Journal of Computer Science and Artificial Intelligence
ISSN: 3078-8242 (Print), ISSN: 3078-2309 (Online) | Vol. 2, No. 1, 2025

Separable Multi-scale Large Kernel Convolutional Remote Sensing Denoising Network

Gui Luo, Xiangguo Sun *

College of Mechanical Engineering, Sichuan University of Science & Engineering, Zigong Sichuan 643002, China

Abstract: Abstract: The abstract of the study stated that remote sensing images contain abundant details of land objects and terrain, and the denoising process should strive to preserve these critical pieces of information. However, traditional CNN methods performed poorly when dealing with high-resolution, multi-scale, and complex scenes, as they struggled to model the long-range dependencies within images. Methods based on Transformer improved this issue through the self-attention mechanism; however, their high computational cost limited their application in resource-constrained environments. To address this, a Multi-Scale Large Kernel Detail Enhancement Network was proposed, aiming to effectively retain the detailed information in remote sensing images. By utilizing pooling to separate high and low-frequency information, the approach adopted separable multi-scale large kernel convolutions to capture extensive spatial information, enhancing high-frequency features while reducing computational complexity. These innovative techniques effectively expanded the receptive field, improving the denoising effect of remote sensing images. Currently, compared with the best results from other methods, MLKNet achieves an average improvement of approximately 3.1 dB in grayscale remote sensing image denoising across three different noise levels, and an average improvement of about 1.17 dB in color remote sensing image denoising under the same conditions.

Keywords: Image Denoising, Multi-scale, Frequency Separation, Large Kernel Convolution, Remote Sensing Images.

1. Introduction

During the generation of remote sensing products, denoising techniques can significantly improve the quality of the generated products. Remote sensing images are widely used in fields such as map making, environmental monitoring, and agricultural management. However, the quality of remote sensing images is affected by various imaging system and environmental factors, like vibrations of satellite platforms, ground artifacts and noise under low light conditions. As shown in Figure 1, the remote sensing image obtained from the "Earth Big Data Science Data Center of the Chinese Academy of Sciences", when the marked parts in it are enlarged, a large amount of noise can be seen.

The removal of these noises by using denoising techniques can effectively improve the efficiency and accuracy of subsequent work [1]. These multiple factors result in complex noise components, making the semantic information of remote sensing images more complicated than that of natural images and seriously affecting the accuracy of image segmentation and small target recognition. Therefore, it is particularly important to remove noises and improve image quality.

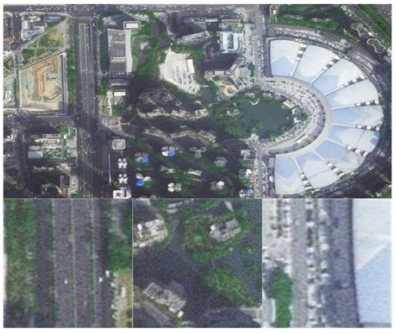

Figure 1. Real noise map

As shown in Figure 2.

Figure 2. Remote sensing image feature map

Firstly, the objects in remote sensing images usually exhibit specific textures and shapes. For example, roads have regular straight edges, and buildings present contours in the form of rectangles or combinations of rectangles. Secondly, aerial images are usually captured from a bird's-eye view at high resolution, featuring high resolution, random orientations, large variations within classes, multi-scale scenes and dense small objects [2]. Thirdly, as shown in the figure, the marking lines on the roads and cars are small, and the information changes rapidly in space.

Based on the characteristics of remote sensing images, this paper makes the following contributions:

(1) Good denoising performance requires extensive contextual information. In remote sensing image denoising, limited context information may lead to the failure of detectors to effectively identify all types of noise. Separable large convolution kernels at different scales are established according to the multi-scale characteristics and dense small objects. Within a relatively large range of small objects, there are many similar structures, and large convolution kernels can provide a larger receptive field to obtain more similar features.

(2) During the convolution process, detailed features are

easily overlooked. Therefore, pooling operations are used to separate high-frequency and low-frequency features, and multi-scale convolution is adopted to strengthen high-frequency features.

2. Related Work

Image denoising is a low-level vision problem in computer vision. According to different principles, it can be divided into traditional image denoising methods and image denoising methods based on deep learning.

