China Instrument and Control Society
Instrumentation
INSTRUMENTATION, Vol. 11, No. 3, September 2024 https://doi.org/10.15878/j.instr.202400198

Article

Image Denoising Using Dual Convolutional Neural Network with Skip Connection

Mengnan Lü[1], Xianchun Zhou[2,*], Zhiting Du[1], Yuze Chen[1], Binxin Tang[1]

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

[2] School of Artificial Intelligence, Nanjing University of Information Science and Technology, Nanjing 210044,China

[*] Corresponding author email: zhouxc2008@163.com

Abstract: In recent years, deep convolutional neural networks have shown superior performance in image denoising. However, deep network structures often come with a large number of model parameters, leading to high training costs and long inference times, limiting their practical application in denoising tasks. This paper proposes a new dual convolutional denoising network with skip connections (DECDNet), which achieves an ideal balance between denoising effect and network complexity. The proposed DECDNet consists of a noise estimation network, a multi-scale feature extraction network, a dual convolutional neural network, and dual attention mechanisms. The noise estimation network is used to estimate the noise level map, and the multi-scale feature extraction network is combined to improve the model's flexibility in obtaining image features. The dual convolutional neural network branch design includes convolution and dilated convolution interactive connections, with the lower branch consisting of dilated convolution layers, and both branches using skip connections. Experiments show that compared with other models, the proposed DECDNet achieves superior PSNR and SSIM values at all compared noise levels, especially at higher noise levels, showing robustness to images with higher noise levels. It also demonstrates better visual effects, maintaining a balance between denoising and detail preservation.

Keywords: image denoising; convolutional neural network; skip connections; multi-scale feature extraction network; noise estimation network

Citation: Lü Mengnan, Xianchun Zhou, Zhiting Du, Yuze Chen, and Binxin Tang. "Image Denoising Using Dual Convolutional Neural Network with Skip Connection." Instrumentation 11, no. 3 (September 2024). https://doi.org/10.15878/j.instr.202400198.

1 Introduction

Image denoising[1] is a fundamental task in computer vision aimed at eliminating noise from low-quality images and restoring their spatial details. However, due to the severe disruption of many key features in the original image by uncertain noise, image denoising remains a challenging problem. In practical applications, image denoising not only enhances the quality and appearance of images but also plays a crucial role in fields such as medical imaging, industrial inspection, and security monitoring. Therefore, researching and improving image denoising algorithms is significant for enhancing the efficiency and accuracy of image processing and analysis.

Currently, image denoising algorithms are primarily categorized into model-based and learning-based approaches. Model-based methods involve modeling natural images or noise distributions, utilizing the modeled distribution as a prior, and obtaining clear images through optimization algorithms. Common prior features include local smoothness, sparsity, non-local self-similarity, and external statistical priors. Image denoising often leverages non-local self-similarity and sparsity to enhance performance. For instance, NLM (Non-Local Means)[2] conducts a block-by-block search for similar areas in the image and averages them to eliminate Gaussian noise. Similarly, BM3D (Block-Matching and 3D Filtering)[3] identifies similar two-dimensional image blocks through similarity determination, integrates them into three-dimensional groups for collaborative filtering, and aggregates them to produce a denoised image at the

original block position in the image. The WNNM (Weighted Nuclear Norm Minimization method)[4] matches each block to generate a similar block matrix, which is then shrunk based on the characteristics of singular values to estimate denoised blocks. This approach effectively preserves texture details while reducing noise, making it suitable for simple scenes and low noise levels. However, these algorithms suffer from issues such as information loss, challenging parameter selection, high computational complexity, and limited adaptability. In contrast, deep learning algorithms exhibit superior processing capabilities and generalization abilities by acquiring complex image features and patterns. Consequently, they often achieve better denoising performance in handling intricate scenes and high noise levels.

Due to the flexible connections and strong learning capabilities inherent in deep network frameworks, they have emerged as a highly effective solution for addressing image denoising problems. With the widespread adoption of deep neural networks, learning-based denoising methods have witnessed rapid advancements[5-8]. For instance, in 2017, Zhang et al. proposed DnCNN (denoising convolutional neural network)[9] specifically designed for denoising Gaussian AWGN (Additive White Gaussian Noise), highlighting the significance of residual learning and batch normalization techniques while showcasing the exceptional generalization ability of convolutional neural networks. However, it is important to note that DnCNN primarily focuses on a specific type of noise; hence, more extensive network depth and complexity may be required to effectively address other complex noise models. In 2018, Zhang et al. introduced FFDNet (Fast and Flexible Denoising Network)[10], which extends its applicability beyond Gaussian noise by incorporating a noise level map as part of the network input, thereby making it suitable for images with varying levels of noise intensity. Nevertheless, it should be acknowledged that FFDNet may not perform optimally when dealing with certain types of noise or images characterized by high levels of noise intensity. Subsequently in 2019, Zhang et al. presented CBDNet (Cross-Scale Bi-Directional Network)[11], which adopts a more realistic approach by considering Poisson-Gaussian noise along with signal-dependent and ISP-induced noises effects during training using both synthetic and real noisy images to enhance adaptation to real-world scenarios. However, it is worth mentioning that this particular network necessitates an extensive amount of training data while also being time-consuming to train due to its demanding hardware resource requirements.

Despite the notable achievements of convolutional neural network techniques in image denoising, current methods still face challenges in effectively extracting and leveraging intricate features present in images, thereby limiting the overall performance of models. This limitation can significantly impede the efficacy of denoising, particularly when dealing with complex tasks.

Research findings suggest that the integration of skip connections[12] and dilated convolutions[13] can effectively facilitate the development of deep image denoising models. Dilated convolutions are capable of expanding the receptive field of convolutional layers, while skip connections not only expedite network training but also enhance feature preservation during feature transposition. Recent advancements in denoising models such as IRCNN[14], BRDNet[15], and RDUNet[16] have demonstrated remarkable performance by leveraging these techniques.

To address these challenges, this paper proposes a novel Dual Convolutional Denoising Network with skip connections (DECDNet) to effectively achieve denoising. The proposed model comprises a noise estimation network, a multi-scale feature extraction network, dual convolutional neural networks, and a dual attention mechanism network. In the upper layer of the dual convolutional neural network, both convolution and dilated convolution are employed in an interleaved manner while incorporating downsampling and upsampling operations. The lower network incorporates dilated convolution layers to expand the receptive field. Both subnetworks utilize skip connections to integrate features from different convolutional layers. Finally, the features from these two subnetworks are concatenated to generate the denoised image. The proposed DECDNet offers several advantages:

1. A noise estimation network with an extensive receptive field is devised to accurately estimate the noise level map, which, together with the noisy image, is inputted into the multi-scale feature extraction network for extracting more informative details.

