related Work

1.1 Attention mechanism and sparse mechanism

1.1.1. Attention mechanism

Complex backgrounds can complicate the function of image and video applications, making training more challenging. To address the issue, attention mechanism has been introduced into the model. Currently, common attention mechanisms used in convolutional neural networks include channel attention, spatial attention, and mixed attention. The typical model for the channel attention mechanism is SENet[13], which involves two processes: squeeze and excitation. The squeeze process condenses global information and learns the significance of each channel in the channel dimension. The excitation process integrates the condensed features into the fully connected layer to predict the importance of each channel. The classical model for spatial attention mechanism is STNet, which can transform various spatial deformation data and automatically capture features of important regions. The hybrid attention mechanism combines both channel attention and spatial attention, with CBAM being a representative model for this approach. It effectively extracts important information from feature maps to enhance the expressiveness and generalization ability of the model.

1.1.2. Sparse mechanism

The sparse mechanism utilizes a combination of ordinary convolution and dilated convolution by alternating ways to create a hybrid convolution,

effectively restraining the unique raster effect of dilated convolution. This effect arises from not all pixels participating in the convolution operation, leading to discontinuities of the feature map. The adoption of the hybrid convolution strategy allows for more continuous extraction of feature information at each layer. While the 3×3 convolution kernel is limited to extracting local features, this design not only preserves detailed spatial information but also effectively prevents discontinuities of the feature extraction.

1.2 Dilated convolution

The process of image denoising is a complex one, involving the interrelation and interaction between contiguous pixels. In this process, the extraction of context information plays a crucial role, making it essential to expand the receptive field. Currently, there are two methods for enlarging the receptive field. The first involves increasing the depth of the network or expanding the size of the convolution kernel, but this results in an increase in model parameters, memory usage, and training difficulty. The second method utilizes dilated convolution which enlarges the receptive field without adding more parameters[14]. As a result, dilated convolution can extract richer feature information while striking a balance between model stability and complexity. Trung et al.[15] successfully applied dilated convolution to medical image denoising and successfully achieved clearer CT images, providing convenience to the medical industry. Therefore, this paper introduces dilated convolution as a means to enhance image denoising effectiveness.

1.3 Skip connection

The principle of skip connection is to directly add input data to the output of a specific layer, ensuring the uninterrupted flow of information while preserving the details of the input image. This structure employs an asymptotic reuse strategy to extract more contextual information. Wu et al.[16] introduced skip connection into the model, enhancing the robustness and expressiveness of model while extracting more detailed information. The skip connection method offers several advantages: effectively alleviating the problem of gradient disappearance and improving model robustness; utilizing a feature reuse strategy to extract more significant context information; significantly reducing the number of parameters and reducing model training difficulty; ensuring that the network can be extended to deeper levels.[17]

2 Proposed Model Construction

The MADRNet utilizes the skip connections, dilated convolutions, sparse mechanisms, and attention mechanisms while employing an end-to-end residual learning approach. This means that the input is a noisy image and the output is a clear image. The structure can be seen in Fig.1. The network parameters of the MADRNet are presented in Table 1. The parameters of each convolution within the table adhere to the logical sequence of operation of the block.

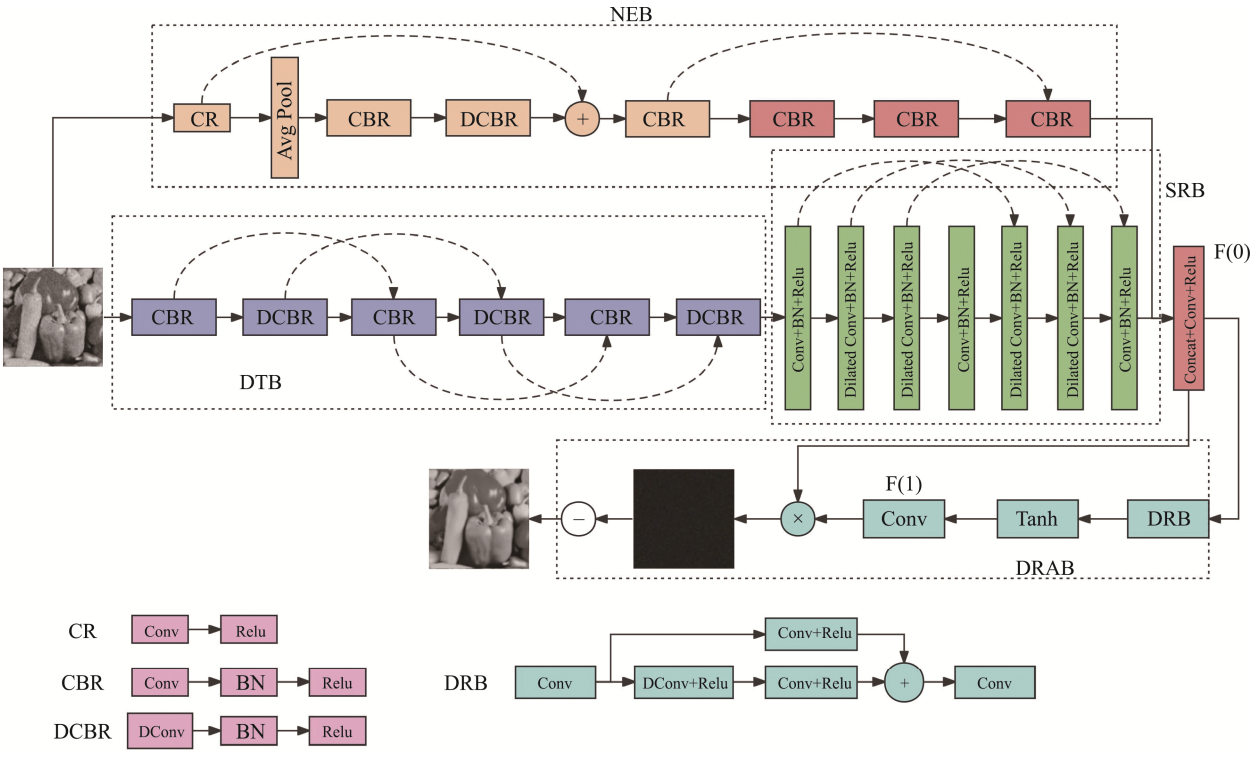


Fig.1 Architecture of MADRNet network

Table 1 The parameters of the MADRNet network

Block	Kernel size	Dilation	Stride	Padding	Channel	Pooling
DTB	3×3	1	1	1	32	
	3×3	2	1	2	32	
	3×3	1	1	1	32	
	3×3	2	1	2	32	
	3×3	1	1	1	32	
	3×3	2	1	2	64	
NEB	3×3	1	1	1	32	Avgpool(2×2)
	3×3	1	1	1	32	
	3×3	2	1	2	32	
	3×3	1	1	1	32	
	3×3	1	1	1	32	
	3×3	1	1	1	32	
	3×3	1	1	1	64	
SRB	3×3	1	1	1	64	
	3×3	2	1	2	64	
	3×3	2	1	2	64	
	3×3	1	1	1	64	
	3×3	2	1	2	64	
	3×3	2	1	2	64	
	3×3	1	1	1	64	
F(0)	3×3	1	1	1	1	
DRB	3×3	1	1	1	32	
	3×3	2	1	2	32	
	3×3	1	1	1	32	
	3×3	1	1	1	32	
	3×3	1	1	1	1	
F(1)	1×1	1	1	1	1	

