Knowledge and Information Systems (2024) 66:581–611
https://doi.org/10.1007/s10115-023-01965-9

REGULAR PAPER

A novel image denoising algorithm combining attention mechanism and residual UNet network

Shifei Ding[1,2] · Qidong Wang[1] · Lili Guo[1,2] · Jian Zhang[1] · Ling Ding[3]

Received: 13 August 2022 / Revised: 29 April 2023 / Accepted: 8 August 2023 /
Published online: 8 September 2023

Abstract

Images are easily polluted by noise in the process of acquisition and transmission, which will affect people's understanding and utilization of knowledge and information in images. Therefore, image denoising, as a classic problem, has received extensive attention from researchers. At present, many image denoising methods based on deep learning have been proposed and achieved good performance. However, most existing methods are insufficient in acquiring and utilizing crucial information in the image when removing noise under complex image denoising tasks such as blind denoising and real-world denoising, resulting in the loss of fine details in the reconstructed image. To overcome this shortcoming, in this paper, we propose a novel image denoising algorithm combining attention mechanism and residual UNet network, named Att-ResUNet. Specifically, we propose a novel UNet-based image denoising framework, which employs residual enhancement blocks and skip connections to form global–local residuals, which can fuse multi-scale global context and local features to more thoroughly capture and remove hidden noise in the image. A channel attention mechanism is introduced, which can better focus on the crucial information in the image and improve the denoising performance. In addition, we use adaptive average pooling for down-sampling, which can preserve more image structure information, reduce the loss of edge details, and adopt a residual learning strategy to enhance the learning and expressive capabilities of the denoising model. Extensive experiments on several publicly available standard datasets demonstrate the superiority of our method over 15 state-of-the-art methods and achieve excellent denoising performance. Compared with mainstream methods, our method outperforms current state-of-the-art methods by up to 0.76 dB and 1.10 dB on PSNR evaluation metrics on BSD68 and Set12 datasets, respectively. Notably, our method achieves an

✉ Lili Guo
liliguo@cumt.edu.cn

✉ Ling Ding
dltjdx2022@tju.edu.cn

1 School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

2 Mine Digitization Engineering Research Center of Ministry of Education of the People's Republic of China, Xuzhou 221116, China

3 College of Intelligence and Computing, Tianjin University, Tianjin 300350, China

average PSNR value of 37.88 dB on the CC dataset in real-world denoising experiments, a significant improvement of 2.14 dB over the most advanced methods.

Keywords Image denoising · UNet · Residual learning · Attention mechanism · Complex denoising tasks

1 Introduction

The image holds a wealth of information and serves as a crucial carrier of information, but it can easily be impure by noise within the process of transmission and compression. Noise can have an adverse influence on following tasks such as recognition and segmentation, and especially in some specific application scenarios, a small amount of noise may cause serious errors or deviations. For example, in the aerospace field, noise pollution will affect the quality of images transmitted from the space station back to the Earth. In the medical field, noise may interfere with physicians. Consequently, reducing noise to acquire high-quality images is of paramount importance, and it has consistently been a fundamental challenge in the field of image processing [1–6].

We can understand the image denoising task in the following way. For the input image $p(t)$ containing noise, its additive noise can be expressed mathematically as follows:

$$p(t) = c(t) + n(t),\ t \in \Omega$$

where t denotes the pixel; $c(t)$ represents the clean image; $n(t)$ signifies the noise term, which symbolizes the impact of noise on the image; and $p(t)$ is the whole image. The task of image denoising can be interpreted as recovering the clean image $c(t)$ from the noisy image $p(t)$.

Separating the noise from the observed image is the key to image denoising. Meanwhile, preserving as much of the original image information as possible is expected. To achieve the goal of image denoising, researchers have proposed many denoising algorithms in recent decades [7–9]. The traditional denoising algorithm mainly adopts the idea of filtering and sparse representation. Among them, the filter methods mainly include mean filtering, median filtering, and Wiener filtering. In the filtering algorithm, Buades et al. took the non-local self-similarity of the image into account and proposed a non-local mean (non-local mean, NLM) [10] filtering method. The NLM algorithm is simple in principle and can achieve a good noise reduction effect, but its computational complexity is high. Dabov et al. suggested the BM3D (block-matching and 3D filtering, BM3D) [11] algorithm. The BM3D algorithm has an excellent performance in denoising effect and processing speed, and BM3D is still widely used for image denoising today. Currently, many methods are improved based on BM3D, but these methods are prone to loss of details and blurred images in the face of complex noise. Gu et al. [12] proposed a Weighted Kernel Norm Minimization (WNNM) model for image denoising. This model adaptively assigns weights to singular values, enhancing image structure representation and denoising performance. However, the computational complexity of solving the WNNM problem and employing the Weighted Kernel Norm Proximal (WNNP) operator can be significant.

The method of sparse representation is derived from wavelet theory. The main idea of this type of model is: The image containing noise can be decomposed to obtain a limited number of signal atoms through sparse transformation, the signal atoms are composed of a dictionary, and the expression of clean signal and noise in the dictionary differently, only clean signals with sparse components can be effectively expressed, and the separation of

noise can be achieved in this way. The representative method is the K-SVD (K-singular value decomposition, K-SVD) [13, 14] algorithm proposed by Aharon et al. The sparse representation method can fit the data well through the dictionary. However, this type of method generally has the problem of a large amount of calculation and is easy to ignore the correlation.

As artificial intelligence and big data technology continue to advance, noise reduction research based on deep learning has emerged as a popular and rapidly developing field. With the support of sufficient data, deep learning approaches can learn mappings from noisy images to clean images, which can better connect image context information. Jian [15] et al. adopted a supervised learning approach for image denoising, which encouraged people to use the deep learning method for image denoising. Burger et al. presented multilayer perceptron (MLP) [16]. MLP has a strong approximation ability and can fit very complex functions. But it also suffers from limited receptive fields, high computational cost, and difficulty in handling spatial and contextual information. Schmidt et al. designed a framework named the cascade of shrinkage field (CSF) [17] model through convolution operation and DFT transform, which improved the computational efficiency and image restoration effect. However, its performance deteriorates when dealing with intricate or highly intricate noise. Mao [18] et al. proposed a deeper network model, using skip connections, and proposed a deep residual encoding–decoding network. This model's advantage is the use of symmetric skip connections, which leads to easier training convergence and higher-quality image restoration. However, this model necessitates training a significant number of parameters, which may cause overfitting. Inspired by the thermal reaction principle in physics, Chen et al. established the TNRD [19] model. TNRD provides a flexible framework for various image restoration tasks with fast and effective results, but it performs poorly in the face of complex image backgrounds and high noise. Intending to solve the gradient dispersion problem caused by deepening the network structure. Zhang et al. proposed a CNN-based Gaussian denoising approach named DnCNN (denoising convolutional neural networks, DnCNN) [20]. Unlike previous methods, this method is not directly output the denoised image, the residual learning strategy is employed and plays a significant role in training the network to predict noise, and the clean image can be attained via doing subtractions between the noisy image and the noise. The proposal of DnCNN provides a new direction for deep learning noise reduction. They subsequently proposed IRCNN [21] and FFDNet [22], in which IRCNN is a deep CNN denoiser before image restoration, which integrates CNN into a model-based HQS optimization method, and FFDNet optimizes the adaptability to noise and the amount of calculation based on DnCNN. These works inspired us to pay attention to noise adaptability when designing denoising methods. At the same time, we also found that these methods still have the problem of easy loss of fine details in complex denoising tasks and image backgrounds.

Recently, more DNN-based techniques have been proposed by researchers to address the challenge of image denoising. Tian et al. [23] suggested an enhanced CNN model named ECNDNet, using residual learning and batch normalization (BN) techniques to solve the challenge of training difficulties and improve the convergence ability of the network. Subsequently, to strengthen the contribution of the shallow layer to the deep network, they proposed an attention-guided CNN for image denoising (ADNet) [24]. These methods can significantly improve the performance of image denoising, but there are still problems that cannot fully extract feature information. Quan et al. [25] used complex-valued convolutional neural networks for image denoising for the first time and achieved good denoising results. To better extract and strengthen features, Tian et al. proposed DudeNet [26], which is a dual-path denoising network that uses sparse mechanisms and complementary methods to

improve denoising performance. Zhang et al. [27] introduced deformable convolution into a denoising network to expand the receptive field. Tian et al. [28] proposed a multi-stage image denoising method combined with wavelet transform (MWDCNN), which can reduce the training cost while achieving better denoising performance. These advanced methods provide us with valuable references and inspiration. Through in-depth analysis and research on these works [20, 25, 27, 28], we found that these methods can effectively extract features and achieve excellent denoising results in most general denoising tasks. For complex denoising tasks (blind denoising and real-world noise removal), there are still shortcomings of not being able to fully obtain and utilize key information, which will lead to the degradation of fine details in the denoised image.

To overcome the above shortcomings, in this paper, we propose a novel UNet-based image denoising framework. Specifically, we propose a residual enhancement block to increase the receptive field, reduce the loss of feature information through residual connections, and use the PReLU activation function to help preserve detailed information and prevent overfitting in complex denoising tasks. Then, the residual enhancement block is combined with skip connections to form a global–local residual, which can learn and capture noise more thoroughly in complex denoising tasks. Furthermore, we use a channel attention mechanism to assist in image denoising. The channel attention mechanism effectively emphasizes crucial information to deal with complex noisy images. To better preserve the structural information and edge details of the image, we use adaptive average pooling for down-sampling. Inspired by DnCNN and ADNet, we also use the strategy of residual learning to learn the distribution of noise, and the final clean image can be acquired by subtracting the noisy image and the noise. Extensive experiments show that our Att-ResUNet outperforms the state-of-the-art denoising methods in quantitative and qualitative analysis and has obvious advantages in complex denoising tasks.

To summarize, the main contributions of our paper are listed as follows:

- We propose a novel image denoising algorithm combining attention mechanism and residual UNet network. We design a residual enhancement block and form a global–local residual with skip connections, which can combine multi-scale global context information and local feature information to capture and remove complex noise more thoroughly.
- The channel attention mechanism is introduced to better focus on crucial details and reduces the loss of fine details in the image during denoising.
- An adaptive average pooling layer is adopted to adjust the image resolution in the down-sampling stage, which can better preserve the structure and edge details of the image during the denoising process.
- Our extensive comparison experiments with a total of 15 state-of-the-art methods and ablation studies show that our method achieves excellent results on multiple publicly available benchmarks. At the same time, significant improvements are achieved in both blind denoising and real-world noise removal tasks.

The remainder of this paper is structured as follows. Section 2 provides a brief overview of related work on the UNet network, the residual structure model, and the channel attention mechanism. Section 3 provides a comprehensive description of the method proposed in this paper. Results from comparative experiments, ablation studies, and qualitative and quantitative analyses are discussed in Sect. 4. At last, Sect. 5 provides a summary of this paper and outlines potential future research directions.