For traditional remote sensing image denoising methods, BM3D [3] adopts regularized inverse transformation and collaborative filtering based on transform-domain shrinkage. The K-SVD [4] algorithm is a dictionary learning-based method. The weighted nuclear norm minimization method proposed by WNNM [5] retains more details compared to BM3D and K-SVD, but it may still over-smooth the image, resulting in blurred edges or loss of details. This is especially likely to occur in areas of the image where there are fine textures or high-frequency details. Traditional denoising methods are relatively simple to implement, but they have difficulty dealing with complex images with strong noise, and their computational efficiency is also low.

In recent years, methods based on deep learning have been able to adapt to different data distributions and noise patterns through learning. They have achieved better denoising effects and possess strong generalization abilities, and have gradually become the mainstream methods in the field of remote sensing image denoising.

DnCNN [6] employs residual learning and batch normalization techniques to accelerate the training process and improve denoising performance. It learns the residual between the noisy image and the clean image, that is, the noise itself, which has promoted the application of deep learning in the field of image denoising and opened a new chapter in this field. In reference [7], deconvolution is added on the basis of DnCNN to enhance detail restoration, thus constructing an approximately U-shaped network structure. FFDNet [8] proposes a model to denoise images with different noise levels and effectively expands the receptive field by means of downsampling. ECNDNet [9] also uses residual learning and batch normalization techniques and uses dilated convolution to improve the ability to acquire information. On this basis, ADNet [10] has studied the application of attention-guided convolutional neural networks in image denoising. Adding attention endows the model with stronger denoising capabilities.

RSIDNet [11] combines the idea of multi-scale and the combination of multiple modules, which greatly improves the denoising ability. However, it also increases the computational load and requires large computational resources to complete training and inference, which may become a limitation on platforms with limited computational resources.

VIT [12] applies the Transformer to the field of image processing. The self-attention mechanism of the Transformer can capture the long-distance spatial dependency relationships in the image and can help the model better understand different parts in the image and the associations between them. Image Processing Transformer (IPT) [13], SwinIR [14], Uformer [15], and MAXIM [16] have enabled the Transformer to achieve good results in the field of denoising. However, the self-attention computation is the main cost of the Transformer. Restormer [17] uses 1×1 pointwise convolution and 3×3 depthwise convolution to form QKV and applies self-attention in a cross-channel manner. Xformer [18] and LGDNet [19] further apply a dual-branch structure of channel plus space. All of them are reducing the large amount of computation brought by self-attention, but the computational load is still relatively large compared to convolution.

To sum up, there has been no research on the application of separable large-kernel convolution in denoising. A series of large separable kernel attention modules have been proposed in LSKA [20] and LCAN [21], which can obtain the receptive field of large-kernel convolution without bringing the large computational load of large-kernel convolution. Compared with ordinary image denoising, in remote sensing image denoising, object recognition in remote sensing images often relies on the context information provided by the surrounding environment, and it is necessary to establish longer distance dependency relationships.

The idea of the Large Coordinate Kernel Attention Network (LCAN) is introduced. More context information is obtained through large-kernel convolution to distinguish features from noise. Multi-scale large-kernel convolution is established by using 1D depthwise separable convolution kernels and 1D depthwise separable dilated convolution, which can achieve performance comparable to that of the standard LKA module [22]. In the task of remote sensing image denoising, the images of small objects may consist of only dozens of pixels. For small objects, edges and details become more important. Therefore, general convolutional networks often face the problem of insufficient high-frequency details. To effectively solve this problem, we adopt the pooling layer to separate high-frequency and low-frequency features.

3. Denoising Model

3.1. Network Structure——MLKNet

As shown in Figure 3, this paper adopts a U-Net model structure deformed based on NAFNet [23]. The overall structure consists of an encoder, a decoder, skip connections, and a bottleneck. The skip connections between the encoder and the decoder use a 1×1 convolution kernel to perform a linear combination on the input features, fusing the shallow detail features with the deep abstract features to better capture the detail features.