2. The dual convolutional denoising model of DECDNet employs downsampling operations and dilated convolutions to enlarge the receptive field while incorporating skip connections to preserve intricate image details.

3. After the dual convolutional denoising model, we introduce a dual attention mechanism to simultaneously focus on both channel and spatial dimensions. This approach effectively learns generalized weights, thereby enhancing preservation of useful image features and enabling the model to handle various complex noise situations more efficiently.

2 Related Work

2.1 Skip Connections

Skip connection structures, also referred to as residual connection structures, are architectural design strategies aimed at addressing optimization and information transfer challenges in deep neural networks.When the number of layers exceeds ten without incorporating skip connections,deep neural networks often encounter optimization challenges[17-19]. Skip connections were initially introduced in ResNet, introducing residual blocks that enable direct information

transfer through skip connections. Inspired by this concept, other network architectures have also adopted skip connections to facilitate information transfer. For example, the Transformer model utilizes skip connections to mitigate gradient vanishing issues during deep network training[20]. By improving gradient flow and reducing optimization difficulties, skip connections enhance feature diversity and expression capability while maintaining a balance between the network's width and depth through efficient information propagation.

2.2 Attention Mechanism

The attention mechanism, initially introduced in the field of machine translation, has emerged as a fundamental theoretical underpinning in deep learning with the continuous advancement of neural network research[21]. Drawing inspiration from the human visual system, which relies on selectively attending to relevant information in an image to facilitate visual judgment while disregarding irrelevant details, the attention mechanism plays a pivotal role. In scenarios involving multiple objects, attention can be flexibly allocated to each target based on its specific requirements.

The attention mechanism principle involves generating a feature weight matrix to identify input features, which are subsequently utilized by the neural network for learning the salient regions within an image. ADNet (Attention-Guided Dilated Network)[22] employs a singular convolutional channel of focus information to guide CNN training models. In parallel, CBAM (Convolutional Block Attention Module)[23], a lightweight attention mechanism module proposed by Woo et al[24]., integrates spatial and channel attention mechanisms to augment feature representation. Motivated by this, we propose a dual attention mechanism that simultaneously emphasizes both spatial and channel dimensions in order to effectively capture intricate image details and textures.

3 The proposed model

3.1 Network architecture

The proposed DECDNet comprises a noise estimation network, a multi-scale feature extraction network, dual convolutional denoising networks, and a dual attention mechanism. The architecture of DECDNet is depicted in Fig.1. Firstly, the noise estimation network acquires the noise level map from the noisy image, which is then combined with the noisy image and inputted into the multi-scale feature extraction network. Subsequently, denoising is conducted through the dual convolutional neural denoising networks. Lastly, the dual attention mechanism network focuses on both channel and spatial dimensions to acquire generalized weights that effectively preserve valuable image features and generate the ultimate denoised image.

3.1.1 Noise Estimation Network

To achieve effective blind image denoising, we construct a Noise Estimation Network (NENT) to accurately estimate the noise level map. As illustrated in Fig.2, the network consists of seven convolutional layers with kernel sizes of 3×3×64. The overall architecture incorporates convolution, upsampling, downsampling, BN, ReLU and Tanh[25] activation functions.

By employing downsampling and upsampling techniques, we expand the network's receptive field to capture more intricate noise features and enhance the accuracy of noise level map estimation. To mitigate potential loss of image detail caused by downsampling, skip connections are introduced within the network to directly transfer and integrate fine features from lower layers. This compensates for possible detail loss and ensures preservation of original image information during noise estimation. The expression for the noise estimation network is presented in Equation (1),

$$O_{NE}=\mathrm{Tanh}(C(C_{BR}(C_{BR}(C_{BR}(C_{BR}(C_{B}(N)))))))) \quad (1)$$

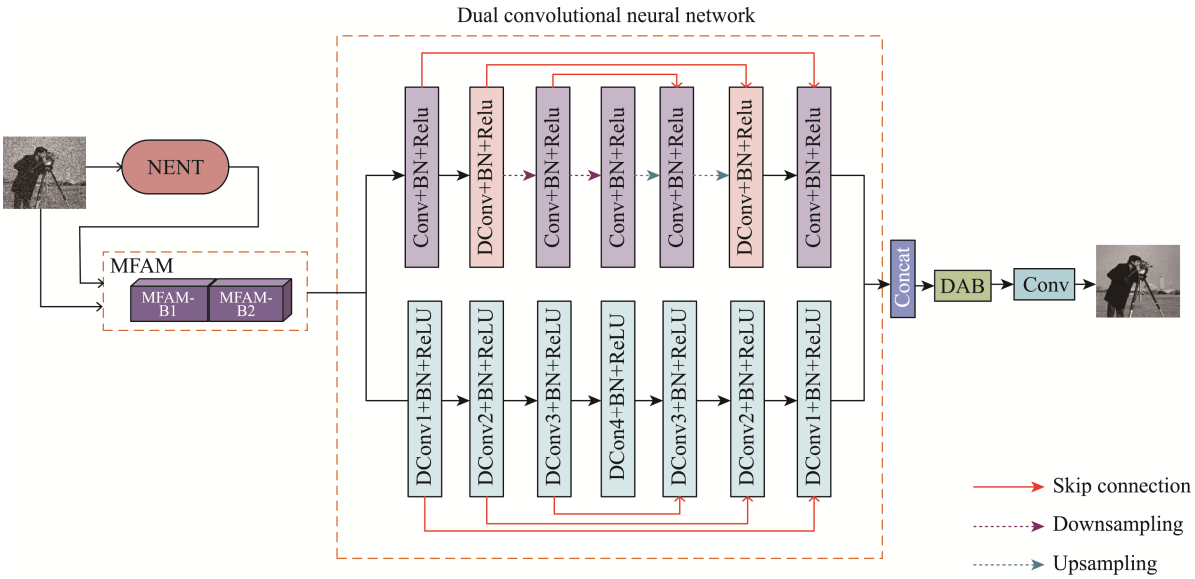

Fig.1 DECDNet network framework

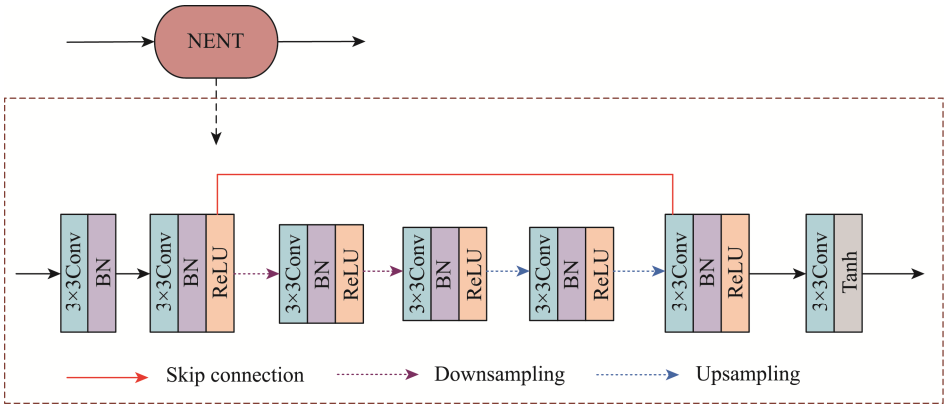

Fig.2 NENT

where O_{NE} denotes the output of the noise estimation network, N represents the noisy image, C, C_{BR}, and C_B are functions.