2 Related work

2.1 UNet network

The UNet [29] network was first proposed as a novel image segmentation model, named because its shape is very similar to the U in the English letter, and then, it became a powerful tool for most medical image segmentation tasks [30–33]. Its superior and special mechanism also encourages many scientific experts to think about and improve the U-based image segmentation network. Nowadays, U-shaped structures have gained increasing attention among researchers and are being widely applied in image generation applications.

The structure of the UNet network model is symmetrical. The overall network structure is a classic encoding and decoding structure, which mainly consists of two stages: the encoding stage and the decoding stage. The left half is used for encoding and compression, and the right half is used for decoding and expansion. In the encoding stage, the encoder performs feature extraction on the input image from different angles through multiple feature channels, and the subsequent network layer further processes and extracts the shallow texture features extracted in the previous stage to obtain high-level semantic features to get a more accurate description and description of things. After multiple encodings, finally, the appropriate feature map will be obtained. The biggest significance of the encoding operation is to extract useful information through learning, remove a lot of interference and useless information, and achieve the purpose of reducing the data dimension. The decoding stage is to up-sample and restore the encoded information and finally obtain the target image.

The encoding–decoding structure and skip connections of the UNet network help to extract and utilize image features more fully, which inspired us to design an image denoising method based on the UNet network.

2.2 Residual structure model

The residual structure model [34] was first proposed by researchers to effectively tackle the challenges of network gradient disappearance and structural degradation as a result of the increase in network depth. The residual structure model is easy to adjust and improve, and appropriately increasing the network depth in the model structure can improve the accuracy [35–38]. The residual block inside the model uses the idea of skip connection, and the output of the latter layer can be combined with the input of the earlier stage, which can significantly decrease the loss of feature information. The advantage of the residual structure motivates us to design the residual enhancement block to combine the UNet network architecture to better adapt to the denoising task.

2.3 Channel attention mechanism

The attention mechanism was originally applied in the task of machine translation, and now in the realm of deep learning, it has proven to be a very effective strategy. The main principles of the attention mechanism are the human visual mechanism and cognitive science. When humans observe and analyze things, they will choose to focus only on information that is useful for decision-making and judgment, while ignoring other irrelevant information. The attention mechanism can reasonably use and distribute restricted information processing resources to maximize the accuracy of the model.

SENet (squeeze-and-excitation networks) [39] is a very classic and effective channel attention mechanism that is found in a variety of computer vision works [40–43]. The SE module can find out the difference between feature channels. It can adaptively alter the weight of each feature channel based on the current task, which can encourage the model to pay closer attention to the image feature information.

Inspired by these facts, we incorporate the channel attention mechanism of SENet into the proposed denoising framework to better focus and preserve important details.

3 Image denoising algorithm combining attention mechanism and residual UNet

This section provides a detailed introduction to the proposed denoising model Att-ResUNet, which includes global–local residual connections composed of residual enhancement blocks and skip connections, channel attention blocks, down-sampling implemented by adaptive average pooling, batch normalization, and residual learning. The proposed residual enhancement block aims to decrease the loss of features, which helps to improve the denoising effect. Moreover, it is worth emphasizing that the residual enhancement block can be combined with the skip connection to form a global–local residual connection to combine multi-scale global context information and local feature information, which can more thoroughly and effectively eradicate the noise concealed in the image. The channel attention block can better focus on the crucial details in the image by introducing the SENet channel attention mechanism while reducing the loss of fine details. Adaptive average pooling is adopted in down-sampling to preserve more structure and edge details of the image. Batch normalization is adopted to speed up training and prevent overfitting, and a residual learning strategy is adopted to improve robustness. Further, we highlight important modules or technologies in several subsections.

3.1 Network architecture

The proposed Att-ResUNet, as shown in Fig. 1, consists of three main parts, a novel UNet-based network, five residual enhancement blocks, and five attention blocks. Att-ResUNet

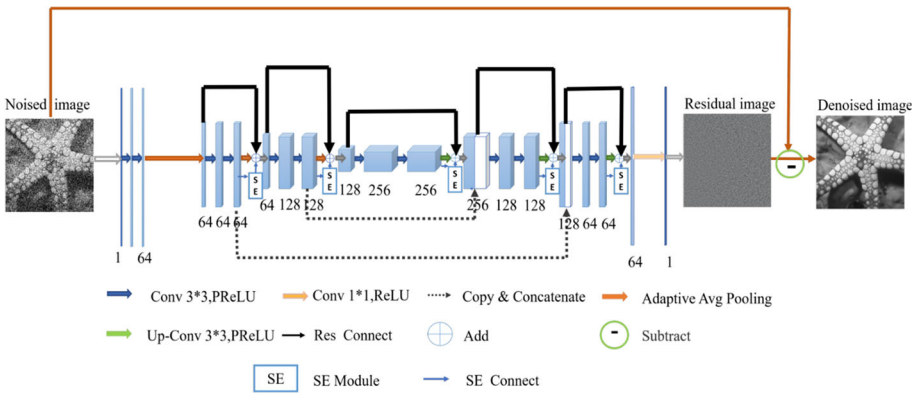

Fig. 1 The network architecture of Att-ResUNet

extracts feature information from input noisy images, predicts the noise mixed in the image, and then, the clean image is attained by subtracting the noise image input and the predicted noise, which is the final output of the model. Specifically, Att-ResUNet outputs a predicted clean image through the combination of multiple convolution modules, residual enhancement block, attention block, and adaptive average pooling layers for down-sampling stages, and multiple deconvolution layers, residual enhancement block, attention block, and convolution layers for up-sampling stages, which can preserve the features extracted during the network down-sampling process to the greatest extent information to ensure that the reconstructed image has better quality. In Att-ResUNet, there are two down-sampling stages and up-sampling stages. The deconvolution output of each up-sampling is combined with the output of the second activation function of the same layer down-sampling, preserving important feature information.

The training procedure of the UNet network model is summarized as two stages: the forward propagation prediction stage and the backpropagation parameter optimization stage. The testing process can be regarded as forward propagation to reconstruct the target image under the optimized model.

The core of the forward propagation stage is up-sampling and down-sampling. The sampling includes five basic units: convolution calculation, batch normalization (batch normalization), PReLU activation function, adaptive average pooling, and deconvolution.

The convolution calculation is expressed as Eq. (1):

$$C_k = W_k * C_{k-1} + b_k \tag{1}$$

Among them, C_k represents the output of the layer k; W_k denotes the convolution kernel for layer k; C_{k-1} signifies the input of the layer k; and b_k represents the bias of the layer k.

Batch normalization is expressed as Eqs. (2), (3), (4), and (5):

$$\mu_{\mathrm{B}} = \frac{1}{m}\sum_{i=1}^{m} C_k \tag{2}$$

$$\sigma_{\mathrm{B}} = \sqrt{\frac{\sum_{i=1}^{m}(C_k - \mu_{\mathrm{B}})^2}{m-1}} \tag{3}$$

$$C_k = \frac{(C_k - \mu_{\mathrm{B}})}{\sigma_{\mathrm{B}}} \tag{4}$$

$$\hat{C} = r * C_k + \beta \tag{5}$$

Among them, μ_B represents the mean of the batch data; σ_B is the variance of the batch data; m denotes the size of the selected batch data; $\hat{C}$ signifies the output after batch β normalization; and r and β are the parameters. BN [20, 24] can alleviate the vanishing and exploding gradient problem, thereby improving training stability and denoising ability.

In the image denoising algorithm proposed in this paper, the applied activation function is PReLU, and the formula is expressed as Eq. (6):

$$\mathrm{PReLU}(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases} \tag{6}$$

Among them, x represents the normalized output. The gradient of the ReLU activation function is 0 in the negative part, which may cause the problem of gradient disappearance. In image denoising tasks, this may affect the learning capability of the network, especially when dealing with noisy images. However, the gradient of the PReLU activation function in the negative part is not 0 (depending on the parameter α), which helps to maintain better

gradient propagation throughout the training process, thereby improving image denoising performance.

Adaptive average pooling in down-sampling is expressed as Eq. (7):

$$C_k = \text{Adaptive} - \text{avgpool}(C_k) \tag{7}$$

Among them, Adaptive − avgpool represents adaptive average pooling. The details of image resolution adjustment in the down-sampling stages using adaptive average pooling are described in Sect. 3.4.

The deconvolution calculation principle used in the up-sampling stage is expressed as Eq. (8):

$$C_{kT} = W_{kT} * C_{kT-1} + b_{kT} \tag{8}$$

Among them, C_{kT} symbolizes the output of the layer k; W_{kT} symbolizes the convolution kernel of the layer k; C_{kT-1} symbolizes the input of the layer k; and b_{kT} symbolizes the bias of the layer k.

In up-sampling, the deconvolution output C_{u} of each layer will be spliced and merged with the down-sampling convolution output C_{d} of the same layer on the position of channels to obtain the input C_f of the up-sampling convolution calculation, as shown in Eq. (9):

$$C_f = \text{concat}(C_{\text{u}}, C_{\text{d}}) \tag{9}$$

The backpropagation stage uses the mean squared error as the loss function to express the loss between the predicted value obtained by forward propagation and the true value and then optimizes by updating the parameters of the network model. The loss is calculated as Eq. (10):

$$\text{MSE} = \frac{1}{m}\sum_{i=1}^{m}\left(y_{\text{ref}}^{(i)} - y_{\text{pred}}^{(i)}\right)^2 \tag{10}$$

Att-ResUNet adopts the strategy of residual learning [20, 24]. The network structure will first predict the noise residual and then subtract this residual to obtain the denoised image from the input noisy image. The advantage of this method is that it can better capture the details of the noise, making the denoising effect better. Therefore, what is used to calculate the loss and participate in the training optimization is the clean image after denoising. In Eq. (10), $y_{\text{pred}}^{(i)}$ symbolizes the denoised image by Att-ResUNet; $y_{\text{ref}}^{(i)}$ symbolizes the real clean image; and m symbolizes the size of the data volume.

The optimization method adopts the Adam algorithm, which can solve problems such as the disappearance of learning rate and slow convergence, and better learning optimization. The principle of the Adam algorithm is expressed as Eq. (11):

$$\begin{cases} v_1 = \beta_1 v_1 + (1-\beta_1)\partial J\left(\theta_j\right) \\ s_1 = \beta_2 s_1 + (1-\beta_2)\partial J\left(\theta_j\right)^2 \\ \theta_{j+1} = \theta_j - \text{lr}\frac{v_1}{\sqrt{S_1+\varepsilon_0}} \end{cases} \tag{11}$$

Among them, v_1 is the first-order cumulative gradient; s_1 is the second-order cumulative squared gradient; $J\left(\theta_j\right)$ represents the loss function; β_1 and β_2 are the decay rate; and lr is the learning rate.