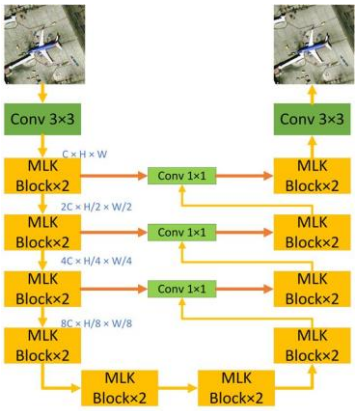

Figure 3. Overall structure of the network

As illustrated in Figure 4, the structure within the block is devised based on the U-Net model. Each block comprises two residual sub-blocks. The first sub-block is made up of two layers of 3×3 convolutions and an activation function. In the

second sub-block, the multi-scale large-kernel convolution fusion module, namely MLK, is incorporated.

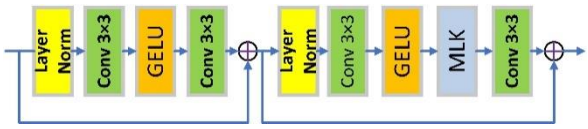

Figure 4. MLK structure diagram

3.2. Feature Separation and Enhancement

In the realm of remote sensing image denoising, augmenting high-frequency detailed information holds paramount significance. Regarding the techniques for segregating high-frequency and low-frequency characteristics, the Fourier transform represents one alternative. Nevertheless, it invariably demands substantial computational resources and exacerbates both the intricacy of the network architecture and the challenges associated with training. To circumvent this issue, a pooling layer [24] is employed. The input features are processed via average pooling to derive low-frequency features. Subsequently, these low-frequency features undergo upsampling. By deducting the upsampled low-frequency features from the initial input features, as illustrated in Figure 5, the high-frequency detail information can be procured. Precisely, the feature map is initially averaged pooled to engender the low-frequency feature Fl. Thereafter, Fl is upsampled to correspond to the size of F. Eventually, subtracting Fl from F begets the requisite high-frequency detail features.

$$F_l = avgpool(F) \tag{1}$$

$$F_h = F - upsample(F_l) \tag{2}$$

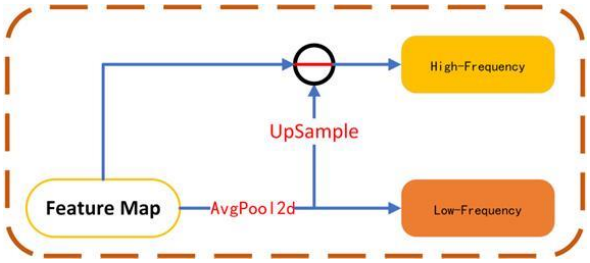

Figure 5. Frequency separation diagram

Dissociate the high-frequency features from the low-frequency ones, and subsequently employ separable large-kernel convolutions with varying sizes to amass the multi-scale high-frequency features. The utilization of convolution kernels of diverse dimensions facilitates the model in apprehending the features across different scales. Given that the manifestations of high-frequency traits (such as edges and details) differ at disparate scales, multi-scale convolution can thus more comprehensively seize these nuances. Implementing separable large-kernel convolution can markedly curtail the computational complexity and procure context information spanning a broader scope.

3.3. Large Separable Kernel Attention

We propose a multi-scale high-frequency feature fusion module with the aim of aggregating multi-scale information. In pursuit of obtaining an enlarged receptive field and alleviating the substantial computational overhead incurred by large convolution kernels, we employ conventional 1D depthwise convolutions and 1D depthwise dilated convolutions to formulate large separable kernel convolutions.

Multiple large separable kernel convolutions are configured in a parallel manner to extract multi-scale features. As depicted in Figure 6, the separable large-kernel convolution module is constituted of five convolution blocks. Features are subjected to depthwise convolutions along the height dimension, which predominantly function to extract local information. Subsequently, a depthwise dilated convolution along the height dimension is executed to expand the receptive field and capture more comprehensive context information, thereby facilitating the establishment of long-range dependencies. Thereafter, this methodology is replicated to procure local features and long-range dependencies in the width direction. Finally, 1×1 convolutions are deployed to actualize the fusion of features among different channels.