The proposed MADRNet combines with MTB, NEB, SRB, and DRAB. The denoising mechanism of the network involves strengthening the learning of noise feature information through the dense interval transmission block and the sparse residual block, followed by extracting basic noise information through the noise extraction block[18]. Subsequently, the output of these blocks collaborates with the dilated residual attention reconstruction module to extract feature information with higher relevance, ultimately utilizing residual learning to reconstruct a clear image. The output of the network is

$$y = X_N - F_{MADRNet}(X_N) \tag{1}$$

Where y denotes clear image, X_N denotes input image, $F_{MADRNet}(X_N)$ denotes noise map image.

2.1 Dense interval transmission block

The skip connection simplifies deep information transmission and enhances noise expression, thereby improving the denoising effect. Due to the presence of numerous skip connections in the dense layer, dense connection block is incorporated into the model. The dense connection block offers several advantages, including facilitating back propagation, reducing the number of parameters, compacting the network structure, and enhancing model expressiveness. This structure improves feature information extraction by reusing the output of each convolutional layer. However, it also has limitations. For instance, reusing all features can cause the deep network to process large amounts of useless information.

Based on the DenseNet[19] network structure, this paper introduces a dense interval transmission block. The interval transmission strategy of skip connection is employed in dense interal transmission block, which can extract more valuable information and reduce parameter redundancy. The structure is illustrated in Fig.2. Within the block, CBR denotes Convolution+BN+Relu, while DCBR represents Dilated convolution+BN+Relu. Through interval transmission, this structure not only preserves the fundamental advantages of classical dense layers but also effectively diminishes parameter redundancy. Specifically, the block not only enlarges the receptive field but also significantly reduces parameter count by alternating between convolution and dilated convolution, thereby more effectively emphasizing key features within the network. The DTB can be expressed in Equation (2). The details of the DTB are expressed in Equation (3).

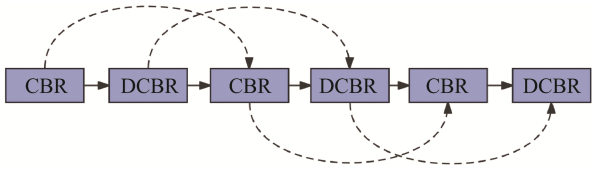


Fig.2 Dense interval transmission block

$$O_{DTB} = F_{DTB}(X_N) \tag{2}$$

$$\begin{aligned} O_1 &= F_{CBR}(X_N) \\ O_2 &= F_{DCBR}(O_1) \\ O_3 &= F_{CBR}(Cat(O_1, O_2)) \\ O_4 &= F_{DCBR}(Cat(O_2, O_3)) \\ O_5 &= F_{CBR}(Cat(O_3, O_4)) \\ O_6 &= F_{DCBR}(Cat(O_4, O_5)) \end{aligned} \tag{3}$$

Where F_{DTB}, O_{DTB}, F_{CBR}, F_{DCBR}, Cat, $O_{(1,2,3,4,5,6)}$ denote the function of the DTB, the output of the DTB, the function of CBR, the function of DCBR, the skip connection, the ouput of each layer, respectively

2.2 Sparse residual block

Considering that the sparse mechanism can extract more abundant spatial features from images, the shallow layer of a convolutional neural network typically represents low-level features, while the deep layer represents high-level features. As the network layers deepen, the extracted features become more accurate, but there is also a risk of losing context information. Therefore, this paper proposes the skip sparse block. In this block, the strategy of combining extended convolution and skip connection is utilized, enabling the model to acquire more detailed information.

The sparse residual block proposed in this paper incorporates three skip connections into the sparse mechanism, and introduces skip connection to the dilated convolution. Dilated convolution can obtain a wider receptive field, thus obtaining richer information and background. Therefore, it is very beneficial to adopt dilated convolution as the input of skip connections because it provides more information and the network structure and parameters remain the same[20].

The sparse residual block effectively mitigates the issue of gradient vanishing and enhances the propagation of feature information.The structure is depicted in Fig.3. SRB is shown in Equtation(4), the specific steps of SRB is shown in Equtation(5) and Equtation(6).

$$O_{CSB} = F_{CSB}(O_{DTB}) \tag{4}$$

$$\begin{aligned} O_{DCBR1} &= F_{CBR}(O_{DTB}) \\ O_{DCBR2} &= F_{DCBR}(O_{DCBR1}) \\ O_{DCBR3} &= F_{DCBR}(O_{DCBR2}) \end{aligned} \tag{5}$$

$$\begin{aligned} O_{CBR} &= F_{CBR}(O_{DCBR3}) \\ O_{DCBR4} &= F_{DCBR}(Cat(O_{CBR}, O_{DCBR1})) \\ O_{DCBR5} &= F_{DCBR}(Cat(O_{DCBR2}, O_{DCBR4})) \\ O_{DCBR6} &= F_{CBR}(Cat(O_{DCBR5}, O_{DCBR3})) \end{aligned} \tag{6}$$

Where F_{CSB}, O_{CSB}, F_{CBR}, F_{DCBR}, Cat, $O_{DCBR(1,2,3,4,5,6)}$ denote the function of the CSB, the output of the CSB, the function of CBR, the function of DCBR, the skip connection, the ouput of each layer, respectively.

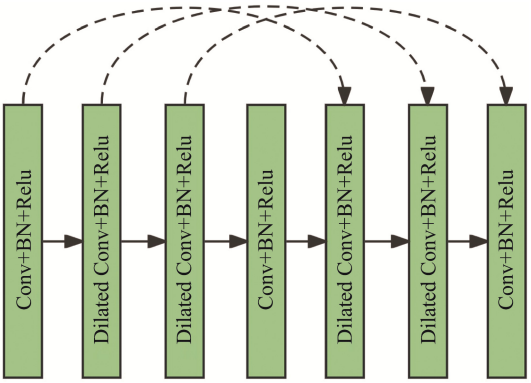


Fig.3 Sparse residual block

2.3 Dilated residual attention construction block

In order to address the issue of insufficient enhancement ability of attention mechanism features in the DRFENet model, this paper propose the Dilated Residual Attention Reconstruction Block(DRAB) which introduces an dilated residual block (DRB) for feature expression enhancement. The structure is depicted in Fig.4. The newly proposed dilated residual block can expand receptive field and encourage model denoising. Unlike the hybrid attention mechanism, the attention mechanism presented in this paper has a simpler structure but effectively improves noise reduction efficiency and quality. The Tanh activation function generally demonstrates a more rapid convergence compared to the Relu and Sigmoid activation function in the training process of neural networks, owing to its advantages such as output range, gradient attributes, and central symmetry. As a result, Tanh is adopted as the activation function in DRAB.