3.2 Residual enhancement block

To decrease the loss of feature information in the denoising process, and make the network more fully learn and utilize feature information to improve denoising performance, while reducing the risk of gradient disappearance, we add residual connections on top of the original convolutional blocks, as indicated by the bold black arrows in Fig. 1. We incorporate residual connections at each convolution stage for down-sampling and up-sampling. A total of five residual enhancement blocks are constructed, which can not only eliminate the risk of gradient explosion in the backpropagation process, but also the residual module can remember the previous feature information, effectively decrease the loss of feature in each stage, and enhance the denoising of the model performance. The specific structure of the residual enhancement block is shown in Fig. 2. The residual enhancement block consists of a residual connection and two convolutional stages, each convolutional stage consists of convolution, batch normalization, and PReLU activation function, where PReLU is more suitable for image denoising tasks because it can learn parameters while allowing the activation function to adapt to different features more flexibly, thereby improving the effectiveness of the model and helping to restore the original image more accurately.

The primary notion of residual connection is identity mapping. The residual structure uses shortcuts to implement skip connections, that is, the input results of the previous stage are directly transferred to the output position of the convolutional layer and added. When the network finds that the parameters of some layers are difficult to learn or interfere with the results during the training process, the parameters will be assigned a value of 0, and an identity map is formed at this time. This method can dynamically and flexibly change the complexity of the model, reduce the loss of feature information during the denoising process, and improve the denoising performance at the same time. This process can be expressed as Eq. (12):

$$H(x) = F(x) + x \Rightarrow F(x) = H(x) - x \tag{12}$$

Among them, x represents the input of the residual module, $F(x)$ is the residual, and $H(x)$ represents the output of the residual module, when $F(x) = 0$, it constitutes the identity map $H(x) = x$.

It needs to be added that the residual enhancement block and the skip connection of UNet can form a global–local residual connection, which can combine multi-scale global context features and local features, and can capture the details of different levels in the image. Global features provide information on the overall structure of an image, while local features focus on details and textures. This combination helps the network to more fully and thoroughly

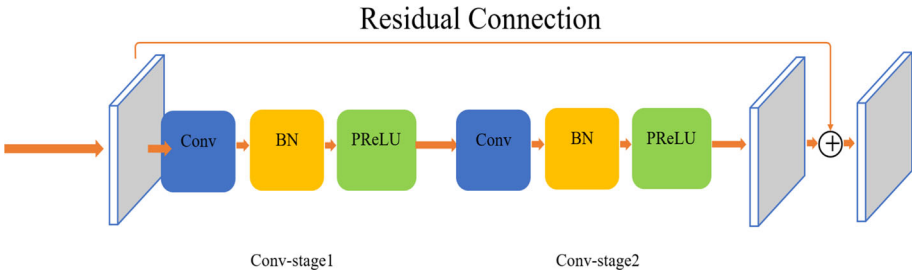

Fig. 2 Illustration of the detailed structure of the residual enhancement block

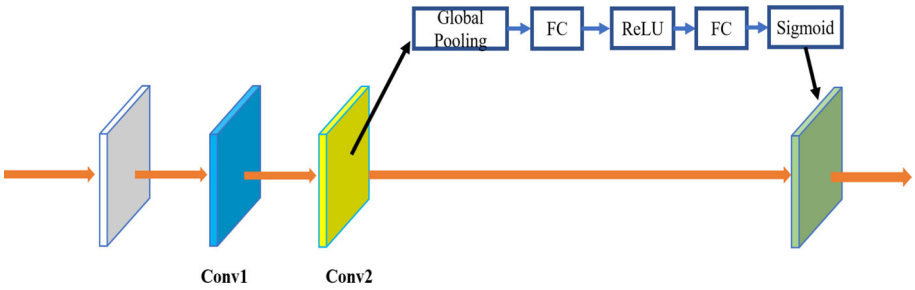

Fig. 3 Illustration of the detailed structure of the attention block

mine the complex noise hidden in the image while retaining the important information of the image and improving the denoising effect.

3.3 Attention block

Ordinary deep neural networks ignore the correlation between channels in the process of denoising, and it is easy to lose fine detail information [20, 21], especially in complex denoising tasks such as blind denoising and real image denoising. To effectively emphasize and preserve the detailed information of the image during the denoising process, we add an attention block based on the residual enhancement block and apply the attention mechanism in both the encoding and decoding stages. The correlation between channels is learned by introducing the SENet channel attention mechanism. By processing the feature map obtained by convolution, a vector consistent with the number of channels is obtained as the evaluation weight of each channel and added to the corresponding channel, respectively, which can strengthen the acquisition of key features, thereby enhancing the network's denoising performance. The introduced attention blocks in our study are shown in Fig. 3. Moreover, in the denoising process, the working process of the attention block can be expressed by Eq. (13):

$$F_{\text{att}} = C_{\text{att}}(O_{\text{cout}}) \tag{13}$$

where O_{cout} represents the output of the convolution module, C_{att} is the attention block, and F_{att} is the output of the attention block, which can pay more attention to crucial information based on the convolutional feature map.

3.4 Adaptive average pooling layer

Another novelty of Att-ResUNet compared to traditional UNet-based networks is the use of an adaptive average pooling layer to achieve down-sampled image resolution instead of max pooling in the down-sampling stage. Adaptive average pooling is less sensitive to outliers and noise than max pooling. This means that when dealing with images affected by noise, average pooling can effectively retain the basic features and edge details of the image. Moreover, adaptive average pooling is relatively simple and efficient computationally. This makes it advantageous when dealing with complex noise tasks in real time.

4 Experiment and analysis

To validate the efficacy of our proposed method, in this section, we design and implement various types of denoising experiments, including non-blind denoising experiments on grayscale images and color images, and experiments for complex denoising tasks such as blind denoising and real-world noise removal. On six test datasets, Att-ResUNet is compared with other advanced algorithms under the same conditions, and the performance of the model is analyzed through subjective feelings and objective evaluation indicators. We also design ablation experiments targeting individual and composite components to verify the contributions of various parts of the network.

4.1 Introduction to datasets

4.1.1 Training dataset

There are three training datasets used in the experiments in this paper. Among them, the public image dataset BSD400 [44] provided by Berkeley University, Waterloo Exploration Database [45], and super-resolution common dataset DIV2K [46] are used for grayscale image experiments. We used all 400 images from the BSD400 dataset, the first 100 images from the Waterloo Exploration Database, and the first 100 images from the 4 $\times$ magnified test set images in DIV2K to form a synthetic dataset of 600 images to train gray Gaussian synthetic noise denoising model. To be clear, in order to train Att-ResUNet for Gaussian denoising in application scenarios with known noise levels, we take three different noise levels into consideration, namely, $\sigma = 15$, 25, and 50. In order to ensure sufficient samples for model training, we perform data augmentation. First, we apply a bicubic interpolation algorithm with reduction factors of 0.7, 0.8, 0.9, and 1 to increase the size of the training dataset and then crop the image patch, and the cropped image patch is set to 40 $\times$ 40 in size, we crop in total 128 $\times$ 2308 image patches to train the model. We name the Att-ResUNet model for Gaussian denoising in application scenarios with known specific noise levels Att-ResUNet-S. To train a single Att-ResUNet model for blind Gaussian denoising, we limit the noise level range to $\sigma \in [0, 55]$, in order to obtain more feature information and enhance the efficiency of training the denoising model, we set the patch size to 50 $\times$ 50 and crop 128 $\times$ 4110 image patches for training the model. The Att-ResUNet model for the blind Gaussian denoising application is referred to as Att-ResUNet-B.

In addition to grayscale image denoising, we also design experiments on color images, and the dataset used for color image denoising experiments is Waterloo Exploration Database. We employ the color version of the Waterloo Exploration Database dataset for training, using 1000 color images as training images. For the case of known noise levels, we consider five noise levels, namely, $\sigma = 15$, 25, 35, 50, and 75. In order to ensure sufficient samples for model training, we perform data augmentation. First, we apply a bicubic interpolation algorithm with reduction factors of 0.7, 0.8, 0.9, and 1 to increase the size of the training dataset and then crop the image patch, and the cropped image patch is set to 50 $\times$ 50 in size, we crop in total 128 $\times$ 2691 image patches for training the model. In addition, we perform blind denoising experiments on color images, the noise level is set in the range [0, 55], and 128 $\times$ 5382 blocks of size 50 $\times$ 50 are cropped for training.

For experiments on real-world noise removal, we use the first 1000 images of the Waterloo Discovery Database dataset and 100 JPEG-compressed 512 $\times$ 512 images from the Nam [47]

dataset to form a training dataset of realistic noisy images. Here, we also uniformly cut all images into 50*50 image blocks for easy processing.

In addition, with the aim of further increasing the variety of data, each training image block above will randomly select one of the following eight transformation methods to perform transformation: keep the original image, rotate 90°, rotate 180°, rotate 270°, the original image is flipped horizontally, the original image is flipped horizontally by 90°, the original image is flipped horizontally by 180°, and the original image is flipped horizontally by 270°.

4.1.2 Test dataset

A total of six test sets are used in our experiments to evaluate the denoising effect of Att-ResUNet, namely, BSD68 [48], Set12 [49], CBSD68 [48], Kodak24 [50], McMaster [51], and CC [52], where BSD68 and Set12 are used for grayscale image testing, CBSD68, Kodak24, and McMaster for color image testing, and CC is used for real-world noise removal testing. It should be added that all test datasets are publicly available test benchmarks, and the test set does not participate in the model training process, that is, the test set and the training set do not overlap. Part of the test dataset is shown in Fig. 4a which shows some samples of BSD68, and Set12 is a classic image quality test dataset. As shown in Fig. 4b, it contains 12 kinds of scenes such as photographers, houses, and starfish. CBSD68 is the color form of BSD68, and some samples are shown in Fig. 4c. Kodak24 contains 24 color image data of size 500 $\times$ 500, 12 of which are displayed in Fig. 4d. McMaster is also a commonly used dataset for image processing, as shown in Fig. 4e.