3.1.2 Multi-Scale Feature Extraction Module

In the field of image denoising research, sparse representation[26] and multi-scale feature analysis[27] have been extensively validated for their effectiveness. To optimize the network's receptive field and enhance feature extraction capabilities, we propose a Multi-Scale Feature Extraction Module (MFAM). This module seamlessly integrates the fundamental principles of sparse representation and multi-scale feature analysis, leveraging sparse representation to extract crucial information while expanding the range of captured features through multi-scale analysis. Consequently, our model ensures an abundance of comprehensive feature information in subsequent processing stages, effectively mitigating performance bottlenecks caused by inadequate shallow feature extraction. MFAM comprises two subnetworks, namely MFAM-B1 and MFAM-B2, which employ dilated convolutions and convolutional layers with varying kernel sizes to extract and fuse features at different scales.

As illustrated in Fig.3, MFAM-B1 employs the concept of sparse representation with a structure composed of three sets of sequentially connected convolutional networks. Each set consists of dilated convolutions, batch normalization (BN), and PReLU activation functions. The dilation rates for the three sets of dilated convolutions are 1, 2, and 3 respectively. By stacking these three dilated convolutions, MFAM-B1 achieves an extensive receptive field that precisely focuses on pixel points and efficiently extracts features. The expression for MFAM-B1 is presented in Equation (2),

$$O_{MFAM-B1}=D_3(D_2(D_1(N))) \tag{2}$$

where N represents the input noisy image, D_3, D_2, and D_1 denote dilated convolution functions with dilation rates of 1, 2, and 3 respectively; $O_{MFAM-B1}$ denotes the output of the MFAM-B1 block.

The MFAM-B2 module receives the sparse feature maps processed by MFAM-B1 as input. As illustrated in Fig.4, the input feature maps are fed into four independent parallel processing branches within MFAM-B2, each consisting of 64 channels. Initially, Conv+BN+ReLU operations are applied to each branch to ensure consistent feature content output at this stage. Subsequently, convolutions with kernel sizes of 1, 3, 5, and 7 are performed on the feature maps of each branch individually to capture multi-scale feature information. Following this step, BN+PReLU operations further optimize the expression capabilities of each branch's feature maps. It is worth noting that the PReLU activation function is selected due to its ability to adaptively learn parameter settings during training, thereby enhancing model accuracy without significantly increasing computational burden. Finally, concatenation is employed to fuse the multi-scale features extracted from the four branches together, resulting in more powerful and comprehensive feature information output. The expression

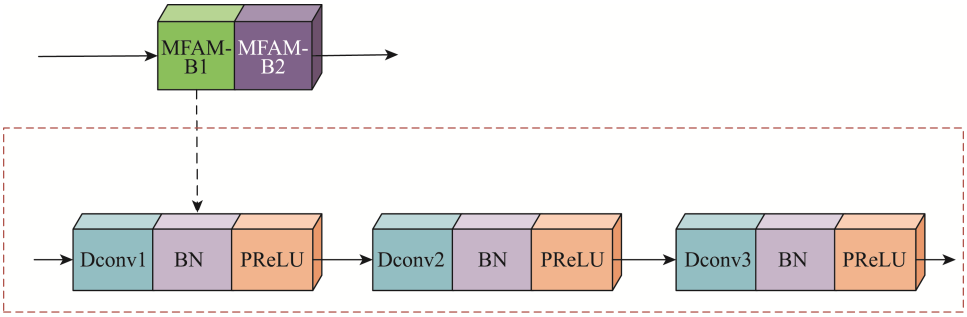

Fig.3 MFAM-B1

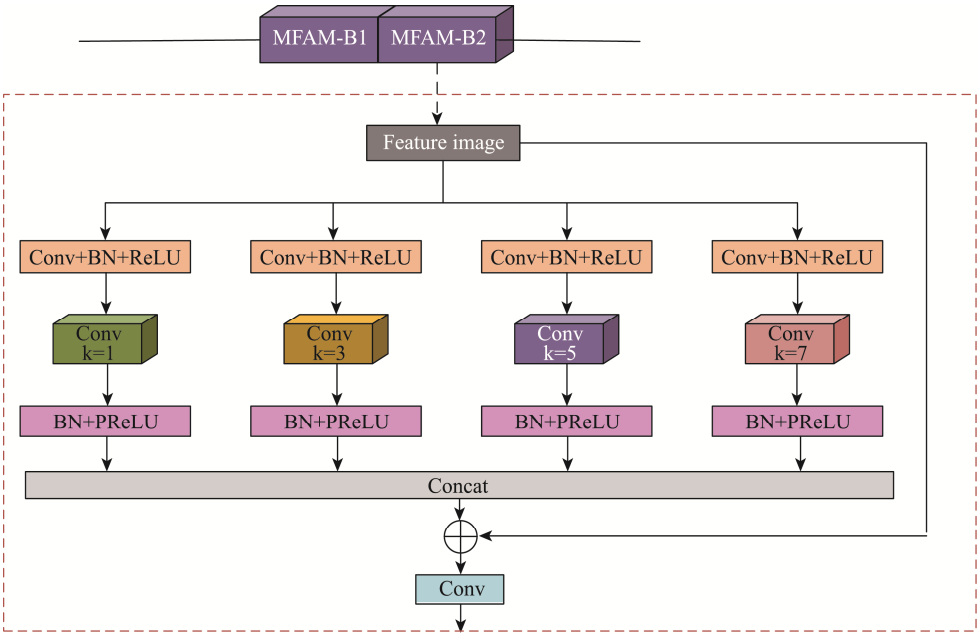

Fig.4 MFAM-B2

for MFAM-B2 can be represented by Equation (3),

$$O_{MFAM-B2} = Concat(O_1, O_2, O_3, O_4) \tag{3}$$

$$\begin{aligned} O_1 &= f_{k=1}(O_{MFAM-B1}) \\ O_2 &= f_{k=3}(O_{MFAM-B1}) \\ O_3 &= f_{k=5}(O_{MFAM-B1}) \\ O_4 &= f_{k=7}(O_{MFAM-B1}) \end{aligned} \tag{4}$$

where $f_{k=1}$, $f_{k=3}$, $f_{k=5}$, and $f_{k=7}$ represent convolution functions with different kernel sizes; Concat represents the multi-scale feature fusion function; and $O_{MFAM-B2}$ represents the output of MFAM-B2.