Within the block, represents the multiplication of feature values. The convolution of 2C×1×1×C is used in the first layer of DRAB, where C is the number of channels in the noise image. The convolution is utilized in DRAB to derive weight information and extract significant noise from intricate backgrounds. Finally, the residual learning technology is used in DRAB to reconstract the clear image. DRAB is shown in Equtation(7). The concrete steps of DRAB is shown in Equtation(8).

$$O_{DRAB} = F_{DRAB}(O^1, X_N) \tag{7}$$

$$\begin{aligned} O^1 &= Cat(O_{NEB}, O_{SRB}) \\ O_{noise} &= O^1 \times Conv(Tanh(DRB(O^1))) \\ O_{DRAB} &= X_N - O_{noise} \end{aligned} \tag{8}$$

Where F_{DRAB}, O_{DRAB}, O_{noise}, , O_{noise} ,O^1, Cat, DRBdenote the function of the DRAB, the output of the DRAB, the output of noise, the fuse of NEB and SRB, the skip connection, the Dilated Residual Block, respectively.

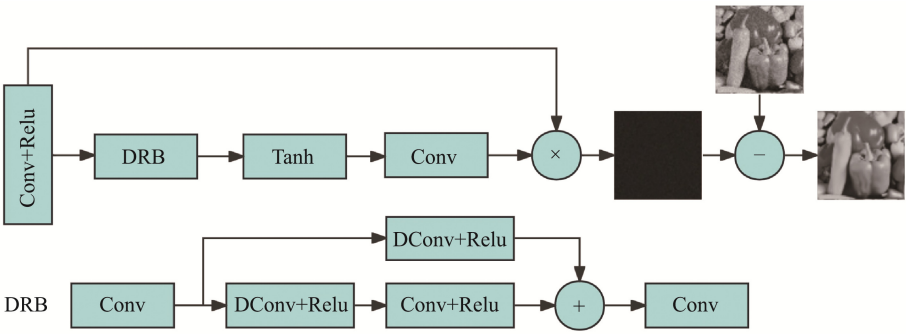


Fig.4 Dilation residual attention reconstruciton block

2.4 Noise extraction block

In contrast to the noise estimation networks in CBDNet and VDN, we design the noise extraction block with 7 convolutional layers that effectively enlarges the receptive field and incorporates two skip connections to enhance details[16]. The output of NEB is combined with the SRB on the DRAB. The structure is depicted in Fig.5. The noise extraction block consists of Conv, Dconv, AvgPool, BN, Relu, and skip connection. The first four layers are dedicated to noise extraction while the last three layers focus on feature enhancement. Multiple skip connections are used to prevent gradient disappearance and explosion, ultimately improving network antagonism and robustness.NEB is shown in Equtation(9).

$$O_{NEB} = F_{NEB}(X_N) \tag{9}$$

Where F_{NEB}, O_{NEB}, X_N denote the function of the NEB, the output of the NEB, the noise image.

2.5 Loss function

The absolute value error (MAE) and mean square error (MSE) are the two most commonly utilized loss functions in the field of image processing. In the realm of image denoising, MAE represents the absolute difference between the residual image and noise. MSE is calculated as the mean of the sum of squares of the differences between residual image and the noise, effectively capturing the average feature weight[21].

The formulas for MAE and MSE are as follows:

$$MAE = \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}|X(i,j)-Y(i,j)| \tag{10}$$

$$MSE = \frac{1}{MN}\sum_{i=1}^{N}\sum_{j=1}^{M}[X(i,j)-Y(i,j)]^2 \tag{11}$$

We use MSE as the loss function to predict residual image. Before model training, it is necessary to process the data with noise. Noise image

$$X_N = X_C + \sigma \tag{12}$$

Where X_C, σ denote the clear image, noise level.

The expression of loss function is as follows

$$L(\omega,b) = \frac{1}{N}\sum_{n=1}^{N}[MASNet(X_N)-\sigma]^2 \tag{13}$$

Where $MASNet(X_N)$, N, ω, b denote the residual image, the number of noise image blocks, weight, bias.

We verify the advantages and disadvantages of MAE and MSE loss functions on the MADRNet model, and the results are shown in Table 2.

3 Results and Discussions

3.1 Evalution indicators

We utilize peak signal-to-noise ratio (PSNR) and structural similarity index(SSIM)[22] as the evaluation principles for image denoising. It is necessary to compute the mean square error (MSE) before calculating PSNR. The formulas for MSE and PSNR are as follows:

$$PSNR = 10\times\log_{10}(\frac{MAX^2}{MSE}) \tag{14}$$

Where M, N, MAX denotes the size of image, and the max pixel of image.

The structural similarity index is calculated based on the luminance, contrast, and structural similarity between two images. The formula for the structural similarity index is

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{15}$$

Where x, y denote the denoising image and clear image. μ_x, μ_y denote the mean of x and y. σ_x, σ_y denote the standard deviation of x and y. σ_{xy} denote the covariance of x and y .c_1, $c_2$denote constant.

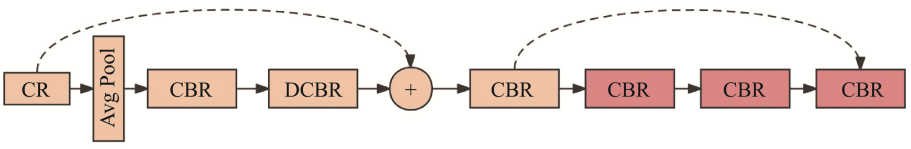


Fig.5 Noise extraction block

Table 2 Denoising performance of different loss function on BSD68

Methods	BSD68	
Noise level	25dB	50dB
MAE	29.15	26.29
MSE	29.27	26.35

3.2 Datasets

In order to verify the diversity of MADRNet model applications, this paper conducts experiments on gray image denoising, color image denoising, and real image denoising. The training sets for gray image denoising and color image denoising experiments include the Berkeley Segmentation Data Set (BSD) and Waterloo Exploration Database. The BSD training dataset consists of 400 images with a size of 180×180, while the Waterloo Exploration Database includes 3859 images. For the gray image denoising experiment, noise intensities of 15dB, 25 dB, and 50 dB are added. For color image denoising, we introduced for known noise levels(σ=15, 25, 35, 50). The noise training dataset is cropped to a size of 50×50 with a step size of 40.The test datasets consist of BSD68 and Set12 datasets for the gray image denoising experiment, CBSD68, Kodak24, and McMaster datasets for the color image denoising experiment. The noise generation method used in the test dataset is consistent with that used in the training dataset.