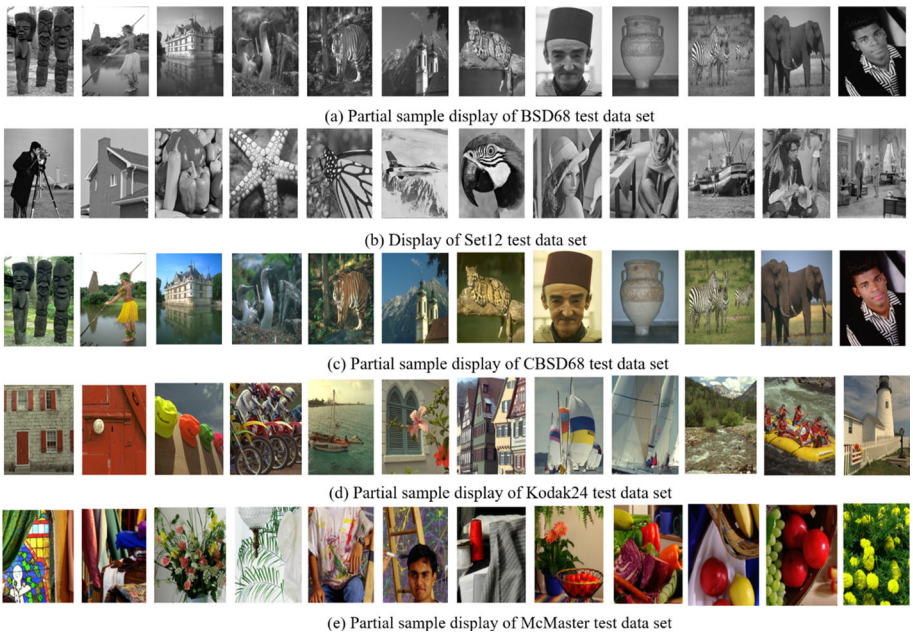

Fig. 4 Illustration showing part of the test datasets

4.2 Evaluation metrics

For the purpose of evaluating the effect of image denoising, we evaluate the performance of the UNet-based image denoising algorithm suggested in this study through two criteria:

(1) Peak signal-to-noise ratio (PSNR) [53]: a measurable standard for assessing the quality of image generation, the unit is dB, the higher of PSNR, the less visual distortion there will be, the calculation formula is expressed as (14).

(2) Structural similarity (SSIM) [54]: a metric of similarity between two images, based on the comparison of three angles between labels and generated results: brightness, contrast, and structure. The image creation quality improves as the SSIM value rises, and the calculation formula is expressed as Eq. (15).

$$\mathrm{PSNR} = 10\log_{10}\frac{\max^2(L(x),\, G(x))}{\frac{1}{M}\sum_{i=1}^{M}\|L(x) - G(x)\|_2^2} \tag{14}$$

$$\mathrm{SSIM} = \frac{\left(2\mu_{L(x)}\mu_{G(x)} + c_1\right)\left(2\sigma_{L(x)G(x)} + c_2\right)}{\left(\mu_{L(x)}^2 + \mu_{G(x)}^2 + c_1\right)\left(\sigma_{L(x)}^2 + \sigma_{G(x)}^2 + c_2\right)} \tag{15}$$

Among them, $L(x)$ symbolizes the ground truth, $G(x)$ symbolizes the generated result, M represents the number of pixels contained in the image, $\max^2(L(x),\, G(x))$ symbolizes the maximum image resolution value obtained by calculation, $\mu_{L(x)}$ and $\sigma_{L(x)}^2$ are the mean and variance of $L(x)$, $\mu_{G(x)}$ and $\sigma_{G(x)}^2$ are the mean and variance of $G(x)$, $\sigma_{L(x)G(x)}$ symbolizes the covariance of $L(x)$ with $G(x)$, and c_1 and c_2 are used to maintain stability.

4.3 Experiment description and parameter settings

The experiments in this article are based on Ubuntu 18.04.3 LTS (64-bit) system, the running memory is 24G, and the GPU is NVIDIA GeForce RTX 2080 Ti (11G). Data preprocessing, program coding implementation, model training, and testing are all in the Python 3.7.11 environment and the Pytorch 1.10.2 deep learning framework. The training epoch is set to 100, the batch size is set to 128, and the initial learning rate is set to 1e-3, after 30 epochs, the learning rate decays to 1e-4, and after 60 epochs, the learning rate decays to 1e-5. The optimization method adopts the Adam algorithm, the loss function employs the mean square error, the model and data are read and written in the form of files, and the final results are recorded and kept in accordance with the specifications.

4.4 Ablation study

In this part, we aim to design ablation experiments to verify the effectiveness of various parts in the network, which includes the study of the role of individual and composite components. In order to observe the role of each module more carefully, we conduct ablation research on the BSD68 and Set12 datasets, respectively. Please note that all our ablation experiments are performed on the denoising task with a noise level of 25.

First, in order to verify the role of residual connection, channel attention mechanism, and residual learning strategy, we conducted detailed comparative experiments on the role of single and composite components on the BSD68 and Set12 datasets, respectively. Specifically, we conducted experiments for all possible situations, recorded the PSNR results, and finally formed Table 1 (Y stands for containing the component, and N stands for not containing the

Table 1 Ablation study of key components on BSD68 and Set12 datasets

Datasets	Residual connection	Channel attention	Residual learning	PSNR
BSD68	N	N	Y	29.2549
	N	Y	N	29.2620
	Y	N	N	29.2948
	Y	Y	N	29.3210
	Y	N	Y	29.3025
	N	Y	Y	29.2940
	N	N	N	29.2302
	Y	**Y**	**Y**	**29.3389**
Set12	N	N	Y	30.5620
	N	Y	N	30.5987
	Y	N	N	30.6588
	Y	Y	N	30.6784
	Y	N	Y	30.6634
	N	Y	Y	30.6345
	N	N	N	30.5425
	Y	**Y**	**Y**	**30.6914**

The best results for each test image or dataset with each noise level are highlighted in bold

component). Observing Table 1, we can find that there are a total of eight possible situations for residual connection, channel attention mechanism, and residual learning strategy, that is, combining residual connection, channel attention mechanism, and residual learning strategy (Att-ResUNet), while not containing these three, containing only one, and a combination of two of them. Overall, the performance of the model combined with residual connection, channel attention mechanism, and residual learning strategy (Att-ResUNet) is the best. It achieved 29.3389 dB on the BSD68 dataset and 30.6914 dB on the Set12 dataset. Compared with the model which is without these three modules has a significant improvement. In addition, the contribution of each component to the entire model is also different, and the following is a more specific analysis.

Residual connection The skip connection in the UNet architecture can combine multi-scale information, but only using the skip connection can not fully acquire and utilize feature information in the denoising task. We introduce residual connections through the residual enhancement block, which can combine multi-scale contextual information and local feature information to achieve a better denoising effect. To verify this, we conducted comparative experiments. By observing Table 1, we found that after adding the residual connection, the performance on the BSD68 and Set12 datasets was improved from 29.2302 and 30.5425 dB to 29.2948 and 30.6588 dB, which is very significant. Moreover, we found that the general performance of the model after removing the residual connection is relatively low. These all prove the effectiveness of residual connections.

Channel attention mechanism In this paper, we introduce the SENet channel attention mechanism into the network to better focus on important feature information and reduce the loss of fine details of the image during the denoising process. By analyzing Table 1, we can find that after only using the channel attention mechanism in the original model, the result on the Set12 dataset is improved from 30.5425 to 30.5987 dB, and, after the channel

Table 2 Ablation study of different activation functions on BSD68 and Set12 datasets

Datasets	Models	PSNR
BSD68	ReLU	29.3034
	LeakyReLU	29.2940
	PReLU	**29.3389**
Set12	ReLU	30.6508
	LeakyReLU	30.6350
	PReLU	**30.6914**

The best results for each test image or dataset with each noise level are highlighted in bold

attention mechanism is combined with the residual connection, the BSD68 and Set12 datasets 29.3210 dB and 30.6784 dB, respectively, which can achieve better denoising performance compared to 29.2948 dB and 30.6588 dB using only residual connections, and are closer to the final model Att-ResUNet.

Residual learning strategy The residual learning strategy aims to acquire the noise information contained in the noisy image and then subtract the noise image and the noise to obtain a clean image. With the purpose of proving the effectiveness of the residual learning strategy, we make a comparative experiment. Please note that our research object is the strategy adopted by the algorithm model learning, that is, whether the model learns noise information or reconstructed feature information. The experimental results are displayed in Table 1. Among them, the PSNR value of the strategy of directly predicting clean images is 29.2302 dB and 30.5425 dB, while the PSNR value of using the residual learning strategy is 29.2549 dB and 30.5620 dB, which indicates that the residual learning strategy is more effective under the same conditions. Residual learning is more reasonable for applications in image restoration. In addition, by observing Table 1, we can find that the residual learning strategy can well assist the residual connection and the channel attention mechanism to make a greater contribution to the model as a whole. For example, when only using the channel attention mechanism on the BSD68 dataset, the PSNR is 29.2620 dB, while the PSNR after using the residual learning strategy is 29.2940 dB.

Then, to verify the role of the PReLU activation function and adaptive average pooling in Att-ResUNet, we organized ablation experiments to analyze the influence of activation functions and down-sampling strategies. Tables 2 and 3 present the experimental results,

Table 3 Ablation study of different pooling layers on BSD68 and Set12 datasets

Datasets	Models	PSNR
BSD68	Max-pooling	29.2879
	Average pooling	29.3250
	Adaptive average pooling	**29.3389**
Set12	Max-pooling	30.6457
	Average pooling	30.6745
	Adaptive average pooling	**30.6914**

The best results for each test image or dataset with each noise level are highlighted in bold

respectively. Experimental results show that the PReLU activation function and adaptive average pooling layer are more advantageous in our denoising framework.

4.5 Experimental results

In this section, we present and discuss the analysis of our experimental results from four aspects: comparison with state-of-the-art methods in the denoising task of Gaussian synthetic noise in grayscale and color images, studies of blind denoising and real-world noise removal. Moreover, there are analytical studies on the convergence speed and performance comparison of different models.

4.5.1 Comparison with state-of-the-art methods on Gaussian synthetic denoising tasks

For this part, we analyze the denoising effect of Att-ResUNet from both subjective perception and objective evaluation metrics. The objective evaluation indicators are PSNR and SSIM. We compare our model with 15 state-of-the-art denoising algorithms: 3D block matching (BM3D) [11], Weighted Kernel Norm Minimization Method (WNNM) [12], multilayer perceptron (MLP) [16], cascaded shrinking field (CSF) [17], TNRD [19], EPLL [55], DnCNN [20], IRCNN [21], FFDNet [22], ECNDNet [23], ADNet [24], CDNet [25], DudeNet [26], RDDCNN [27], and MWDCNN [28].

The average PSNR outcomes of various algorithms on the BSD68 dataset are displayed in Table 4, and the best model is displayed in bold font. It can be seen that our algorithm can achieve higher PSNR values than other algorithms at noise levels 15 and 25 and can achieve competitive results at noise level 50.