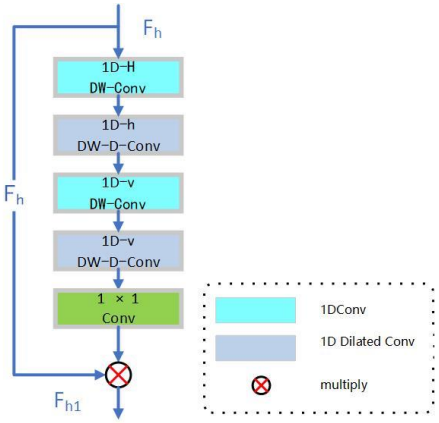

Figure 6. Separable large kernel convolution

3.4. Multi-scale High-frequency Fusion

In the academic paper [25], the construction of dual-branch multi-scale spatial information was put forward. It resorted to 3×3 branch convolutions; however, the receptive field thus engendered was comparatively circumscribed. Through the utilization of large-kernel convolutions to devise a three-branch framework, more expansive context information can be acquired under three heterogeneous convolution kernel modalities. As depicted in Figure 7, the high-frequency features dissociated by means of pooling Fh are inputted into the three-branch convolution blocks, denoted as K1, K2, and K3, generating F1, F2, and F3. Notably, K1, K2, and K3 signify convolution modules of diverse magnitudes, which are competent to achieve an equivalent receptive field to those equipped with convolution kernels of 7×7, 11×11, and 23×23. Subsequently, Fh is subjected to processing through three parallel convolution blocks and thereafter combined with the segregated low-frequency feature Fl. Upon the stochastic addition, the information pertaining to different scales is fused via 1×1 convolutions.

$$F_1 = K_1(F_h) + F_l \tag{3}$$

$$F_2 = K_2(F_h) + F_l \tag{4}$$

$$F_3 = K_3(F_h) + F_l \tag{5}$$

$$X = Conv_{1\times1}(F_1 + F_2 + F_3) \tag{6}$$

Subsequently, the feature maps are fed into the Simplified Coordinate Attention module (S-CA). Specifically, the input X is partitioned along the channel dimension into X1 and X2, thereby curtailing the computational complexity without

compromising on performance. The S-CA represents a streamlined variant of the Coordinate Attention (CA) mechanism [26], wherein the normalization and activation functions have been eliminated. Subsequent experimental validations have attested that the abrogation of the normalization and activation functions exerts a minimal, almost negligible influence on the functionality and efficacy of the S-CA module.

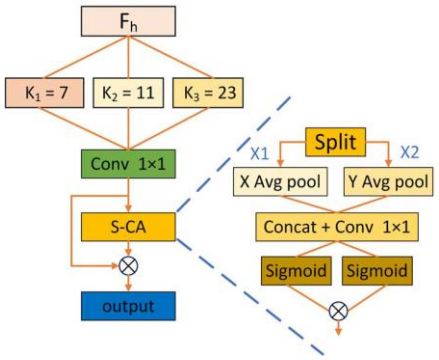

Figure 7. Multi-scale fusion architecture

4. Experimental Results and Analysis

The MLKnet model was trained on the public dataset NWPU-RESISC45 [27]. Each remote sensing image in the dataset has a pixel size of 256 × 256. This dataset contains 45 types of color remote sensing images, with 700 images of each type, totaling 31,500 images [28]. 29,610 images were selected as the training images. Gaussian noise with standard deviations of 15, 25, and 50 was added to the selected 29,610 training images. 1,000 selected images were used for testing, and 420 images from the UCMerced_LandUse dataset were also selected to test the network [29].

4.1. Network Training and Parameter Configuration

The experimental platform is based on the Ubuntu system. The Nvidia GeForce RTX3080Ti GPU is used for the training and testing of the model. The network employs Nvidia CUDA 11.8 and cuDNN 8.0 to accelerate the training speed of the GPU, and the network framework is built on the PyTorch platform. The initial learning rate for training the denoising model is [Here you didn't provide a specific value for the initial learning rate in your original text]. The momentum is 0.9, and the weight decay is 0.99. The batch size and the number of epochs is 4 and 40 respectively. The Adam optimizer is adopted. Other settings follow those of the NAFNet.

4.2. Model Comparison and Evaluation

Two widely recognized and objective image quality assessment metrics, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), were adopted to comprehensively and accurately measure the performance and effectiveness of the denoising algorithm. The level of the PSNR value directly reflects the magnitude of the difference at the pixel level between the image after denoising and the original image. Specifically, a higher PSNR value indicates that the denoising algorithm has removed noise more effectively while retaining the details of the image, thus achieving a better denoising effect. SSIM is used to evaluate the similarity in visual perception between the denoised image and the original image. When the SSIM value approaches 1, it means that the denoised image is not only close to the original image at the pixel level, but also maintains a high degree of consistency in visual structure and content, that is, the denoising effect is more natural and realistic.