3.1.3 Dual Convolutional Neural Network

Previous studies have demonstrated that increasing the width of the network can significantly enhance overall network performance[28]. Building upon this finding, we propose a dual convolutional neural network architecture comprising two parallel subnetworks to augment denoising capabilities by expanding the network width. This novel structure encompasses an upper and lower subnetwork, synergistically amplifying the model's expressive capacities.

The upper subnetwork employs convolution, dilated convolution, upsampling, and skip connections to effectively capture global features and provide contextual information. The alternating utilization of standard and dilated convolutions facilitates the integration of information across a wider range, while batch normalization is applied after each convolution layer to optimize computational complexity. Both the upper and lower subnetworks consist of seven 3×3×64 convolution layers, striking a balance between network complexity and denoising performance.

We introduces three sets of symmetric skip connections in both the upper and lower subnets of the convolutional neural network. Despite its shallow depth, these skip connections play a pivotal role in our design. Firstly, they facilitate smoother information flow within the network, mitigating information loss during layer-by-layer propagation and enhancing the network's expressive capacity. Secondly, they alleviate gradient vanishing issues by ensuring effective gradient transmission even in shallow networks, thereby improving model training effectiveness. Additionally, the utilization of three sets of symmetric skip connections enables effective fusion of features at different levels by integrating detailed low-level features with abstract high-level information, resulting in superior performance for image denoising tasks. Lastly, these skip connections enhance model stability by preventing unstable gradient changes during training and ensuring smoothness and stability throughout the entire training process.

The downsampling operations in the upper subnetwork and the dilated convolutions in the lower subnetwork are strategically designed to concurrently expand their receptive fields, bringing them closer in size for capturing more comprehensive contextual information and finer local features. Drawing inspiration from hybrid dilated convolution (HDC)[29,30], we assigned distinct dilation rates to each sequential dilated convolution as 1, 2, 3, 4, 3, 2, 1 respectively to mitigate potential gridding effects and significantly enhance the denoising performance of our network.

3.1.4 Dual Attention Mechanism

This paper introduces a dual attention mechanism DAB after applying a denoising neural network. The purpose of this design is to ensure that DAB pays equal attention to both channel and spatial dimensions in order to learn generalized weights, resulting in more precise image details and textures. Compared to single mechanisms like CBAM or ECA, DAB can significantly enhance information flow within the network, leading to improved performance in image denoising tasks.

Furthermore, DAB exhibits high flexibility and adaptability, making it highly effective across different complex network structures and task scenarios[31]. As illustrated in Fig.5, the Channel Attention Block (CAB) [32] and Spatial Attention Block (SAB)[33] operate concurrently, $\odot$ denotes element-wise multiplication and $\oplus$ denotes element-wise addition. CAB computes channel attention and information weights to generate a channel map (C_M), employing Equation (5),

$$C_M = x_{in} \odot \sigma(f^2 \xi(f^1 P(x_{in}))) \tag{5}$$

where $\xi(\cdot)$ and $\sigma(\cdot)$ represent ReLU and Sigmoid functions, respectively, and $P(\cdot)$ represents GAP. SAB generates a spatial map (S_M) with the formula shown in Equation (6),

$$S_M = x_{in} \odot \sigma(f^1 C(Avgpool(x_{in}), Maxpool(x_{in}))) \tag{6}$$

where $C(\cdot)$ represents the convolution operation,and $f(\cdot)$ is the combination symbol for convolutional operations. Finally, CM and SM are added to obtain the output of DAB.

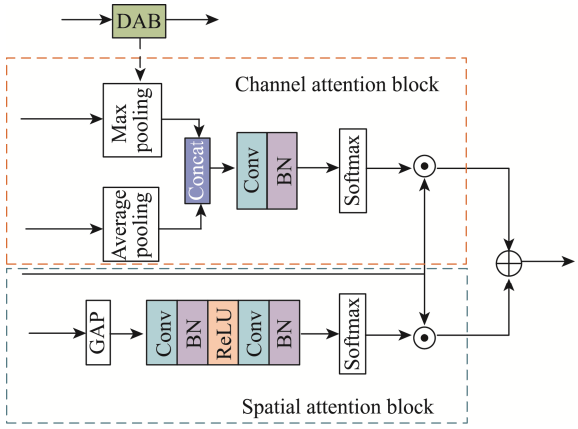

Fig.5 Dual Attention Block

3.2 Loss Function

To train DECDNet for Gaussian white noise removal, the Mean Squared Error (MSE) was selected as the loss function. Let x_j represent the clean image and y_j denote the noisy image. The noise mapping x=y_j–F(x_j) is predicted through residual learning for a given training set. By minimizing this loss function, optimal parameters are obtained.The expression for $L(\theta)$ can be represented by Equation (7),

$$L(\theta) = \frac{1}{2K}\sum_{j=1}^{Z}\left\|F(y_j;\theta) - x_j\right\|^2 = \frac{1}{2K}\sum_{j=1}^{Z}\left\|\hat{x} - x_j\right\|^2 \tag{7}$$

where x_j, $\hat{x}$, y_j represent the clean images, predicted images,and noisy images,respectively, j represents the jth pair of images, θ are the trainable network parameters, and K is the number of noisy image patches.

4 Experiments and results

4.1 Denoising Evaluation

We conducted a comprehensive evaluation of the proposed model's denoising performance, both quantitatively and qualitatively. For quantitative analysis, we employed objective metrics such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Measurement (SSIM) to assess the quality of the denoised images. In addition, for qualitative analysis, we compared the visual effects of denoised images obtained using different methods, while considering subjective evaluation metrics based on human visual perception of image quality.

4.2 Datasets

To eliminate Gaussian white noise, we employed the DIV2K dataset[34] to train our model. This dataset consists of 800 high-resolution color images for training and 100 for validation purposes. The training images were rescaled to a size of 512×512 and converted to grayscale in order to train the grayscale image denoising model effectively. Random cropping was performed on these images, resulting in patches sized at 180×180, which were then used for training the grayscale image denoising model. By utilizing this relatively large patch size, DECDNet is able to capture more contextual features, thereby enhancing its denoising performance particularly under high noise levels. Additionally, each clean image patch was subjected to additive white Gaussian noise (AWGN) with a noise level range of [0, 50], thus generating noisy patches.

We utilized the Set12 and BSD68 datasets to evaluate grayscale image denoising[35,36], while the Kodak24 and McMaster datasets were employed for color image denoising evaluation. The BSD68 dataset comprises 68 grayscale images with pixel dimensions of 321×481, whereas the Set12 dataset consists of 12 grayscale images with pixel dimensions of 256×256. Moreover, the Kodak24 dataset contains 24 color images with a resolution of 500×500, while the McMaster dataset includes 18 color images.