Tables 5, 6, and 7 list the PSNR value comparisons of different models for the 12 test

Table 4 Comparison of peak signal-to-noise ratio (dB) of different models on the BSD68 dataset

Methods	$\sigma = 15$	$\sigma = 25$	$\sigma = 50$
BM3D [11]	31.07	28.57	25.62
WNNM [12]	31.37	28.83	25.87
EPLL [55]	31.21	28.68	25.67
MLP [16]	–	28.96	26.03
CSF [17]	31.24	28.74	–
TNRD [19]	31.42	28.92	25.97
DnCNN [20]	31.73	29.23	26.23
IRCNN [21]	31.63	29.15	26.19
FFDNET [22]	31.62	29.19	26.29
ECNDNET [23]	31.71	29.22	26.23
ADNet [24]	31.74	29.25	26.29
CDNet [25]	31.74	29.28	**26.36**
DudeNet [26]	31.78	29.29	26.31
RDDCNN [27]	31.76	29.27	26.30
MWDCNN [28]	31.77	29.28	26.29
Att-ResUNet	**31.82**	**29.34**	**26.37**

Best model is displayed in bold font

Table 5 Comparison of peak signal-to-noise ratio (dB) of different models on the Set12 dataset ($\sigma = 15$)

Images	BM3D	EPLL	TNRD	DnCNN	IRCNN	FFDNet	ADNet	DudeNet	MWDCNN	Ours
C.man	31.91	31.85	32.19	32.61	32.55	32.43	**32.81**	32.71	32.53	*32.77*
House	34.93	34.17	34.53	34.97	34.89	35.07	**35.22**	35.13	35.09	*35.14*
Peppers	32.69	32.64	33.04	33.30	33.31	33.25	*33.49*	33.38	33.29	**33.50**
Starfish	31.14	31.13	31.75	32.20	32.02	31.99	32.17	*32.29*	32.28	**32.31**
Monarch	31.85	32.10	32.56	33.09	32.82	32.66	33.17	*33.28*	33.20	**33.42**
Airplane	31.07	31.19	31.46	31.70	31.70	31.57	**31.86**	*31.78*	31.74	**31.86**
Parrot	31.37	31.42	31.63	31.83	31.84	31.81	31.96	31.93	*31.97*	**31.99**
Lena	34.26	33.92	34.24	34.62	34.53	34.62	*34.71*	34.66	34.64	**34.72**
Barbara	33.10	31.38	32.13	32.64	32.43	32.54	*32.80*	32.73	32.65	**33.03**
Boat	32.13	31.93	32.14	32.42	32.34	32.38	*32.57*	32.46	32.49	**32.59**
Man	31.92	32.00	32.23	32.46	32.40	32.41	*32.47*	32.46	32.46	**32.50**
Couple	32.10	31.93	32.11	32.47	32.40	32.46	*32.58*	32.49	32.52	**32.60**
Average	32.37	32.14	32.50	32.86	32.77	32.77	*32.98*	32.94	32.91	**33.04**

The best PSNR results for each test image and each noise level are highlighted in bold font, and the second PSNR results are italicized

Table 6 Comparison of peak signal-to-noise ratio (dB) of different models on the Set12 dataset ($\sigma = 25$)

Images	BM3D	EPLL	TNRD	DnCNN	IRCNN	FFDNet	ADNet	DudeNet	MWDCNN	Ours
C.man	29.45	29.26	29.72	30.18	30.08	30.10	*30.34*	30.23	30.19	**30.36**
House	32.85	32.17	32.53	33.06	33.06	33.28	**33.41**	33.24	33.33	**33.41**
Peppers	30.16	30.17	30.57	30.87	30.88	30.93	**31.14**	30.98	30.85	*31.09*
Starfish	28.56	28.51	29.02	29.41	29.27	29.32	29.41	29.53	*29.66*	**29.75**
Monarch	29.25	29.39	29.85	30.28	30.09	30.08	30.39	30.44	*30.55*	**30.65**
Airplane	28.42	28.61	28.88	29.13	29.12	29.04	*29.17*	29.14	29.16	**29.26**
Parrot	28.93	28.95	29.18	29.43	29.47	29.44	*29.49*	29.48	29.48	**29.60**
Lena	32.07	31.73	32.00	32.44	32.43	32.57	32.61	32.52	*32.67*	**32.72**
Barbara	30.71	28.61	29.41	30.00	29.92	30.01	*30.25*	30.15	30.21	**30.62**
Boat	29.90	29.74	29.91	30.21	30.17	30.25	*30.37*	30.24	30.28	**30.41**
Man	29.61	29.66	29.87	30.10	30.04	*30.11*	30.08	30.08	30.10	**30.16**
Couple	29.71	29.53	29.71	30.12	30.08	30.20	*30.24*	30.15	30.13	**30.31**
Average	29.97	29.69	30.06	30.43	30.38	30.44	*30.58*	30.52	30.55	**30.69**

The best PSNR results for each test image and each noise level are highlighted in bold font, and the second PSNR results are italicized

Table 7 Comparison of peak signal-to-noise ratio (dB) of different models on the Set12 dataset ($\sigma = 50$)

Images	BM3D	EPLL	MLP	TNRD	DnCNN	IRCNN	ADNet	DudeNet	MWDCNN	Ours
C.man	26.13	26.10	26.37	26.62	27.03	26.88	*27.31*	27.22	26.99	**27.46**
House	29.69	29.12	29.64	29.48	30.00	29.96	*30.59*	30.27	30.58	**30.63**
Peppers	26.68	26.80	26.68	27.10	27.32	27.33	*27.69*	27.51	27.34	**27.75**
Starfish	25.04	25.12	25.43	25.42	25.70	25.57	25.70	*25.88*	25.85	**26.10**
Monarch	25.82	25.94	26.26	26.31	26.78	26.61	26.90	26.93	*27.02*	**27.05**
Airplane	25.10	25.31	25.56	25.59	25.87	25.89	25.88	25.88	*25.93*	**25.97**
Parrot	25.90	25.95	26.12	26.16	26.48	26.55	*26.56*	26.50	26.48	**26.72**
Lena	29.05	28.68	29.32	28.93	29.39	29.40	29.59	29.45	*29.63*	**29.75**
Barbara	27.22	24.83	25.24	25.70	26.22	26.24	*26.64*	26.49	26.60	**27.26**
Boat	26.78	26.74	27.03	26.94	27.20	27.17	*27.35*	27.26	27.23	**27.42**
Man	26.81	26.79	27.06	26.98	27.24	27.17	27.17	27.19	*27.27*	**27.31**
Couple	26.46	26.30	26.67	26.50	26.90	26.88	27.07	26.97	*27.11*	**27.23**
Average	26.72	26.47	26.78	26.81	27.18	27.14	*27.37*	27.30	27.34	**27.56**

The best PSNR results for each test image and each noise level are highlighted in bold font, and the second PSNR results are italicized

Table 8 Structural similarity comparison of methods under different noise levels

Dataset	Noise level	BM3D	DnCNN	FFDNet	IRCNN	ADNet	DudeNet	Ours
BSD68	15	0.8719	0.8907	0.8901	0.8882	0.8916	0.8918	**0.8925**
	25	0.8012	0.8279	0.8288	0.8248	0.8294	0.8294	**0.8302**
	50	0.6867	0.7189	0.7239	0.7169	0.7216	0.7167	**0.7242**
Set12	15	0.8948	0.9025	0.9025	0.9006	0.9049	0.9036	**0.9079**
	25	0.8493	0.8617	0.8632	0.8598	0.8654	0.8633	**0.8675**
	50	0.7664	0.7828	0.7899	0.7804	0.7898	0.7839	**0.7905**

The best results for each test image or dataset with each noise level are highlighted in bold

images in Set12 when the noise level is 15, 25, and 50, respectively. The best PSNR results for each test image and each noise level are highlighted in bold font, and the second PSNR results are italicized.

It can be seen from Tables 5, 6, and 7 that under the three noise levels, the method in this paper has achieved the best average PSNR results. The method in this paper only achieves suboptimal results on the 'C.man' and 'House' images when the noise level is 15, and on the 'Peppers' image, when the noise level is 25. In other cases, the algorithm in this paper can get the highest PSNR value. From the average PSNR of 12 images, the method of this paper is 0.06 ~ 1.10 dB ahead of all the comparison methods, and in the case of medium and high noise levels, the effect of this method is more prominent, which shows that Att-ResUNet is more suitable for dealing with high-intensity severe noise.

During the process of image denoising, it is inevitable to lose the structural details of the image. PSNR can objectively evaluate the quality of image generation, but it cannot fully evaluate the detailed preservation of the image. For this reason, we also select SSIM as the evaluation metric and compare the algorithm in this study with several state-of-the-art models in the aspect of structural similarity. As displayed in Table 8, it is clear that on the BSD68 and Set12 datasets, under the three noise levels, the algorithm in this study has achieved the best structural similarity, which shows that the method in this study can be used in the denoising process, and it can better reduce the loss of image detail information.

In terms of subjective perception analysis, Figs. 5 and 6 are the horizontal compared images of the original image, noisy map, DnCNN denoising map, IRCNN denoising map, FFDNET denoising map, and the algorithm denoising map in this paper at $\sigma = 25$ and $\sigma = 15$, respectively. For the purpose of further observing the details in the image, we encircle the part with a border and zoom in three times. Figures 5 and 6 show that the lines denoised by the algorithm in this paper appear smoother and more uniform, and are closer to the contour of the original image. From the direct comparison of the denoising effects of these images, it is obvious that the algorithm in this study can not only restore the clear edges of the original image but also achieve better structural similarity and visual results.

For the application scenario of color image denoising, we select five advanced algorithms for comparison: CBM3D [11], DnCNN, FFDNet, IRCNN, and DudeNet and test on three public datasets under five noise levels. The experimental results are as follows as displayed in Table 9. From the table, it is evident that our suggested Att-ResUNet is superior to the compared algorithms in various situations. In particular, we have conducted blind denoising experiments and found that the denoising effect of Att-ResUNet-B is similar to that of the Att-ResUNet. And with the rise of noise level, the denoising effect of Att-ResUNet-B is better than