For real image denoising experiments, this paper utilizes the Smartphone Image Denoising Data set (SIDD) as the training dataset. SIDD contains 320 pairs of real noisy images along with nearly noise-free images which are then cropped into sizes of 128×128. Additionally, SIDD validation Setand Darmstadt Noise Data Set (DND) datasets are selected as test datasets[23].

3.3 Experiment settings

The MASNet model utilizes the ADAM optimizer with parameters 0.9 and 0.999, while setting the batch size to 64 and planning to train 70 epochs. The learning rate changes as follows: Our initial learning rate set 10^{-3} at the first 20 epochs, then every 10 epochs eps was one-tenth of the previous one until reaching 10^{-8}. Device information is listed in Table 3.

Table 3 device imformation

Device Name	Device Information
System	Windows 11
CPU	Intel Core i9-11900K@3.50GHz
GPU	NVIDIA GeForce RTX 3090
Memory	16 GB
Language framework	Anconda+Python3.9.18+Pytorch1.12.0

3.4 Experiments results

In this chapter, we conduct ablation experiments to demonstrate the importance of each block. Then, we proceed with a comparison experiment to verify the denoising efficiency of MADRNet by comparing it with traditional image denoising algorithms and outstanding convolutional neural network models in gray image dataset, color image dataset, and real noise image dataset. The traditional image denoising algorithms include NLM algorithm and BM3D algorithm, while the algorithms based on convolutional neural networks consist of DnCNN, image restoration CNN (IRCNN)[24], image enhancement and denoising CNN (ECNDNet), ADNet, VDN[25] and dilated skip convolutional image denoising CNN (DRFENet). We add the same intensity of noise to all image denoising algorithms and evaluate them using the same training dataset and test dataset.

3.4.1. Ablation experiment

To assess the necessity of each block in the MADRNet model, we conducted a series of ablation experiments: (1) Remove the noise extraction block from MADRNet; (2) Remove the skip connection strategy in the noise extraction block; (3) Remove the shallow network structure DTB from MADRNet; (4) Remove the skip connection strategy from sparse residual block; (5) Remove the dilated residual block from the DRAB; (6) Remove the DTB and NEB from MADRNet; (7) Implement completely the MADRNet. To ensure experimental accuracy, all models were evaluated on DSB68 and Set12 datasets, consisting of 68 images and 12 images, respectively. The average value of ten experiments was chosen as the experimental result, and the same noise map was utilized to attack the images in each batch of experiments. The experimental findings are presented in Table 4.

Table 4 Average PSNR of ablation experiment when noise intensity is 25 dB

Methods	Set12	BSD68
MADRNet without NEB	30.46248	29.19159
NEB without skip connection	30.52175	29.22284
MADRNet without DTB	30.21215	29.01259
SRB without skip connection	30.58612	29.28281
MADRNet without DRB	30.54215	29.24184
MADRNet without DTB and NEB	30.19259	29.00271
MADRNet	30.65312	29.33478

It can be discerned from Table 3 that upon the sole removal of DTB, the denoising performance of the model deteriorates significantly, indicating that DTB assumes the most fundamental and core role within the model. It can utilize the interval transfer strategy to establish connections among different layers and enhance the extraction of noise features. NEB can reinforce local feature information. SRB can enhance the perception capability between the upper and lower layers of the model. DRAB contributes to strengthening the noise separation of the model[26].

Fig.6 illustrates the comparison of ablation

experiment results, highlighting the significance of each block. In Fig.7, the change curve of average PSNR with the number of iterations is presented. It is evident from the figure that, except for ablation experiment 2, the models in other ablation experiments did not converge as the number of iterations increased. Although ablation experiment 2 converged, its denoising effect was not as effective as MADRNet. Therefore, based on the experimental results above, it can be concluded that removing the noise extraction block leads to insufficient shallow noise extraction, reduced global feature learning ability, and compromised model integrity. The removal of dense interval transmission block will result in significant information redundancy and weakened the generalization ability. Skip connections are crucial for contextual feature transfer and model robustness, their removal would interrupt these processes. Additionally, removing DRB diminishes the tail end capability of extracting significant information and weakens the inference ability of the model. When both DTB and NEB are eliminated, the model loses the most fundamental ability of extracting noise features and the efficiency of noise removal is significantly compromised.

3.4.2. Gray image denoising

We conducted gray image denoising experiments with three different noise intensities on the Set12 and BSD68 test datasets respectively. Table 5 presents the average PSNR and SSIM values of different denoising algorithms in the test dataset when noise intensities is 15, 25, and 50 dB. The red marks indicate optimal data. Compared to the current advanced image denoising model DRFENet, our proposed model achieves a higher PSNR value by 0.16dB. Therefore, the model presented in this paper holds significant significance and demonstrates more reliable stability.

Fig.8 illustrates the comparison of the Couple image results after applying different denoising algorithms at a noise intensity of 25dB. The red box highlights the selected local magnification, which is displayed in the