4.2.1. Denoising of Gray Remote Sensing Images

Figure 8 illustrates a comparative visualization between a subset of noisy images and their denoised counterparts. The noise present in these images is genuine, and the grayscale renditions are derived through the conversion of color images. The source of the noisy images is the "Big Earth Data Science Data Center of the Chinese Academy of Sciences". Evidently, as can be discerned from the figure, the gray noisy images processed by the MLKNet network have attained a satisfactory denoising effect. Furthermore, the overall structural integrity of the images has been effectively maintained, and the restoration of detailed features has also been accomplished to a considerable extent.

Figure 8. Denoising rendering of MLKNet grayscale image

Table 1 exhibits the average outcomes of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) for the MLKnet put forward in this manuscript, in conjunction with those of K-SVD, BM3D, WNNM, DnCNN, ECNDNet, ADNet and RSIDNet on the gray remote sensing image dataset. The optimal results are accentuated in bold. As can be deduced from Table 1, under the three noise levels, the MLKnet network, there is an average improvement of 3.11 dB in the two datasets. Compared with the denoising results of color remote sensing images, the MLKnet has a greater improvement in the denoising effect on gray remote sensing images. Figure 9 shows the denoising comparison chart of multiple denoising methods on the UCMerced_LandUse dataset with a noise level of 25. It can be intuitively seen from the figure that the MLKnet demonstrates its capability in denoising gray remote sensing images, achieving a good denoising effect and retaining more detailed features.

Table 1. Average PSNR and SSIM Results of Different Denoising Methods on Grey Remote Sensing Image Dataset

Data set	method	σ = 15 PSNR/SSIM	σ = 25 PSNR/SSIM	σ = 50 PSNR/SSIM
NWPU-RESISC45	BM3D	31.52/0.9316	29.05/0.8862	25.82/0.7977
	K-SVD	29.42/0.8950	26.89/0.8146	22.59/0.6171
	WNNM	31.44/0.8509	29.38/0.8030	26.54/0.6972
	DnCNN	31.90/0.9345	29.51/0.8934	26.71/0.8158
	ADNet	31.83/0.9367	29.53/0.8990	26.71/0.8260
	ECNDNet	31.72/0.9363	29.36/0.8936	26.74/0.8273
	RSIDNet	31.94/0.9385	29.64/0.9007	26.82/0.8295
	MLKnet (our)	35.35/0.9451	31.98/0.8935	28.49/0.7961
UCMerced_LandUse	BM3D	31.31/0.9361	28.779/0.8935	25.43/0.8081
	K-SVD	29.31/0.9007	26.50/0.8193	22.06/0.6257
	WNNM	31.55/0.8822	28.99/0.8174	25.88/0.7047
	DnCNN	31.79/0.9422	29.28/0.9046	26.13/0.8289
	ADNet	31.64/0.9402	29.19/0.9041	26.19/0.8298
	ECNDNet	31.60/0.9394	29.08/0.8990	26.22/0.8314
	RSIDNet	31.84/0.9429	29.38/0.9065	26.34/0.8353
	MLKnet (our)	36.13/0.9498	33.03/0.9093	29.64/0.8361

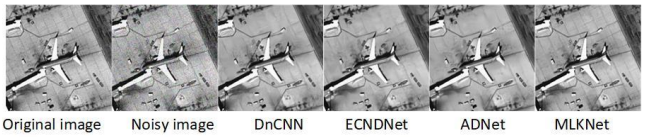

Figure 9. Comparison of denoising results of gray images with different denoising methods

4.2.2. Denoising of Color Remote Sensing Images

Figure 7 presents a comparison between a subset of noisy images and their corresponding denoised counterparts. The noise present in these images is of a realistic nature, and the source of the noisy images is the "Big Earth Data Science Data Center of the Chinese Academy of Sciences". It is conspicuously observable from the figure that the colored noisy images processed by the MLKNet network have attained a favorable denoising outcome. Furthermore, the overall structural integrity of the images has been effectively maintained, while the detailed features have been satisfactorily retained.