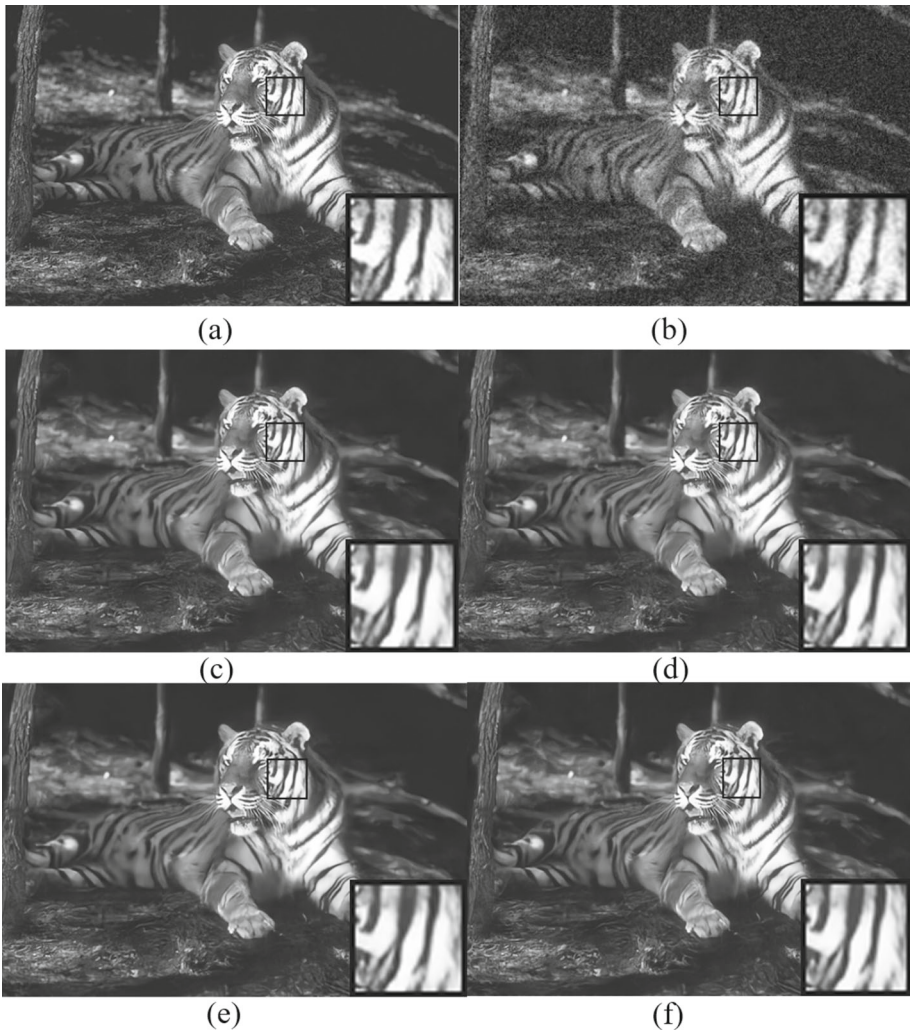

Fig. 5 Comparison of denoising effects of different methods on the same image from BSD68, $\sigma = 25$. **a** original image, **b** noise image, **c** DnCNN/PSNR = 29.14 dB, SSIM = 0.8394, **d** IRCNN/PSNR = 29.06 dB, SSIM = 0.8365, **e** FFDNet/PSNR = 29.11 dB, SSIM = 0.8399, and **f** Att-ResUNet/PSNR = 29.28 dB, SSIM = 0.8432

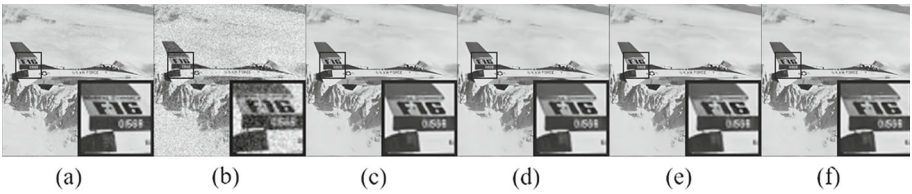

Fig. 6 Comparison of denoising effects of different methods on the same image from Set12, $\sigma = 15$. **a** original image, **b** noise image, **c** DnCNN/PSNR = 31.67 dB, SSIM = 0.9077, **d** IRCNN/PSNR = 31.66 dB, SSIM = 0.9064, **e** FFDNet/PSNR = 31.55 dB, SSIM = 0.9074, and **f** Att-ResUNet/PSNR = 31.86 dB, SSIM = 0.9142

Table 9 Comparison of denoising results of color images by different methods under different noise levels

Dataset	Noise level	CBM3D	DnCNN	FFDNet	IRCNN	DudeNet	Ours	Ours-B
CBSD68	15	33.52	33.98	33.80	33.86	33.86	**34.03**	33.93
	25	30.71	31.31	31.18	31.16	31.16	**31.32**	**31.33**
	35	28.89	29.65	29.57	29.50	29.71	**29.78**	**29.82**
	50	27.38	28.01	27.96	27.86	27.86	**28.01**	**28.14**
	75	25.74	–	26.24	–	26.40	**26.51**	**26.63**
Kodak24	15	34.28	34.73	34.55	34.56	34.56	**34.86**	34.83
	25	31.68	32.23	32.11	32.03	32.03	**32.32**	**32.41**
	35	29.90	30.64	30.56	30.43	30.69	**30.90**	**30.98**
	50	28.46	29.02	28.99	28.81	28.81	**29.13**	**29.33**
	75	26.82	–	27.25	–	27.39	**27.68**	**27.75**
McMaster	15	34.06	34.80	34.47	34.58	34.58	**34.98**	34.97
	25	31.66	32.47	32.25	32.18	32.18	**32.58**	**32.73**
	35	29.92	30.91	30.76	30.59	30.86	**31.25**	**31.38**
	50	28.51	29.21	29.14	28.91	28.91	**29.41**	**29.65**
	75	26.79	–	27.29	–	–	**27.88**	**28.04**

The best results for each test image or dataset with each noise level are highlighted in bold

that of Att-ResUNet, which also indicates that our model has strong blind denoising ability and at high noise level excellent performance. Figures 7 and 8 show the horizontal comparison of the original image, the noise map, the IRCNN denoising map, and the denoising map of the algorithm in this paper under $\sigma = 25$ and $\sigma = 15$. To further observe the details in the image, we encircle the area with a green border and zoom in three times. Through intuition, it can be found that our algorithm can achieve better visual effects.

4.5.2 Research on the application of blind denoising

To further explore the blind denoising ability of our model, we compare the method in this study with the DnCNN-B and CDNet models which have excellent blind denoising effects. Tables 10 and 11 list the denoising results of the method proposed in this paper and the DnCNN-B and CDNet models under the noise intensity of 15–50 on the Set12 and BSD68 datasets, respectively. The average PSNR and average SSIM based on the Set12 and BSD68 test sets for a total of eight noise intensities are recorded. From the data analysis in Tables 10 and 11, it can be seen that whether it is Att-ResUNet-B or DnCNN-B and CDNet proposed in this study when the noise intensity increases sequentially, the denoising effect will be weakened, which is a normal phenomenon. Because the noise intensity has an inverse relationship with the denoising effect. In terms of horizontal comparison, under any noise intensity, the algorithm in this paper can achieve higher PSNR and SSIM values than DnCNN-B and CDNet, indicating that our method still guarantees certain robustness under different noise intensities, and Att-ResUNet-B can still maintain good structural similarity when the noise level reaches 35. On the whole, it can be seen that the algorithm of this research has certain advantages in denoising effect compared with DnCNN-B and CDNet under any

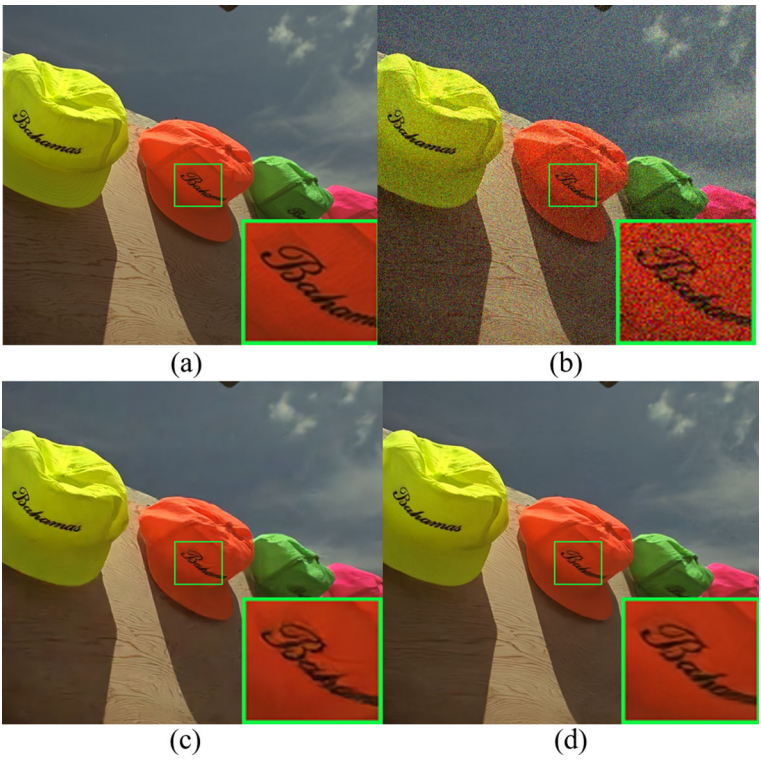

Fig. 7 Comparison of denoising effects of different methods on the same image from Kodak24, $\sigma = 25$. **a** original image, **b** noise image, **c** IRCNN/PSNR = 35.01 dB, and **d** Att-ResUNet/PSNR = 35.59 dB

noise intensity, has better blind denoising ability, and is more suitable for processing blind denoising tasks.

4.5.3 Research on noise removal in the real world

Considering that real-world noise is more complex and challenging to manage, we design and implement real-world noise removal experiments to verify the ability of Att-ResUNet to handle complex denoising tasks. The dataset we use is the publicly available real-world noise dataset CC. The CC dataset contains 15 images taken by digital cameras. Different from the original data, these 15 images are all processed into a size of 512×512, which can be divided into five groups of camera sources according to different sources. We compare Att-ResUNet with four state-of-the-art models, namely, DnCNN, ADNet, DudeNet, and MWDCNN. Table 12 compares the results of five different denoising methods. Observing Table 12, we can find that our method significantly outperforms the state-of-the-art methods in 15 cases. It is worth noting that our method achieves an average denoising result of 37.88 dB, an improvement of 2.14 dB over the state-of-the-art method, a very big boost. Furthermore, we can find that Att-ResUNet achieves a PSNR value of 40.00 dB in five cases, which is an excellent denoising result. Based on the above analysis, it can be proved that Att-ResUNet has remarkable performance in the task of removing real-world noise.