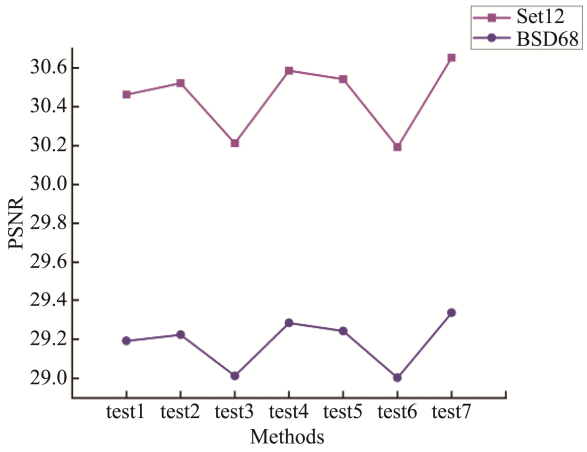


Fig.6 Comparison of ablation experiment

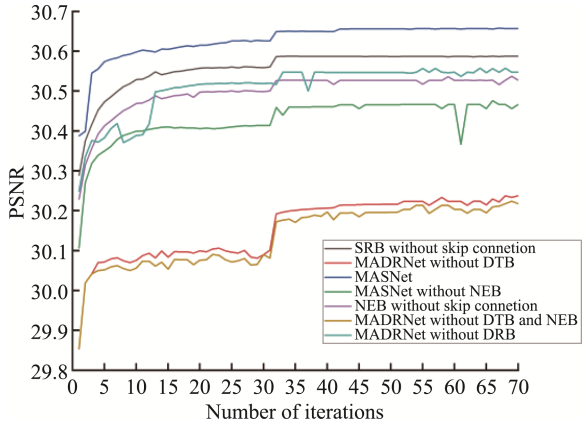
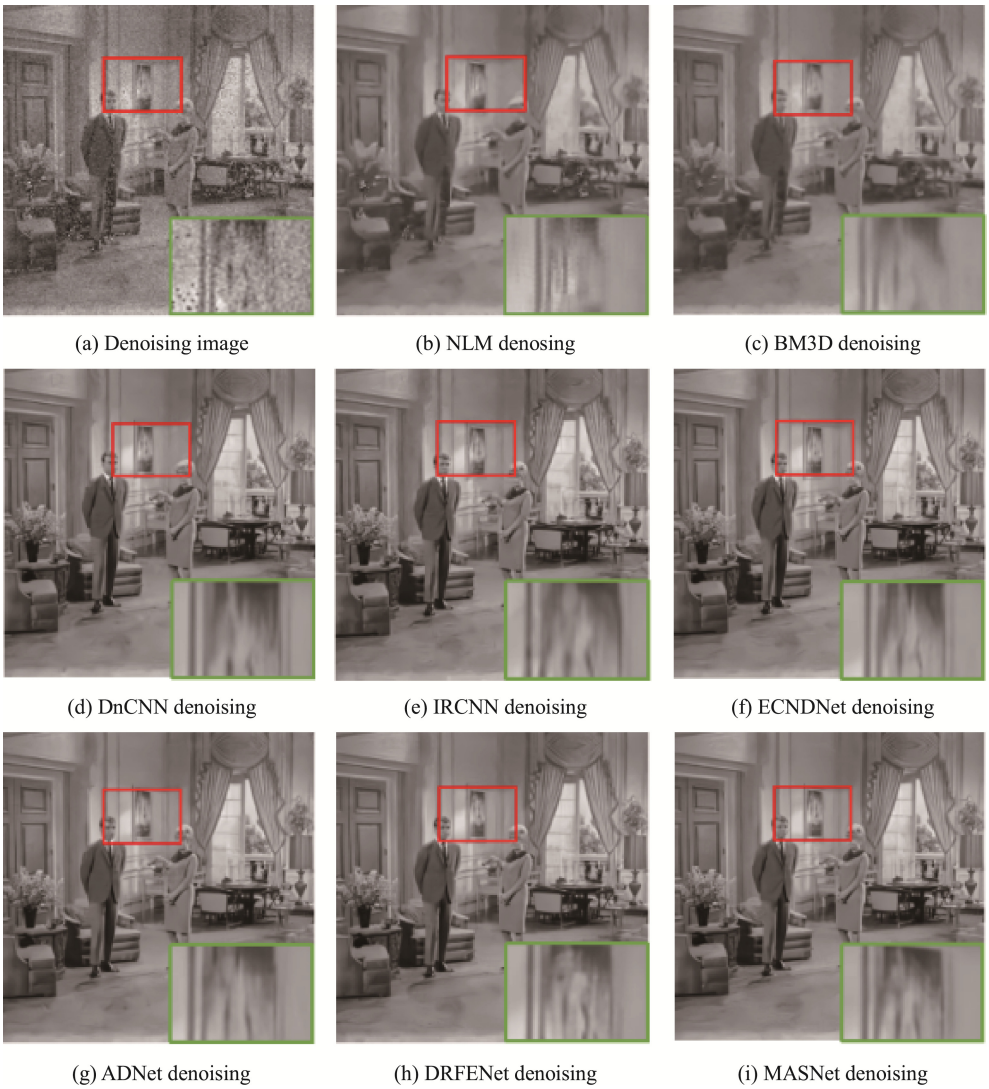


Fig.7 the curve of PSNR with the number of iterations

Table 5 PSNR of different denoising algorithms on Set12 when noise intensity is 25 dB and 50 dB

Datasets	Methods	σ=15dB	σ=25dB	σ=50dB
Set12	NLM	31.31/0.8851	28.27/0.8421	25.50/0.7594
	BM3D	32.46/0.8944	29.88/0.8501	26.65/0.7689
	DnCNN	32.87/0.9027	30.44/0.8617	27.20/0.7828
	IRCNN	32.71/0.8997	30.24/0.8587	27.01/0.7785
	ECNDNet	32.89/0.9034	30.45/0.8621	27.21/0.7831
	ADNet	32.94/0.9037	30.48/0.8631	27.29/0.7874
	DRFENet	32.98/0.9041	30.51/0.8639	27.31/0.7885
	Ours	33.14/0.9049	30.65/0.8647	27.44/0.7885
BSD68	NLM	29.98/0.8629	27.99/0.8024	25.12/0.6878
	BM3D	31.07/0.8720	28.55/0.8121	25.97/0.6968
	DnCNN	31.73/0.8901	29.19/0.8266	26.20/0.7184
	IRCNN	31.54/0.8878	29.04/0.8225	26.08/0.7115
	ECNDNet	31.75/0.8904	29.20/0.8278	26.21/0.7189
	ADNet	31.79/0.8909	29.22/0.8289	26.23/0.7202
	DRFENet	31.82/0.8911	29.23/0.8297	26.25/0.7230
	Ours	31.94/0.8919	29.33/0.8304	26.37/0.7236



(a) Denoising image (b) NLM denosing (c) BM3D denoising

(d) DnCNN denoising (e) IRCNN denoising (f) ECNDNet denoising

(g) ADNet denoising (h) DRFENet denoising (i) MASNet denoising

Fig.8 Result of Boat image for different denoising algorithm when noise intensity is 25 dB

lower right corner. It is evident that both NLM and BM3D algorithms exhibit poor denoising effects, particularly in processing edge information. In contrast, other deep learning image denoising algorithms outperform NLM and BM3D, delivering superior PSNR and better recovery of edge information with almost identical visual effects. Fig.9 depicts the comparison of Parrot image results after employing various denoising algorithms at the noise intensity of 50dB. The red box marks the selected local magnification, which is displayed in the right corner.