4.3 Experimental settings

The experiments were conducted on a computer equipped with a 16-core Intel(R) Core(TM) i7-11700KF CPU running at 2.50 GHz, 32 GB RAM, and an NVIDIA GeForce RTX 4070Ti GPU. The proposed DECDNet was trained for grayscale and color image denoising, taking approximately 48 hours to complete the training process. Model parameters were optimized using the Adam optimizer with an exponential decay rate of β_1=0.9, β_2=0.999, a constant ε=1.0×10^{-8}, and Mean Squared Error (MSE) as the loss function. DECDNet underwent training for 70 epochs to remove Gaussian white noise, starting with an initial learning rate of 1×10^{-4} which was halved every 10^5 iterations while maintaining a batch size of 16.

4.4 Denoising Performance Comparison

In this section, we present the denoising technique for grayscale and color images corrupted by Additive White Gaussian Noise (AWGN). To create noisy images, AWGN with noise levels of 15, 25, and 50 was added to

the original images. The top two denoising results are indicated by red and blue numbers in Tables 1-5.

4.4.1 Grayscale Image Denoising Evaluation

For grayscale image denoising, we compared DECDNet with BM3D[3], WNNM[4], DnCNN[9], FFDNet[10], IRCNN[14], DudeNet[37], and DRANet[38]. Denoising tests were conducted on grayscale noisy images from the BSD68 and Set12 datasets, where AWGN noise levels of 15, 25, and 50 were added to the original images. The corresponding PSNR values are presented in Table 1. Our model consistently outperformed other models by achieving higher average PSNR across all tested noise levels, particularly at higher noise levels. At a noise level of 15, DECDNet exhibited an average PSNR that was superior by 0.01 dB compared to DRANet due to its larger receptive field which reduced sensitivity towards weak noise in low-frequency image regions. Moreover, at noise levels of 25 and 50, DECDNet achieved PSNR improvements of 0.06 dB and 0.15 dB respectively, indicating enhanced robustness against higher noise levels. Additionally, DECDNet possesses a more compact network structure than DRANet.

The average SSIM of various denoising methods on the Set12 dataset is presented in Table 2, with evaluation conducted using noisy images at noise levels of 15, 25, and 50. DECDNet's average SSIM performance matches that of DRANet at a noise level of 15, indicating comparable recognition capabilities. However, when faced with more challenging higher noise levels, our model significantly outperforms DRANet, thereby further validating its superior recognition capabilities.

The grayscale image denoising performance on the BSD68 dataset was also assessed at noise levels of 15, 25, and 50. Table 3 presents the average PSNR and SSIM values obtained from various denoising methods. Our proposed model demonstrated superior performance compared to others at noise levels of 25 and 50, while remaining highly competitive at a noise level of 15.

The denoising results of various methods on image Test027 from the BSD68 dataset at a noise level of 50 are presented in Fig.6. A detailed comparison (highlighted by the green box) of a magnified region (indicated by the red box) reveals that BM3D, WNNM, IRCNN, and DnCNN partially reduced noise but exhibited slight blurring. FFDNet, DudeNet, and DRANet suffered from significant loss of image textures and details. In contrast, DECDNet achieved superior visual effects by effectively balancing denoising and detail preservation. These results demonstrate the excellent performance of our proposed DECDNet model both subjectively and objectively.

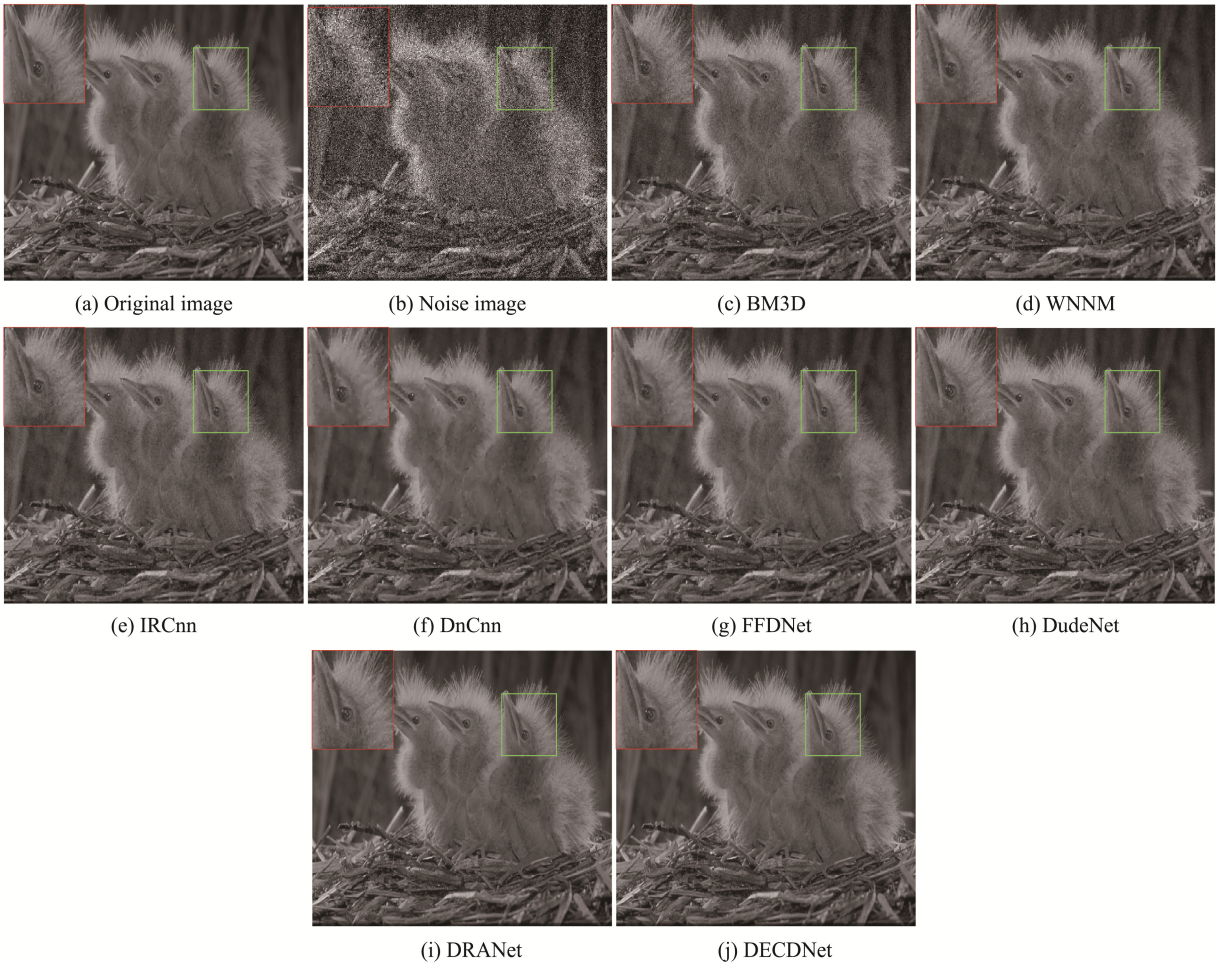

Fig.6 Denoising results of test027 image in BSD68 with different networks (σ=50)