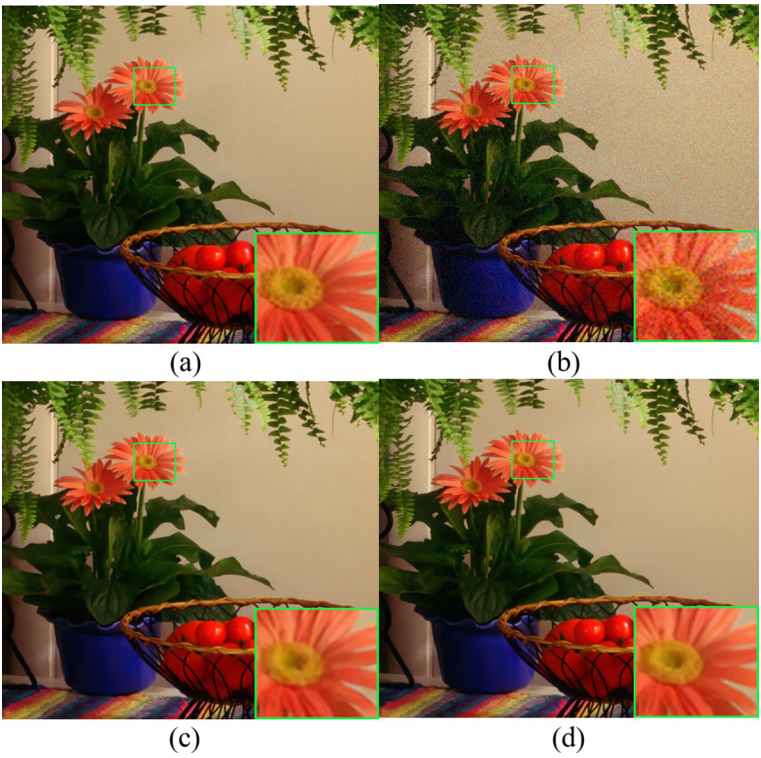

Fig. 8 Comparison of denoising effects of different methods on the same image from McMaster, $\sigma = 15$. **a** original image, **b** noise image, **c** IRCNN/PSNR = 35.29 dB, and **d** Att-ResUNet/PSNR = 35.67 dB

Table 10 Comparison of model blind denoising results under different noise levels on Set12

Datasets	Noise	DnCNN-B		CDNet		Att-ResUNet-B	
	level	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Set12	15	32.67	0.9000	32.79	0.9022	**32.85**	**0.9029**
	20	31.37	0.8791	31.43	0.8798	**31.56**	**0.8825**
	25	30.35	0.8599	30.47	0.8616	**30.61**	**0.8675**
	30	29.52	0.8422	29.56	0.8426	**29.70**	**0.8478**
	35	28.82	0.8256	28.86	0.8211	**28.98**	**0.8312**
	40	28.20	0.8101	28.25	0.8103	**28.38**	**0.8168**
	45	27.66	0.7954	27.74	0.7962	**27.87**	**0.8039**
	50	27.18	0.7816	27.34	0.7894	**27.38**	**0.7905**

The best results for each test image or dataset with each noise level are highlighted in bold

Table 11 Comparison of model blind denoising results under different noise levels on BSD68

Datasets	Noise level	DnCNN-B		CDNet		Att-ResUNet-B	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
BSD68	15	31.62	0.8868	31.64	0.8911	**31.74**	**0.8916**
	20	30.21	0.8537	30.22	0.8538	**30.36**	**0.8545**
	25	29.16	0.8244	29.18	0.8228	**29.21**	**0.8247**
	30	28.35	0.7982	28.37	0.7981	**28.40**	**0.7986**
	35	27.68	0.7744	27.61	0.7764	**27.74**	**0.7765**
	40	27.12	0.7528	27.12	0.7529	**27.18**	**0.7535**
	45	26.64	0.7331	26.68	0.7342	**26.72**	**0.7344**
	50	26.23	0.7164	26.27	0.7140	**26.28**	**0.7166**

The best results for each test image or dataset with each noise level are highlighted in bold

Table 12 Comparison of real-world denoising PSNR results by different methods on the CC dataset

Dataset settings	DnCNN	ADNet	DudeNet	MWDCNN	Ours
Nikon-d600 ISO_3200_1	33.62	33.94	33.72	33.91	**35.47**
Nikon-d600 ISO_3200_2	34.48	34.33	34.70	34.88	**37.02**
Nikon-d600 ISO_3200_3	35.41	38.87	37.98	37.02	**41.35**
Nikon-d800 ISO_1600_1	37.95	37.61	38.10	37.93	**39.26**
Nikon-d800 ISO_1600_2	36.08	38.24	39.15	37.49	**41.77**
Nikon-d800 ISO_1600_3	35.48	36.89	36.14	38.44	**39.23**
Nikon-d800 ISO_3200_1	34.08	37.20	36.93	37.10	**40.28**
Nikon-d800 ISO_3200_2	33.70	35.67	35.80	36.72	**37.65**
Nikon-d800 ISO_3200_3	33.31	38.09	37.49	37.25	**41.39**
Nikon-d800 ISO_6400_1	29.83	32.24	31.94	32.24	**33.62**
Nikon-d800 ISO_6400_2	30.55	32.59	32.51	32.56	**33.51**
Nikon-d800 ISO_6400_3	30.09	33.14	32.91	32.76	**33.39**
Canon-5d ISO_3200_1	37.26	35.69	36.66	36.97	**40.13**
Canon-5d ISO_3200_2	34.87	36.11	36.70	36.01	**37.14**
Canon-5d ISO_3200_3	34.09	34.49	35.03	34.80	**37.03**
Average-PSNR	33.86	35.69	35.72	35.74	**37.88**

The best results for each test image or dataset with each noise level are highlighted in bold

4.5.4 Research on model convergence speed and performance comparison

The convergence speed of the denoising model is also a crucial factor to evaluate the performance of the model. A denoising model with fast convergence can achieve good denoising performance with fewer training epochs. On the contrary, denoising models with slow convergence require more training time and computing resources and have poor stability. Intending to study the convergence speed and performance of the proposed Att-ResUNet, we set up

experiments to compare the convergence speed and performance changes of different models and drew curves as shown in Figs. 9 and 10 for analysis.

Specifically, we reproduce DnCNN, ADNet, and DudeNet and train them strictly according to the training set and strategy indicated in the reference. We train each model under the Gaussian denoising task with a noise level of 25 and achieve their proposed optimal PSNR results to ensure fairness. Please note that different methods (including our proposed Att-ResUNet) use training datasets. The strategies are different, and the training and test sets are non-overlapping. We saved the weights of the first 60 epochs for each model and then tested and recorded the average PSNR value of the model on the test set under each epoch of 1–60

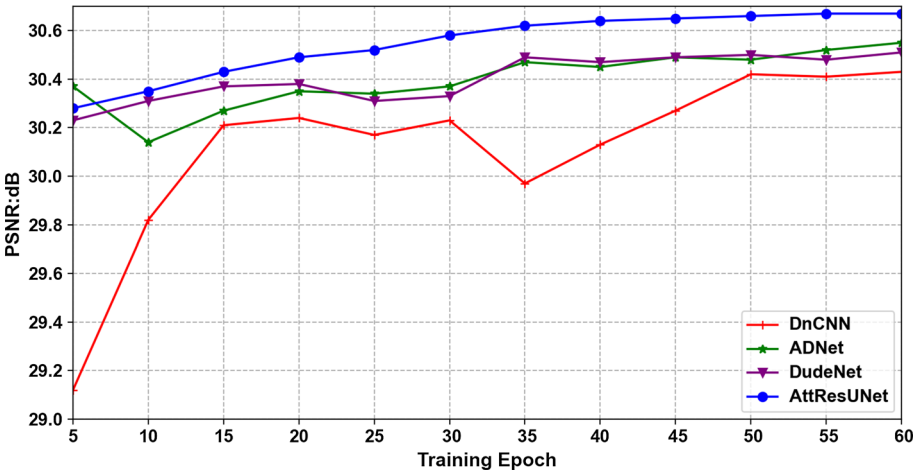

Fig. 9 Comparison of convergence speed and performance changes of different methods on the Set12 dataset ($\sigma = 25$)

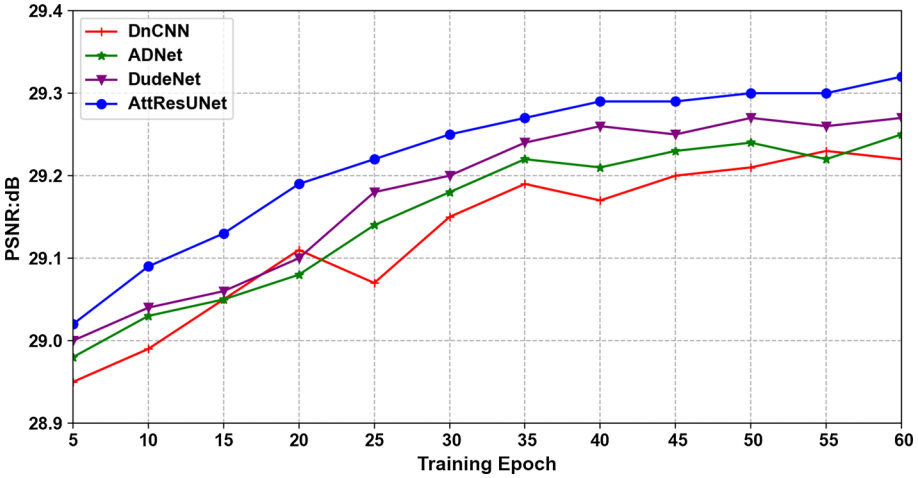

Fig. 10 Comparison of convergence speed and performance changes of different methods on the BSD68 dataset ($\sigma = 25$)

on the two test datasets of Set12 and BSD68, respectively. Finally, according to the obtained results, the transformation curves are drawn as shown in Figs. 9 and 10. The abscissa in the figure represents Epoch, a total of 60 Epochs, and the ordinate represents the average PSNR (dB) of the current method on the test set. Obviously, the convergence trend of the algorithm in this paper is similar to that of the early DnCNN on BSD68, but when the iteration cycle reaches 30, the algorithm in this paper has tended to be stable. At this time, the average PSNR on the Set12 test set is 30.58 dB. While the DnCNN network still has an unstable trend in the 30th epoch and beyond. In addition, it can be found that Att-ResUNet achieves good denoising performance at the fastest speed on both test datasets. The performance change curves of ADNet and DudeNet are similar, and both have the problem of slow convergence. Based on the above analysis, it can be shown that the method we propose has a faster convergence speed and good stability.

5 Conclusion

In this study, we propose a novel image denoising model combining channel attention mechanism and residual UNet network to solve the problem that existing methods cannot fully capture and utilize important information in complex denoising tasks. Our proposed residual enhancement block can combine skip connections to form global and local residuals to combine multi-scale global context information and local feature information. The channel attention mechanism can better focus on important details and preserve fine details of images. An adaptive average pooling layer can preserve more edge information during downsampling. At the same time, the residual learning strategy can speed up the training and help to improve the denoising ability of the model. Experimental results comparing with 15 methods show that Att-ResUNet exhibits superior performance compared to other related models and is competitive under multiple denoising tasks. In particular, Att-ResUNet achieves an average PSNR result of 37.88 dB on the real-world denoising CC dataset, an improvement of 2.14 dB compared to the state-of-the-art method. Att-ResUNet has significant advantages in performance methods, but the calculation amount of the model is a bit larger than that of DnCNN and other models, and there is room for further improvement. In the future work, it is necessary for us to improve the model and try to design a lighter model. Increase efficiency while maintaining high performance.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grant Nos. 62276265, 61976216, and 61672522.