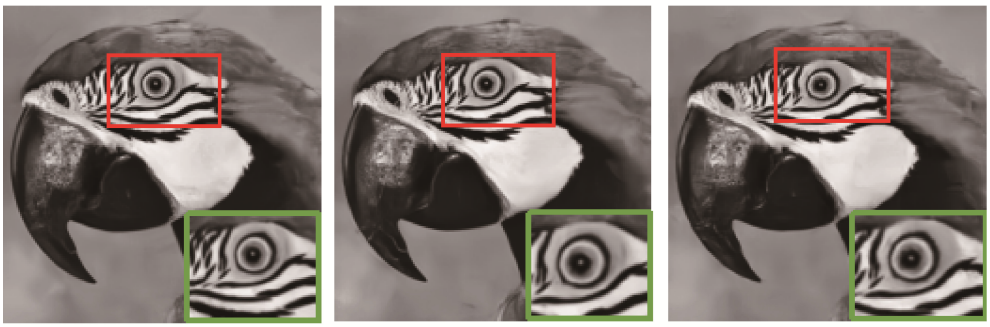
The comparison of processing times for different denoising algorithms on the Set12 dataset is presented in Table 6. The experiment involves processing 12 images from the Set12 dataset and calculating the average processing time as the final result. It is evident from the table that the MASNet model outperforms other deep learning denoising algorithms in terms of image processing time, and it is comparable to traditional image denoising algorithms such as NLM and BM3D. The proposed model in this paper addresses image processing time primarily due to the introduction of the skip connection strategy and DTB. The interval transmission strategy employed in DTB enhances the extraction of noise features, effectively reduces the redundancy of parameters, and shortens the image processing time when combined with the hopping connection strategy in the network.

3.4.3. Color image denoising

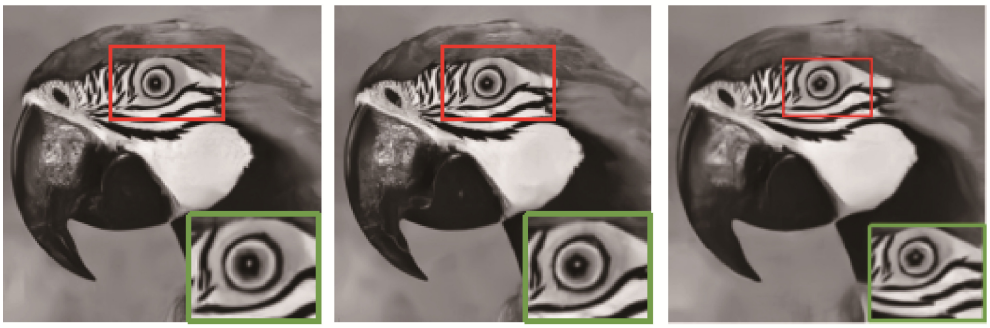
Table 7 presents the comparative experiments on denoising color noise images using color datasets CBSD68, Kodak24, and McMaster. It is evident that MADRNet achieves superior denoising performance across all four noise levels, with an average PSNR improvement of 0.28 dB, 0.17dB, and 0.1dB compared to DnCNN, ADNet, and DRFENet respectively. This demonstrates the strong generalization ability of MADRNet in color image denoising by effectively extracting complex detail information and texture features from color noise images.

Fig.10 illustrates the comparison of kodim23 images in the Kodak24 dataset after being processed by different denoising algorithms at a noise intensity of 25dB. The selected partial magnification is marked by the red box.

(a) Original image (b) NLM denosing (c) BM3D denoising



(d) DnCNN denoising (e) IRCNN denoising (f) ECNDNet denoising



(g) ADNet denoising (h) DRFENet denoising (i) MADRNet denoising

Fig.9 Result of Parrot image for different denoising algorithm when noise intensity is 50 dB

Table 6 Comparison of average runtime

Methods	NLM	BM3D	DnCNN	IRCNN	ECNDNet	ADNet	BRDNet	Ours
Time/s	0.047	0.073	0.19	0.37	0.62	0.47	0.51	0.079

Table 7 Average of PSNR for different denoising algorithms on CBSD68, Kodak24 and McMaster datasets

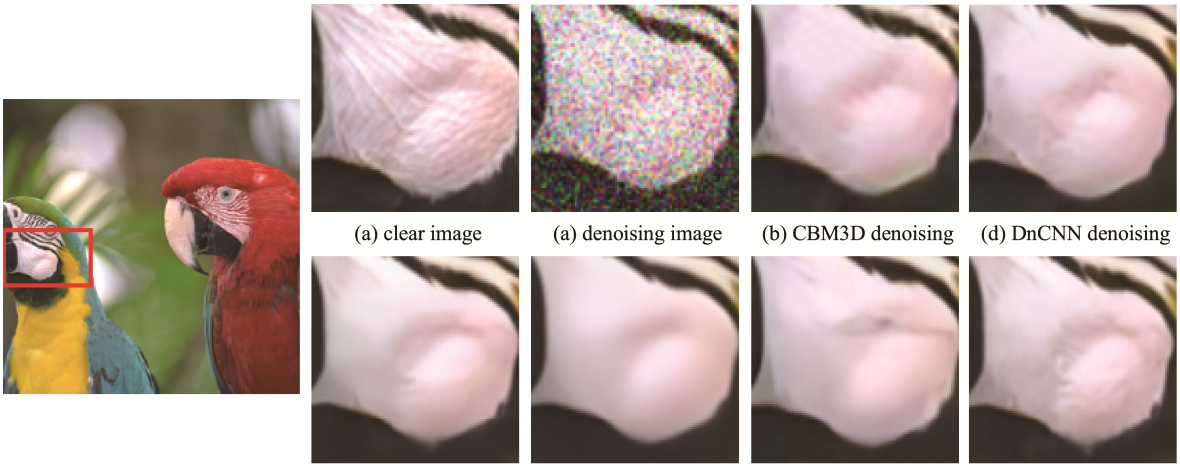
Datesets	Methods	σ=15 dB	σ=25 dB	σ=35 dB	σ=50 dB
CBSD68	CBM3D	33.41	30.71	28.90	27.35
	DnCNN	33.87	31.32	29.66	28.02
	IRCNN	33.77	31.19	29.47	27.89
	ADNet	33.99	31.31	29.66	28.07
	DRFENet	34.01	31.34	29.69	28.12
	Ours	34.14	31.47	29.78	28.19
Kodak24	CBM3D	34.22	31.68	29.90	28.46
	DnCNN	34.68	32.23	30.66	29.04
	IRCNN	34.51	32.02	30.43	28.79
	ADNet	34.77	32.27	30.69	29.11
	DRFENet	34.81	32.33	30.73	29.16
	Ours	34.92	32.42	30.85	29.21
McMaster	CBM3D	34.02	31.66	29.91	28.51
	DnCNN	34.78	32.44	30.89	29.23
	IRCNN	34.66	32.21	30.64	28.94
	ADNet	34.89	32.55	31.01	29.37
	DRFENet	35.01	32.66	31.06	29.49
	Ours	35.17	32.77	31.17	29.55

The outstanding results can be attributed to the network structure of MADRNet, which employs the skip connection strategy to extract more fine-grained features. At the shallow level, we have developed DTB to establish connections between different layers. The feature information obtained from DTB is then passed on to SRB, which captures a multitude of shallow features through mixed convolution and skip connection strategies. Simultaneously, NEB extracts and enhances the noise information extracted in the network. Furthermore, the feature information learned from these two scales is transmitted to DRAB. Through DRB and a lightweight attention mechanism, it effectively enhances the noise separation effect of the model[12].