Figure 10. Denoising rendering of MLKNet color map

Table 2. Average PSRN and SSIM Results of Different Denoising Methods on Color Remote Sensing Image Dataset

Dataset	approach	σ = 15 PSNR/SSIM	σ = 25 PSNR/SSIM	σ = 50 PSNR/SSIM
NWPU-RESISC45	BM3D	33.95/0.9602	31.16/0.9277	27.23/0.8499
	K-SVD	31.05/0.9186	28.34/0.8776	24.68/0.7363
	WNNM	31.45/0.8508	29.35/0.8035	26.56/0.6974
	DnCNN	34.25/0.9631	31.59/0.9356	28.41/0.8777
	ADNet	34.14/0.9621	31.54/0.9347	28.40/0.8774
	ECNDNet	34.01/0.9602	31.36/0.9330	28.34/0.8755
	RSIDNet	34.28/0.9635	31.61/0.9360	28.49/0.8791
	MLKnet (our)	35.27/0.9614	31.93/0.9225	28.62/0.8509
UCMerced_LandUse	BM3D	33.22/0.9585	30.67/0.9299	27.05/0.8609
	K-SVD	30.85/0.9319	28.58/0.8867	24.46/0.7534
	WNNM	31.54/0.8820	29.95/0.8175	25.87/0.7052
	DnCNN	33.18/0.9602	30.79/0.9347	27.70/0.8774
	ADNet	32.99/0.9588	30.71/0.9338	27.70/0.8774
	ECNDNet	32.75/0.9572	30.42/0.9315	27.61/0.8762
	RSIDNet	33.26/0.9609	30.82/0.9358	27.83/0.8809
	MLKnet (our)	35.70/0.9675	32.52/0.9376	29.27/0.8844

Table 2 presents the average Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) results of the MLKnet proposed in this paper, along with those of K-SVD, BM3D, WNNM, DnCNN, ECNDNet, ADNet and RSIDNet on the color remote sensing image dataset. The best results are marked in bold. It can be seen from the table that, compared with other denoising methods, the MLKnet network demonstrates good performance in denoising ability. Under three different levels of noise, it achieves the best denoising results and has an average improvement of 1.17 dB on the two datasets. Figure 11 shows the denoising comparison chart of multiple denoising methods on the UCMerced_LandUse dataset with σ = 25. DnCNN fails to adequately preserve details. ECNDNet and ADNet preserve

the image structure but are unable to effectively eliminate fine noise. Meanwhile, the MLKnet retains more detailed features while achieving good denoising results.

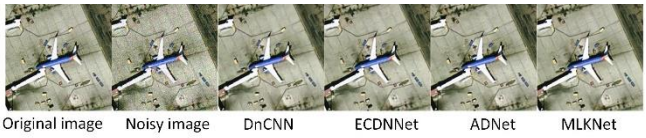

Figure 11. Comparison of denoising results of different denoising methods in color images

4.3. Ablation Experiments

To investigate the performance of individual modules within the network, ablation experiments were devised. The experimental setup involved testing on the NWPU-RESISC45 and UC-Merced_LandUse datasets under a color image noise level of 50. In the S-CA module, both the Batch Normalization (BatchNorm) and activation functions were removed. The experimental results are presented in Table 3 below. As can be seen from the table, eliminating the activation function and BatchNorm has a negligible impact on the results.

Table 3. Comparison of ablation results

Dataset	Activation +BN	PSNR	SSIM
RC45	√	28.62	0.8509
	×	28.60	0.8501
UC	√	29.27	0.8844
	×	29.23	0.8835

5. Conclusions

In this article, a method that incorporates frequency separation and separable large-kernel convolutions is proposed to remove Gaussian noise from remote sensing images. Compared with other networks, this method selects multiple large-kernel convolutions to extract multi-scale detail information, enabling the network to obtain broader context information. To reduce the massive computational load brought about by obtaining a larger receptive field, separable large-kernel convolutions are adopted. Instead of using Fourier transform to separate high and low frequency features, pooling is employed to obtain high-frequency features, and large-kernel convolutions are used to establish multi-scale high-frequency information. Compared with other methods, this method has higher computational efficiency and can effectively restore the detail information of images at the same time. Both the qualitative and quantitative results in the comparative study demonstrate the superiority of the remote sensing image denoising method proposed in this paper. It achieves an improvement of more than 1 dB in both gray remote sensing denoising and color remote sensing denoising.