Table 1 The averaged PSNR (dB) results of different networks on Set12, with the two best PSNR results displayed in red and green fonts

Methods	BM3D	WNNM	DnCnn	FFDNet	IRCNN	ADNet	DRANet	DECDNet
σ = 15	32.37	32.70	32.86	32.77	32.77	32.94	33.00	33.01
σ = 25	29.97	30.26	30.43	30.44	30.38	30.52	30.69	30.75
σ = 50	26.72	27.05	27.18	27.32	27.14	27.30	27.62	27.77

Table 2 The averaged SSIM results of different networks on Set12, with the two best PSNR results displayed in red and green fonts

Methods	BM3D	WNNM	DnCnn	FFDNet	IRCNN	ADNet	DRANet	DECDNet
σ = 15	0.896	0.894	0.903	0.903	0.901	0.905	0.906	0.906
σ = 25	0.851	0.846	0.862	0.864	0.860	0.865	0.868	0.870
σ = 50	0.766	0.766	0.783	0.791	0.780	0.791	0.800	0.807

Table 3 The averaged PSNR and SSIM comparisons of different networks on the BSD68 dataset

	Methods	σ = 15	σ = 25	σ = 50
PSNR	BM3D	31.07	28.57	25.62
	WNNM	31.37	28.83	25.87
	DnCnn	31.72	29.23	26.23
	FFDNet	31.63	29.19	26.29
	IRCnn	31.63	29.15	26.19
	DudeNet	31.78	29.29	26.31
	DRANet	31.79	29.36	26.47
	DECDNet	31.78	29.39	26.55
SSIM	BM3D	0.872	0.802	0.687
	WNNM	0.878	0.810	0.698
	DnCnn	0.891	0.828	0.719
	FFDNet	0.890	0.830	0.726
	IRCNN	0.888	0.825	0.717
	ADNet	0.892	0.829	0.722
	DRANet	0.892	0.833	0.732
	DECDNet	0.891	0.835	0.737

Table 4 The averaged PSNR (dB) results of different networks on the Kodak24 and McMaster datasets

Datasets	Methods	σ = 15	σ = 25	σ = 50
Kodak24	CBM3D	34.28	31.68	28.46
	CDnCnn-S	34.48	32.03	28.85
	FFDNet	34.63	32.13	28.98
	BRDNet	34.88	32.41	29.22
	ADNet	34.76	32.26	29.1
	AirNet	34.68	32.21	29.06
	DECDNet	34.92	32.48	29.39
McMaster	CBM3D	34.06	31.66	28.51
	CDnCnn-S	33.44	31.51	28.61
	FFDNet	34.66	32.35	29.18
	BRDNet	35.08	32.75	29.52
	ADNet	34.93	32.56	29.36
	AirNet	34.70	32.44	29.26
	DECDNet	35.00	32.80	29.65

Table 5 The averaged SSIM results of different networks on the Kodak24 and McMaster datasets

Datasets	Methods	σ = 15	σ = 25	σ = 50
Kodak24	IRCnn	0.920	0.877	0.793
	FFDNet	0.922	0.878	0.794
	ADNet	0.924	0.882	0.798
	AIRNet	0.924	0.882	0.799
	DECDNet	0.926	0.886	0.809
McMaster	IRCnn	0.920	0.882	0.807
	FFDNet	0.922	0.886	0.815
	ADNet	0.927	0.894	0.825
	AIRNet	0.925	0.891	0.822
	DECDNet	0.927	0.896	0.836

4.4.2 Color Image Denoising Evaluation

The denoising performance of DECDNet was compared with CBM3D[3], CDnCnn-S[5], FFDNet[10], BRDNet[15], ADNet[22], and AIRNet[38] for color image denoising. The Kodak24 and McMaster datasets were utilized to evaluate the denoising performance of our model on color images. Table 4 presents the PSNR values obtained by different methods, demonstrating that our proposed model outperforms all other compared methods at noise levels of 25 and 50. Particularly, at a noise level of 50, our proposed model achieves higher PSNR values than the second-best method, BRDNet, by 0.17 dB and 0.13 dB respectively on both datasets. However, at a noise level of 15, our model's PSNR value on the McMaster dataset is slightly lower than that of BRDNet by 0.08 dB due to BRDNet's employment of a dual CNN structure with stronger learning capability under low noise power conditions. Table 5 displays the average SSIM values which indicate that our model consistently outperforms others across different noise levels. Remarkably, at a noise level of 50, our model surpasses AIRNet by an SSIM value increment of 0.01 on Kodak24 and ADNet by an increment of 0.011 on McMaster dataset respectively.The visual results of color image noise removal at a noise level of 50 are shown in Fig.7, using an image from the McMaster dataset.It can be observed that some comparison methods tend to over-smooth while DECDNet better preserves details.

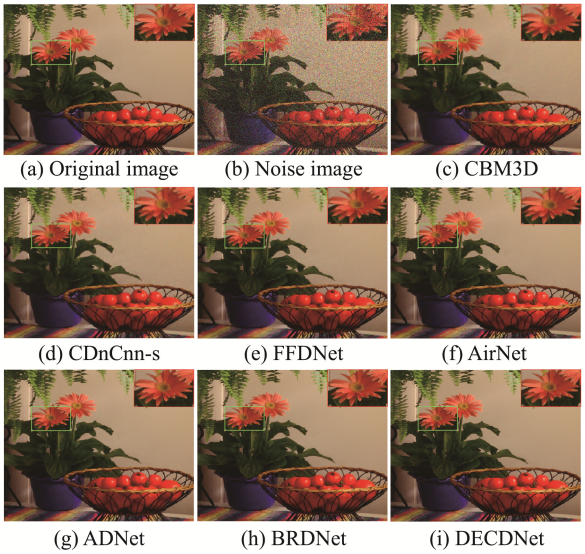

Fig.7 Denoising results of image in McMaster with different networks (σ=50)

4.5 Network complexity analysis

In our experiments, the network complexity was evaluated from the perspectives of model running time and numbers of network parameters. The DnCNN[9], FFDNet[10], IRCNN[14], ADNet[22], AirNet[36] and DRANet[37] were utilized for comparison. All denoising models were implemented in the PyCharm (2021) environment.