3.4.4. Real image denoising

To validate the robustness and accuracy of the new model, the real image denoising experiment was conducted. The results in Table 8 demonstrate the performance of different approaches on the SIDD validation set and DND. The new model exhibits superior PSNR and structural similarity index on SIDD, as well as the best PSNR on DND.

Fig.11 illustrates the denoising outcomes of various models on the SIDD verification set. It is evident that noise has been effectively eliminated. In comparison with other models, our proposed model avoids excessive smoothing in local noisy areas while preserving intricate details. Consequently, this model of paper demonstrates commendable robustness and accuracy[27].



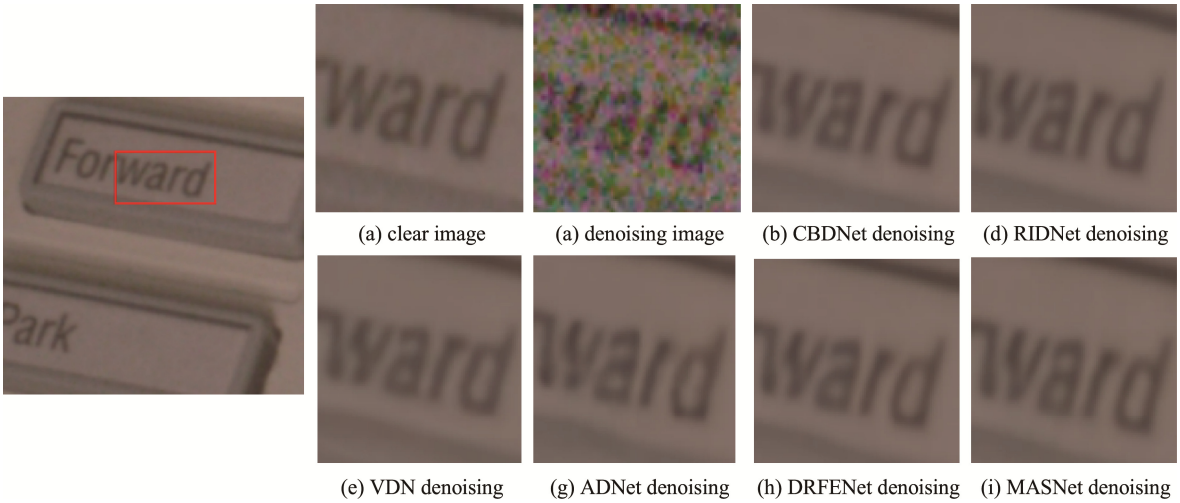
(a) clear image (a) denoising image (b) CBM3D denoising (d) DnCNN denoising

(e) IRCNN denoising (g) ADNet denoising (h) DRFENet denoising (i) MASNet denoising

Fig.10 Result of kodim23 image for different denoising algorithm when noise intensity is 25 dB

Table 8 Comparison of real image denoising experiments

Test Data	SIDD validation					
Methods	CBDNet	RIDNet	VDN	ADNet	DRFENet	Ours
PSNR	38.66	38.70	39.06	39.11	39.19	39.28
SSIM	0.901	0.904	0.907	0.907	0.909	0.912
Test Data	DND					
Methods	CBDNet	RIDNet	VDN	ADNet	DRFENet	Ours
PSNR	38.05	39.25	39.33	39.37	39.45	39.55
SSIM	0.937	0.944	0.952	0.951	0.953	0.951



(a) clear image (a) denoising image (b) CBDNet denoising (d) RIDNet denoising

(e) VDN denoising (g) ADNet denoising (h) DRFENet denoising (i) MASNet denoising

Fig.11 Result of different denoising models on the real noisy image

3.4.5. Discussion and analysis

The MADRNet exhibits an outstanding denoising effect on gray images denoising, as depicted in Fig.8 and Fig.9. MADRNet model adopts the interval transmission strategy based on skip connection. The MADRNet model can not only restore the image more comprehensively but also render the image details smoother, and image possesses an extremely favorable three-dimensional perception.

For color image denoising, the MADRNet model can handle the relationship among image colors and maintain color consistency and naturalness, as shown in Fig.10. MADRNet utilizes the strategy of expanded convolution and skip connection to capture a considerable amount of shallow information for learning the fine detail features of color noise images and enhancing the expression of noise. Meanwhile, NEB can extract a significant amount of noise information and establish the communication between the upper and lower layers, which is highly beneficial for denoising color images.

The real noise image stems from the real environment, and the noise type is complex and non-uniform. The MADRNet demonstrates high adaptability, effectively preserving details and texture, as shown in Fig.11. For real image denoising, DRAB plays a crucial role in MADRNet. This module can adaptively learn the noise distribution and can effectively enhance the noise separation ability of the model.

3.5 Data analysis

To further validate the robustness of the model presented in this paper, we conducted 60 experiments with diverse noise distributions to test the model. Each group of experiments shared the same noise graph. We employed the mean and standard deviation to verify the stability and denoising performance of the algorithm. Table 9 presents the mean and standard deviation of distinct models when the noise intensity is 15, 25, and 35. Evidently, the standard deviation of MADRNet is generally low, indicating that the model proposed in this paper exhibits enhanced performance and superior robustness.

Fig.12 showcases the comparison of error bars of PSNR for various models on dataset Set12 at a noise level of 25. From the figure, it can be seen that by adding error bars, we can evaluate the performance of the proposed model more precisely. Although the increments of PSNR might appear insignificant, we can confirm the statistical significance of these increments through error bar validation, thereby ensuring the validity of the proposed approach in terms of PSNR.

Table 9 The mean PSNR and standard deviation of distinct models when the noise intensity is 15, 25, and 35.

Methods		DnCNN	IRCNN	ECNDNet	ADNet	DRFENet	MADRNet
σ =15	Mean	32.87	32.71	32.89	32.94	32.98	33.14
	Standard deviation	0.000721	0.000737	0.000739	0.000728	0.000737	0.000724
σ =25	Mean	30.44	30.26	30.46	30.48	30.51	30.64
	Standard deviation	0.000985	0.000961	0.000976	0.000944	0.000959	0.000935
σ =35	Mean	27.20	27.01	27.21	27.29	27.31	27.44
	Standard deviation	0.00121	0.00145	0.00131	0.00171	0.00117	0.00198

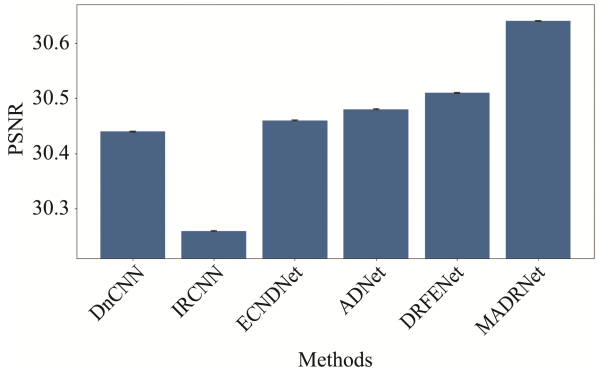


Fig.12 The comparison of error bars of PSNR for various models on dataset Set12 at a noise level of 25