References

[1] Wang, H. Y., Yang, H. T., Wang, J. Y., et al. (2024). A review of research on remote sensing image denoising methods [J]. Journal of Computer Engineering & Applications, 60(15)

[2] Li Y, Li X, Dai Y, et al. LSKNet: A Foundation Lightweight Backbone for Remote Sensing [J]. arxiv preprint arxiv: 2403. 11735, 2024.

[3] Dabov K, Foi A, Katkovnik V, et al. Image restoration by sparse 3D transform-domain collaborative filtering[C]//Image processing: algorithms and systems VI. SPIE, 2008, 6812: 62-73.

[4] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. IEEE Transactions on signal processing, 2006, 54(11): 4311-4322.

[5] Gu S, Zhang L, Zuo W, et al. Weighted nuclear norm minimization with application to image denoising[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 2862-2869.

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

[7] Jin, H. Z., Zhang, X. Y., Ye, Z. W., et al. (2022). Image denoising model based on approximate U-shaped network structure. Journal of Computer Applications, 42(8), 2571-2577.

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

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

[10] Tian C, Xu Y, Li Z, et al. Attention-guided CNN for image denoising [J]. Neural Networks, 2020, 124: 117-129.

[11] Han L, Zhao Y, Lv H, et al. Remote sensing image denoising based on deep and shallow feature fusion and attention mechanism [J]. Remote Sensing, 2022, 14(5): 1243.

[12] Alexey D. An image is worth 16x16 words: Transformers for image recognition at scale [J]. arxiv preprint arxiv: 2010.11929, 2020.

[13] Chen H, Wang Y, Guo T, et al. Pre-trained image processing transformer[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 12299-12310.

[14] Liang J, Cao J, Sun G, et al. Swinir: Image restoration using swin transformer[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 1833-1844.

[15] Wang Z, Cun X, Bao J, et al. Uformer: A general u-shaped transformer for image restoration[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 17683-17693.

[16] Tu Z, Talebi H, Zhang H, et al. Maxim: Multi-axis mlp for image processing[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5769-5780.

[17] Zamir S W, Arora A, Khan S, et al. Restormer: Efficient transformer for high-resolution image restoration[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022: 5728-5739.

[18] Zhang J, Zhang Y, Gu J, et al. Xformer: Hybrid x-shaped transformer for image denoising [J]. arxiv preprint arxiv: 2303. 06440, 2023.

[19] Ding, Y. W., Shi, H. B., Li, J., et al. (2024). Image denoising network based on local and global feature decoupling. Journal of Computer Applications, 44(8), 2571-2579.

[20] Lau K W, Po L M, Rehman Y A U. Large separable kernel attention: Rethinking the large kernel attention design in cnn [J]. Expert Systems with Applications, 2024, 236: 121352.

[21] Hao F, Wu J, Lu H, et al. Large coordinate kernel attention network for lightweight image super-resolution [J]. arxiv preprint arxiv: 2405. 09353, 2024.

[22] Guo M H, Lu C Z, Liu Z N, et al. Visual attention network [J]. Computational Visual Media, 2023, 9(4): 733-752.

[23] Chen L, Chu X, Zhang X, et al. Simple baselines for image restoration[C]//European conference on computer vision. Cham: Springer Nature Switzerland, 2022: 17-33.

[24] Yang K, Hu T, Dai K, et al. CRNet: A Detail-Preserving Network for Unified Image Restoration and Enhancement Task [J]. arxiv preprint arxiv: 2404. 14132, 2024.

[25] Ouyang D, He S, Zhang G, et al. Efficient multi-scale attention module with cross-spatial learning[C]//ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.

[26] Hou Q, Zhou D, Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 13713-13722.

[27] Cheng G, Han J, Lu X. Remote sensing image scene classification: Benchmark and state of the art [J]. Proceedings of the IEEE, 2017, 105(10): 1865-1883.

[28] Li, C., Li, X. T., Li, H. X., et al. (2024). A remote sensing image denoising method fused with multi-scale features. Electronics Optics & Control, 31(6), 74-80.

[29] Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification[C]//Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. 2010: 270-279.