We first evaluated the runtime of different denoising methods, and the results are listed in Table 6. Three grayscale images and color images of different sizes were randomly selected for evaluating their runtime at a noise level of 25. The average time from 20 runs was calculated to determine the runtime for each model on each image. Notably,models such as DRANet and AirNet, which demonstrated superior PSNR performance, required longer denoising times, particularly for larger image sizes. Additionally, these models with intricate structures also necessitated lengthier training periods.

Table 6 Running time (in seconds) of the evaluated denoising methods on three grayscale and color images with different sizes

Devices	Methods	256×256		512×512		1024×1024	
		Gray	Color	Gray	Color	Gray	Color
GPU	DnCnn	0.034	0.035	0.039	0.039	0.059	0.059
	FFDNet	0.033	0.032	0.033	0.032	0.034	0.032
	IRCnn	0.032	0.032	0.032	0.032	0.032	0.032
	ADNet	0.033	0.035	0.037	0.047	0.053	0.095
	AirNet	-	0.150	-	0.505	-	2.515
	DRANet	0.151	0.153	0.345	0.355	1.050	1.059
	DECDNet	0.053	0.055	0.081	0.084	0.182	0.188

The number of parameters for various denoising methods on grayscale and color images is listed in Table 7. It can be observed that our DECDNet outperforms DnCNN, FFDNet, IRCNN, and ADNet in terms of quantitative results despite having fewer parameters. Notably, even with a reduced parameter count, our DECDNet surpasses models like AirNet and DRANet which have more network parameters at specific noise levels.

Table 7 The numbers of parameters (in K) of different denoising models

Methods	Parameters	
	Gray	Color
DnCnn	666	668
FFDNet	485	852
IRCnn	186	188
ADNet	519	521
AirNet	-	8930
DRANet	1612	1617
DECDNet	1254	1263

4.6 Ablation study

To validate the efficacy of our proposed framework, we conducted an investigation on the integration of the noise estimation network and multi-scale feature extraction network, as well as the functionality of the dual attention mechanism by training three distinct networks for grayscale image denoising. Fig.8 illustrates the average PSNR (dB) achieved on the Set12 dataset at a noise level of 25 during various iterations. DECDANet solely comprises a dual denoising network and a dual attention mechanism, while DECDBNet incorporates only the noise estimation network, multi-scale feature extraction network, and dual denoising network. As depicted in Fig.8, it is evident that the complete DECDNet model outperforms both alternative models significantly, indicating that each module contributes to enhancing denoising performance. Table 8 presents ablation experiment results obtained from evaluating on BSD68 dataset across three different noise levels (σ=15, 25, 50). The outcomes demonstrate that when σ=15, PSNR value improves by 0.51 and 0.22 compared to DECDANet and DECDBNet respectively; when σ=25

improvements are observed by 0.41 and 0.14; when σ=50 improvements are noted by 0.19 and 0.06 respectively - effectively substantiating that integrating both noise estimation network with multi-scale feature extraction network along with employing dual attention mechanism enhances overall denoising performance.

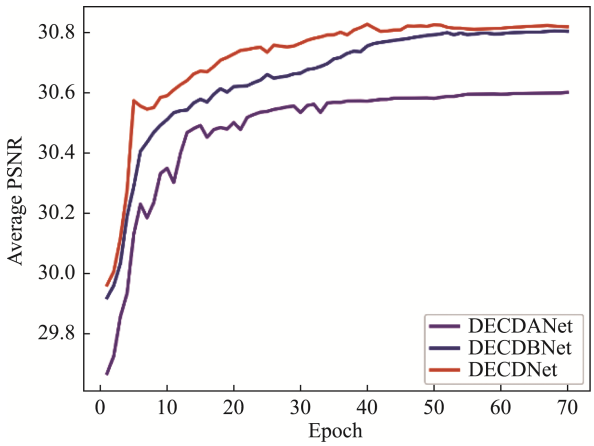

Fig.8 The averaged PSNR (dB) of various different network architectures on Set12 datasets at noise level of 25

Table 8 The averaged PSNR comparisons of different networks on the BSD68 datase

σ	15	25	50
DECDANet	31.27	28.98	26.36
DECDBNet	31.56	29.25	26.49
DECDNet	31.78	29.39	26.55

5 Conclusion

This paper presents the design and implementation of NECDNet, a novel denoising network model that effectively balances denoising performance and model complexity. The proposed model integrates a noise estimation module, a multi-scale feature extraction module, a dual convolutional neural network structure, and a dual attention mechanism to form an integrated denoising system with high synergy. The noise estimation network accurately estimates the noise level in input images, while the multi-scale feature extraction network enhances adaptability to diverse image content by capturing rich image details. The unique architecture of the dual convolutional neural network, combined with downsampling and skip connections, facilitates information flow and promotes feature reuse for improved denoising performance. Compared to existing models, DECDNet achieves highly competitive results across various denoising tasks, particularly under high noise levels. It demonstrates robustness in noisy environments while preserving visual quality in the denoised images, striking a perfect balance between noise removal and detail preservation requirements. Experimental results also highlight the need for further research on enhancing the model's denoising capability under low noise levels. Moreover, despite its excellent performance, training of our proposed model still requires original as well as noisy images; hence future research aims to explore self-supervised or unsupervised image denoising methods to address this limitation.

Author Contributions:

Mengnan Lü: As the first author, conceived the research idea and designed the overall framework. Conducted in-depth literature review and formulated research questions. Collected and analyzed the majority of the data. Wrote the main body of the manuscript and led the revision process.

Xianchun Zhou: Provided guidance on research direction and methodology. Reviewed and critiqued the research design and manuscript at different stages.

Zhiting Du: Assisted in data collection and organization. Performed preliminary data analysis and contributed to data interpretation.

Yuze Chen: Participated in discussions on research findings and provided valuable feedback for manuscript improvement.

Binxin Tang: Conducted supplementary experiments and validations. Helped in creating visualizations of data and enhancing the presentation of results.

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 25 June 2024; Accepted 11 July 2024; Published online 28 December 2024

References

[1] Chen H A, Lu X Y, Shan Y M, et al.(2023). Research on the method of imaging noise removal for high-speed camera[J]. *Chinese Journal of Scientific Instrument*,44(02):211-220.

[2] Buades A, Coll B, Morel J M.(2005). A non-local algorithm for image denoising[C]. *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). IEEE*,2:60-65.

[3] Cheng D Q, Chen J, Kou Q Q, et al.(2022).A Super-Resolution Reconstruction Method for Lightweight Mine Images by Fusing Hierarchical Features and Attention Mechanism[J]. *Journal of Instrumentation*, 43(08) :73-84.

[4] 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): 11695-11715.

[5] Liu Y, Qin Z Y, Anwar S, et al.(2021).Invertible denoising network: A light solution for real noise removal[C].*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 19-25, 2021, Online. IEEE Press*,2021: 13365-13374.

[6] Sharif S M A, Naqvi R A.(2021). BISWAS M.Beyond joint demosaicking and denoising: An image processing pipeline for a pixel-bin image sensor[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 19-25, 2021, Online. IEEE Press*,2021: 233-242.