4 Conclusions

In this paper, MADRNet is proposed for image denoising. To enhance the global awareness of model and extract more fine-grained information, we build the MADRNet model by adopting the interval transmission strategy based on skip connection and the strategy integrating the dilated convolution and the skip connection.The new network includes a dense interval transmission block, sparse residual block, dilated residual attention reconstruction block, and noise extraction block. We improve the classical dense connection block, introducing skip connection strategy in the sparse residual block, adding an dilated residual block in the attention mechanism, and using NEB and skip connection to improve the transmission of context information. Ablation experiments and comparison experiments show that our new network can effectively recover local details, has execent model generalization ability and robustness, and possesses mighty global perception ability. Compared with other advanced deep learning denoising models, our new network performs better in peak signal-to-noise ratio, structural similarity index measurement and denoising time. The model proposed in this paper possesses comprehensive functionality and demonstrates exceptional performance in gray, color, and real images denoising. However, it is ineffective in preserving the structural features of the image. Future research will focus on exploring how to

preserve the structural features of images more effectively.

Author Contributions:

Zhiting Du: Conceptualization, Methodology, Software, Writing - Original Draft. Xianchun Zhou: Supervision. Mengnan Lv: Data Curation. Yuze Chen: Formal analysis. Binxin Tang: Writing - review & editing.

Funding Information:

This research was funded by National Nature Science Foundation of China, grant number 61302188.

Data Availability:

The authors declare that the main data supporting the findings of this study are available within the paper and its Supplementary Information files.

Conflict of Interest:

The authors declare no competing interests.

Dates:

Received 23 May 2024; Accepted 9 July 2024; Published online 30 September 2024

References

[1] Li C, Xu X, Guo Y C. (2023). Blind Denoising Algorithm For Single Image Based on Deep Neural Network[J]. *Electronic Measurement Technology*, 46(21), pp.183-192.

[2] Xu X B, Chen B H, Zhao N N, et al. (2022). Improved Gaussian Mean Region Denoising Based on GA-BP[J]. *Journal of Electronic Measurement and Instrumentation*, 36(02), pp.107-113.

[3] Ji C, He L, Dai W. (2022). Double-Norm Constrained Image Denoising Algorithm Based on Dictionary Learning Sparsity and FCM Structure Clustering. *IEEE Access*, 10, pp.128304-128317.

[4] Dauwe A, Goossens B, Luong H.Q, Philips W. (2008). A fast non-local image denoising algorithm. In *Image Processing: Algorithms and Systems VI* , 6812, pp.324-331.

[5] Dabov K, Foi A, Katkovnik V, et al. (2007). Image denoising by sparse 3-D transform-domain collaborative filtering[J]. *IEEE Transactions on image processing*, 16(8), pp.2080-2095.

[6] Guo X E, Liu F, Tian, X T. (2022). Noise Modeling and Denoising of Images Collected by On-Board Track Inspection System[J]. *Multimedia Tools Appl*, 81(8), pp.11695-11715.

[7] Zhang K, Zuo W, Chen Y, et al. (2017). Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising[J]. *IEEE transactions on image processing*, 26(7), pp.3142-3155.

[8] Zhang K, Zuo, W, Zhang L. (2018). FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Transactions on Image Processing*, 27(9), pp.4608-4622.

[9] Tian C, Xu Y, Fei L, et al. (2019). Enhanced CNN for image denoising. *CAAI Transactions on Intelligence Technology*, 4(1), pp.17-23.

[10] Guo S, Yan Z, Zhang K, et al. (2019). Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp.1712-1722.

[11] Tian C, Xu Y, et al. (2020). Attention-guided CNN for image denoising. *Neural Networks*, 124, pp.117-129.

[12] Zhong R, Zhang Q. (2022). DRFENet: An Improved Deep Learning Neural Network via Dilated Skip Convolution for Image Denoising Application. *Applied Sciences*, 13(1), pp.28.

[13] Hu J, Shen L, et al. (2018). Squeeze and excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.7132-7141.

[14] Zheng M, Zhi K, et al. (2022). A hybrid CNN for image denoising. *Journal of Artificial Intelligence and Technology*, 2(3), pp.93-99.

[15] Trung N.T, Trinh D.H, et al. (2020). Dilated residual convolutional neural networks for low-dose CT image denoising. In *2020 IEEE Asia Pacific Conference on Circuits and Systems*, pp.189-192.

[16] Wu W, Liu S, Xia Y, Zhang Y. (2024). Dual residual attention network for image denoising. *Pattern Recognition*, 149, pp.110291.

[17] Hashimot F, Onishi, Y, Ote K, Tashima H, et al. (2024). Deep learning-based PET image denoising and reconstruction: a review. *Radiological Physics and Technology*, pp.1-23.

[18] Thakur R.K, Maji S.K. (2022). Agsdnet: Attention and gradient-based sar denoising network. IEEE Geoscience and Remote Sensing Letters, 19, pp.1-5.

[19] Huang G, Liu Z, Van Der Maaten L, et al. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.4700-4708.

[20] Zhang Q, Xiao J, Tian C, et al. (2023). A robust deformed convolutional neural network (CNN) for image denoising. *CAAI Transactions on Intelligence Technology*, 8(2), pp.331-342.

[21] Tian C, Zheng M, Zuo W, et al. (2023). Multi-stage image denoising with the wavelet transform. *Pattern Recognition*, 134, pp.109050.

[22] Sara U, Akter M, Uddin M S. (2019). Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study[J]. *Journal of Computer and Communications*, 7(3), pp.8-18.

[23] Ma K., Duanmu Z, Wu Q, et al.(2016). Waterloo exploration database: New challenges for image quality assessment models. *IEEE Transactions on Image Processing*, 26(2), pp.1004-1016.

[24] Zhang K, Zuo W, Gu S, Zhang L. (2017). Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp.3929-3938.

[25] Yue Z, Yong H, Zhao Q, Meng D, Zhang L. (2019). Variational denoising network: Toward blind noise modeling and removal. *Advances in neural information processing systems*, 32.

[26] Zhang J, Cao L, Wang T, Fu W, et al. (2022). NHNet: A non-local hierarchical network for image denoising. *IET Image Processing*, 16(9), pp.2446-2456.

[27] Izadi S, Sutton D, Hamarneh G. Image denoising in the deep learning era[J]. *Artificial Intelligence Review*, 2023, 56(7): 5929-5974.