Author contributions Prof. Shifei Ding helped in supervision. Dr. Qidong Wang helped in conceptualization and methodology. Dr. Lili Guo helped in supervision. Dr. Jian Zhang worked in software and writing. Dr. Ling Ding helped in supervision. All authors reviewed the manuscript.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. He W, Zhang H, Shen H, Zhang L (2018) Hyperspectral image denoising using local low-rank matrix recovery and global spatial–spectral total variation. IEEE J Sel Top Appl Earth Observ Remote Sens 11(3):713–729

2. Shi Q, Tang X, Yang T, Liu R, Zhang L (2021) Hyperspectral image denoising using a 3-D attention denoising network. IEEE Trans Geosci Remote Sens 59(12):10348–10363
3. Pan E, Ma Y, Mei X, Fan F, Huang J, Ma J (2022) Sqad: spatial-spectral quasi-attention recurrent network for hyperspectral image denoising. IEEE Trans Geosci Remote Sens 60:1–14
4. Zhao W, Lu H (2017) Medical image fusion and denoising with alternating sequential filter and adaptive fractional order total variation. IEEE Trans Instrum Meas 66(9):2283–2294
5. Chen M, Pu YF, Bai YC (2021) Low-dose CT image denoising using residual convolutional network with fractional TV loss. Neurocomputing 452:510–520
6. Geng M, Meng X, Zhu L, Jiang Z, Gao M, Huang Z, Lu Y (2022) Triplet cross-fusion learning for unpaired image denoising in optical coherence tomography. IEEE Trans Med Imaging 41(11):3357–3372
7. Buades A, Coll B, Morel JM (2005) A review of image denoising algorithms, with a new one. Multiscale Model Simul 4(2):490–530
8. Thakur RS, Yadav RN, Gupta L (2019) State-of-art analysis of image denoising methods using convolutional neural networks. IET Image Proc 13(13):2367–2380
9. Tian C, Fei L, Zheng W, Xu Y, Zuo W, Lin CW (2020) Deep learning on image denoising: an overview. Neural Netw 131:251–275
10. Buades A, Coll, B, Morel, JM (2005) A non-local algorithm for image denoising. In: 2005 IEEE Computer society conference on computer vision and pattern recognition (CVPR'05), vol 2, IEEE, pp 60–65
11. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Trans Image Process 16(8):2080–2095
12. Gu S, Zhang L, Zuo W, Feng X (2014) Weighted nuclear norm minimization with application to image denoising. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2862–2869
13. Aharon M, Elad M, Bruckstein A (2006) K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans Signal Process 54(11):4311–4322
14. Vardan P, Yaniv R, Jeremias S, Michael E (2018) Theoretical foundations of deep learning via sparse representations: a multilayer sparse model and its connection to convolutional neural networks. IEEE Signal Process Mag 35(4):72–89
15. Jain V, Murray JF, Roth F, Turaga S, Zhigulin V, Briggman KL, Seung H S (2007) Supervised learning of image restoration with convolutional networks. In: 2007 IEEE 11th International Conference on Computer Vision, IEEE, pp 1–8
16. Burger HC, Schuler CJ, Harmeling S (2012) Image denoising: Can plain neural networks compete with bm3d? In: 2012 IEEE conference on computer vision and pattern recognition, IEEE, pp 2392–2399
17. Schmidt U, Roth S (2014) Shrinkage fields for effective image restoration. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2774–2781
18. Mao X, Shen C, Yang YB (2016) Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Advances in neural information processing systems. https://doi.org/10.48550/arXiv.1603.09056
19. Chen Y, Pock T (2016) Trainable nonlinear reaction diffusion: a flexible framework for fast and effective image restoration. IEEE Trans Pattern Anal Mach Intell 39(6):1256–1272
20. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: residual learning of deep cnn for image denoising. IEEE Trans Image Process 26(7):3142–3155
21. 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
22. Zhang K, Zuo W, Zhang L (2018) FFDNet: toward a fast and flexible solution for CNN-based image denoising. IEEE Trans Image Process 27(9):4608–4622
23. Tian C, Xu Y, Fei L, Wang J, Wen J, Luo N (2019) Enhanced cnn for image denoising. CAAI Trans Intell Technol 4(1):17–23
24. Tian C, Xu Y, Li Z, Zuo W, Fei L, Liu H (2020) Attention-guided CNN for image denoising. Neural Netw 124:117–129
25. Quan Y, Chen Y, Shao Y, Teng H, Xu Y, Ji H (2021) Image denoising using complex-valued deep CNN. Pattern Recognit 111:107639
26. Tian C, Xu Y, Zuo W, Du B, Lin CW, Zhang D (2021) Designing and training of a dual CNN for image denoising. Knowl Based Syst 226:106949
27. Zhang Q, Xiao J, Tian C, Chun-Wei Lin J, Zhang S (2022) A robust deformed convolutional neural network (CNN) for image denoising. CAAI Trans Intell Technol. https://doi.org/10.1049/cit2.12110
28. Tian C, Zheng M, Zuo W, Zhang B, Zhang Y, Zhang D (2023) Multi-stage image denoising with the wavelet transform. Pattern Recognit 134:109050

29. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention, Springer, pp 234–241
30. Li C, Tan Y, Chen W, Luo X, Gao Y, Jia X, Wang Z (2020) Attention Unet++: a nested attention-aware U-Net for liver CT image segmentation. In: 2020 IEEE international conference on image processing, IEEE, pp 345–349
31. Amer A, Ye X, Zolgharni M, Janan F (2020) ResDUnet: residual dilated UNet for left ventricle segmentation from echocardiographic images. In: 2020 42nd Annual international conference of the IEEE engineering in medicine & biology society (EMBC), IEEE, pp 2019–2022
32. Han Z, Jian M, Wang GG (2022) ConvUNeXt: an efficient convolution neural network for medical image segmentation. Knowl Based Syst 253:109512
33. Lin A, Chen B, Xu J, Zhang Z, Lu G, Zhang D (2022) Ds-transunet: dual swin transformer u-net for medical image segmentation. IEEE Trans Instrum Meas 71:1–15
34. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas, pp 770–778
35. Zhang Y, Li J, Wei S, Zhou F, Li D (2021) Heartbeats classification using hybrid time-frequency analysis and transfer learning based on ResNet. IEEE J Biomed Health Inform 25(11):4175–4184
36. Zhang Z, Liu Q, Wang Y (2018) Road extraction by deep residual u-net. IEEE Geosci Remote Sens Lett 15(5):749–753
37. Sun T, Ding S, Guo L (2022) Low-degree term first in ResNet, its variants and the whole neural network family. Neural Netw 148:155–165
38. Dentamaro V, Giglio P, Impedovo D, Moretti L, Pirlo G (2022) AUCO ResNet: an end-to-end network for Covid-19 pre-screening from cough and breath. Pattern Recogn 127:108656
39. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7132–7141
40. Roy SK, Dubey SR, Chatterjee S, Chaudhuri BB (2020) FuSENet: fused squeeze-and-excitation network for spectral-spatial hyperspectral image classification. IET Image Proc 14(8):1653–1661
41. Li Y, Liu Y, Cui WG, Guo YZ, Huang H, Hu ZY (2020) Epileptic seizure detection in EEG signals using a unified temporal-spectral squeeze-and-excitation network. IEEE Trans Neural Syst Rehabil Eng 28(4):782–794
42. Li G, Fang Q, Zha L, Gao X, Zheng N (2022) HAM: hybrid attention module in deep convolutional neural networks for image classification. Pattern Recognit 129:108785
43. Cheng J, Tian S, Yu L, Gao C, Kang X, Ma X, Lu H (2022) ResGANet: residual group attention network for medical image classification and segmentation. Med Image Anal 76:102313
44. Martin D, Fowlkes C, Tal D, & Malik J (2001) A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision, IEEE, vol. 2, pp 416-423
45. Ma K, Duanmu Z, Wu Q, Wang Z, Yong H, Li H, Zhang L (2016) Waterloo exploration database: new challenges for image quality assessment models. IEEE Trans Image Process 26(2):1004–1016
46. Agustsson E & Timofte R (2017) Ntire 2017 challenge on single image super-resolution: dataset and study. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 126–135
47. Xu J, Li H, Liang Z, Zhang D, & Zhang L (2018) Real-world noisy image denoising: a new benchmark. arXiv preprint arXiv:1804.02603
48. Roth S, Black MJ (2005) Fields of experts: A framework for learning image priors. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), Vol 2, Citeseer, pp 860–867
49. Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A (2009) Non-local sparse models for image restoration. In: 2009 IEEE 12th international conference on computer vision, IEEE, pp 2272–2279
50. Franzen R (1999) Kodak lossless true color image suite, vol 4, http://r0k.us/graphics/kodak
51. Zhang L, Wu X, Buades A, Li X (2011) Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. J Electron Imaging 20(2):023016
52. Nam S, Hwang Y, Matsushita Y, & Kim S J (2016) A holistic approach to cross-channel image noise modeling and its application to image denoising. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1683–1691
53. Huynh-Thu Q, Ghanbari M (2008) Scope of validity of psnr in image/video quality assessment. Electron Lett 44(13):800–801
54. Hore A, Ziou D (2010) Image quality metrics: PSNR vs. SSIM. In: 2010 20th international conference on pattern recognition, IEEE, pp 2366–2369

55. D Zoran, Weiss Y (2011) From learning models of natural image patches to whole image restoration. In: 2011 International conference on computer vision, pp 479–486

Shifei Ding (Member, IEEE) was born in Qingdao, China, in 1963, and received his Ph.D. degree from Shandong University of Science and Technology, China, in 2004. He is a professor and Ph.D. supervisor at China University of Mining and Technology. His research interests include intelligent information processing, pattern recognition, machine learning, data mining, and granular computing.

Qidong Wang was born in Zaozhuang, China, and received the B.S. degree from University of Jinan, Jinan, China in 2021. He is currently pursuing M.S. degree in China University of Mining and Technology. His research interests include deep learning and image analysis.

Lili Guo (Member, IEEE) was born in Linyi, China, in 1990, and received the Ph.D. degree in college of intelligence and computing from Tianjin University, Tianjin, China, in 2021. She is currently a lecturer with school of computer science and technology, China University of Mining and Technology, China. Her research interests are in the fields of speech emotion recognition, deep learning, and acoustic signal processing.

Jian Zhang was born in Taian, China, in 1990, and received the Ph.D. degree in college of Computer Science of Technology from China University of Mining and Technology, Xuzhou, China, in 2020. He is currently a lecturer with school of computer science and technology, China University of Mining and Technology, China. His research interests are in the fields of deep learning and multi-label learning.

Ling Ding received her B.S. degree and M.S. degree from Asia Pacific University of Technology and Innovation in 2017 and 2019, respectively. She is currently working toward the Ph.D. degree in Tianjin University; her supervisor is Dr. Di Jin. Her research interests include deep learning, graph machine learning, clustering, etc.