[7] Zamir S W, Aroraa, Khan S, et al.(2021).Multi-stage progressive image restoration[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 19-25, 2021, Online. IEEE Press*, 2021:14821-14831.

[8] Zhang K, Li Y W, Zuo W M, et al.(2022). Plug-and-play image restoration with deep denoiser prior[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10): 6360-6376.

[9] Zhang K, Zuo W, Chen Y, Meng D, Zhang L.(2017).Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising[J]. *IEEE Trans on Image Processing*,26(7): 3142-3155.

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

[11] Guo S, Yan Z F, Zhang K, et al. (2019).Toward convolutional blind denoising of real photographs[C].Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Jun. 16-20, 2019, Long Beach, USA. *IEEE Press*, 2019:1712-1722.

[12] Lakshmi Parvathi M, Hirthik M, Sampath Kumar, Vani Damodaran.(2023).Despeckling of optical coherence tomography images using encoder-decoder network with skip connections[C]. *AIP Conf. Proc.* 25 April 2023; 2603(1): 020020.

[13] Zhao W Q, Liu L, Hu J W, et al.(2023). Detection of transformer oil leakage based on deep separable atrous convolution pyramid[J]. *CAAI Transactions on Intelligent Systems*, 18(5): 966-974.

[14] Zhang K, Zuo W, Gu S, et al.(2017). Learning Deep CNN Denoiser Prior for Image Restoration[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2017:2808-2817.

[15] Tian C, Xu Y, Zuo W. (2020). Image denoising using deep CNN with batch renormalization[J]. *Neural networks: the official journal of the International Neural Network Society*, 121: 461-473.

[16] Gurrola-Ramos J, Dalmau O, Alarcon T E.(2021). A Residual Dense U-net Neural Network for Image Denoising[J]. *IEEE Access*, 9: 31742-31754.

[17] Wang H, Cao P, Wang J, et al.(2022). UCTransNet: Rethinking the Skip Connections in U-Net from a Channel-Wise Perspective with Transformer[C]. *Proceedings of the AAAI Conference on Artificial Intelligence*,36(3): 2441-2449.

[18] OYEDOTUN O K, EL RAHMAN SHABAYEK A, AOUADA D, et al.(2017). Training very deep networks via residual learning with stochastic input shortcut connections[C]. *Proceedings of International Conference on Neural Information Processing*. Guangzhou, China: Springer Verlag, 2017: 23-33.

[19] OYEDOTUN O K, ISMAEIL K A, AOUADA D. (2023). Why is everyone training very deep neural network with skip connections? [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9): 5961-5975.

[20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. (2021). An Image is worth 16×16 words: Transformers for image recognition at scale[C]. *Proceedings of International Conference on Learning Representations. Washington D. C., USA: IEEE Press*, 2021:5278-5284.

[21] Chen D Q, Chen J, Kou Q Q, et al. (2022). Lightweight super-resolution reconstruction method based on hierarchical features fusion and attention mechanism for mine image[J]. *Chinese Journal of Scientific Instrument*, 8: 73-84.

[22] Yan S, Long Y, Fu R H, et al. (2022). A Method for Denoising Seismic Signals with a CNN Based on an Attention Mechanism[J], *IEEE Transactions on Geoscience and Remote Sensing*, 60(1): 1-15.

[23] ZHOU Y Q, SONG Y Q, WU H, et al. (2023). Highly Accurate Gesture Recognition Based on DenseNet and Convolutional Attention Module[J]. *Journal of Electronics and Information*, 6(1): 1-10.

[24] S.Woo,J.Park,J.Lee,I.S.Kweon.(2018).CBAM:convolutional blockattentionmodule[J].*EuropeanConferenceonComputerVision,*11211,3–19.

[25] Shen S L, Zhang N, Zhou A, et al. (2022). Enhancement of neural networks with an alternative activation function tanhLU[J]. *Expert Systems with Applications*,199(12): 117181.

[26] Vishwakarma,Amit.(2023).Denoising and Inpainting of Sonar Images Using Convolutional Sparse Representation[J]. *IEEE Transactions on Instrumentation and Measurement*, 72 : 1-9.

[27] Zhang S, Liu C, Zhang Y, Liu S, Wang X. (2023). Multi-Scale Feature Learning Convolutional Neural Network for Image Denoising[J]. *Sensors*, 23(18): 7713.

[28] Sahu A, Rana KPS, Kumar V. (2023). An application of deep dual convolutional neural network for enhanced medical image denoising[J]. *Med Biol Eng Comput*, 61(5): 991-1004.

[29] Yu J, Yang X, Gao F, et al. (2017). Deep Multimodal Distance Metric Learning Using Click Constraints for Image Ranking[J]. *IEEE Transactions on Cybernetics*, 47 (12): 4014-4024.

[30] Wang P, Chen P, Yuan Y, et al. (2018). Understanding Convolution for Semantic Segmentation[C]. *2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE*, 2018:1451-1460.

[31] Jiang, Y, Zhang, C, Liu, J. (2023).CS-PCN: Context-Space Progressive Collaborative Network for Image Denoising[C]. *IEEE International Conference on Multimedia and Expo (ICME)*, 2023:2759-2764.

[32] Liu T, Luo R, Xu L, Feng D, Cao L, Liu S, Guo J. (2022). Spatial Channel Attention for Deep Convolutional Neural Networks[J]. *Mathematics*, 10(10): 1750.

[33] Gong M, Jiang F, Qin A K, et al. (2021). A spectral and spatial attention network for change detection in hyperspectral images[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 60: 1-14.

[34] Eirikur Agustsson, Radu Timofte.(2017).NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study[C]. *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017: 1122-1131.

[35] Roth S, Black M J. (2005). Fields of Experts: A Framework for Learning Image Priors[C]. *IEEE Computer Society Conference on Computer Vision & Pattern Recognition. IEEE,* 2005, 2: 860-867.

[36] L Zhang, X Wu, A Buades,X Li.(2011).Color demosaicking by local directional interpolation and nonlocal adaptive thresholding[J].*Journal of Electronic Imaging*, 20 (2): 1-17.

[37] C. Tian, Y. Xu, W. Zuo, B. Du, C. Lin, D. Zhang. (2021). Designing and training of a dual CNN for image denoising, Knowl[J]. *Based Syst*, 226: 106949.

[38] Wu W, Liu S, Xia Y, Zhang Y. (2024). Dual residual attention network for image denoising[J], *Pattern Recognition*, 149(8): 110291.

[39] B. Li, X. Liu, P. Hu, Z. Wu, J. Lv, X. Peng.(2022).All-in-one image restoration for unknown corruption[C]. *IEEE Conference on Computer Vision and Pattern Recognition*, 2022: 17431-17441.