A Denoising Network Based on Frequency-Spectral-Spatial-Feature for Hyperspectral Image

Siqi Wang, Liyuan Li, Xiaoyan Li, Jingwen Zhang, Lixing Zhao, Xiaofeng Su, and Fansheng Chen, *Senior Member, IEEE*

***Abstract*—The quality of hyperspectral images seriously impedes subsequent high-level vision tasks such as image segmentation, image encoding, and target detection. However, the frequency, spectral, and spatial properties of the hyperspectral noise pictures are not utilized fully by existing image denoising algorithms. To address this issue, a novel convolutional network based on united Octave and attention mechanism (UOANet) is proposed to extract the frequency-spectral-spatial-feature for denoising the actual noise of HSIs. In particular, the negative residual mapping embedded in Unet is proposed for multiscale abstract representation and two modules are designed for modeling global noisy HSI features in the frequency-spectral-spatial domain. First, with the use of residual Octave convolution module, our model can focus on the intrinsic properties of HSI noise distribution for desirable noise removal. Next, a parallel spatial-spectral attention module is used to fully utilize the rich spectrum data and the various spatial data of each band in HSI, which improves the richness of HSI details after denoising. Experimental results on both synthetic and real HSIs demonstrate the validity and superiority of UOANet compared with the state-of-the-arts under various noise settings.**

***Index Terms*—Frequency-spectral-spatial domain, hyperspectral image denoising, octave network, spatial-spectral attention mechanism.**

I. Introduction

COMPARED with the human eye, which can only observe information in the visible spectrum, the hyperspectral image can collect the signal of the whole electromagnetic spectrum, allowing researchers to obtain the spectral characteristics of various substances in the specific wave band and to analyze physical properties of the substance. Therefore, hyperspectral images are widely used in various fields, including ground object recognition [1], [2], [3], water retrieval [4], [5], and target tracking [6], [7]. However, hyperspectral images inevitably suffer from various corruptions and degenerations. Contaminated observations will seriously impede subsequent high-level vision tasks. As a result, it is of great importance to denoise HSIs before performing high-level tasks.

A. Related Work

This section briefly introduces the recent hyperspectral denoising works and the Octave-based approaches. Hyperspectral denoising has always been an ill-posed problem. Effective conditions that assist the denoising are required to address this issue, and various denoising approaches have been suggested to handle different types of noises. The existing methods can be coarsely divided into two categories: 1) Model-based approaches and 2) CNN-based approaches.

1) Model-Based Approaches: Most of the early hyperspectral remote sensing images are denoised by filtering techniques, which can be divided into spatial domain filtering and transform domain filtering. Spatial filtering is the most direct method for image denoising, and it works by combining the adjacent pixels in the window to achieve local smoothing. For example, Dabov et al. [8] proposed BM3D algorithm, which is used for three-dimensional (3-D) data denoising and can be directly used in hyperspectral image denoising. In order to enhance the denoising impact, Maggioni et al. [9] proposed BM4D, which extend BM3D to BM4D by employing 3-D cubes of voxels, and then stacks into a 4-D group and models bandwidth correlation by the joint processing of multidimensional image data. In addition, 1-D signal or 2-D image filtering method can be extended to denoise Hyperspectral Data Cube. Heo et al. [10] proposed a joint bilateral filter for hyperspectral image denoising. The bilateral filter and the fused image are applied to the hyperspectral image denoising after all bands of the picture have been weighted.

These filter-based approaches are simple and efficient. The key is the filter design and the selection of noise threshold. The periodic noise can be used to separate the signal components accurately. However, the mixed noise without obvious distribution features can easily lead to the spatial loss such as local

Manuscript received 11 April 2023; revised 30 May 2023; accepted 8 June 2023. Date of publication 13 June 2023; date of current version 27 July 2023. This work was supported in part by the Strategic Priority Research Program of the Shanghai Institute of Technical Physics, Chinese Academy of Sciences under Grant XDA19010102 and in part by the National Science Foundation of China under Grant 61975222. *(Corresponding author: Fansheng Chen.)*

Siqi Wang, Liyuan Li, and Jingwen Zhang are with the Key Laboratory of Intelligent Infrared Perception, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China, and also with the University of Chinese Academy of Sciences, Beijing 100049, China (e-mail: daqi23333@mail.ustc.edu.cn; liliyuan@mail.sitp.ac.cn; jingwen_zhang@yeah.net).

Xiaoyan Li is with the Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China (e-mail: lixiaoyan@ucas.ac.cn).

Lixing Zhao is with the Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China (e-mail: zhaolixing21@mails.ucas.ac.cn).

Xiaofeng Su is with the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China (e-mail: fishsu@mail.sitp.ac.cn).

Fansheng Chen is with the Key Laboratory of Intelligent Infrared Perception, Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China, and also with the Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China (e-mail: cfs@mail.sitp.ac.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3285454

oversmoothing and detail blurring. Additionally, both the filter design and the selection of noise threshold are dependent on expert priors, which cannot satisfy the demands of intelligent data processing.

Hyperspectral image has abundant space spectrum and space information, which can help to remove noise while preserving and restoring detail information. In recent years, many scholars have designed several effective regularizations for denoising, such as total variation (TV) based method [11], [12], [13], sparsity-based method [14], [15], [16], nonlocal-based method [17], [18], [19], dictionary-learning-based method [20], [21], [22], and matrix/tensor-based low rank method [23], [24], [25]. Among these approaches, ITS-Reg [26], LLRT [27] and a new iterative projection achieve state-of-the-art performance due to their extensive efforts on modeling intrinsic HSI properties.

In general, since the structural characteristics of HSI are utilized, these regularization-based approaches are more flexible and can be used for different types of noises. The key is the design of prior term and the selection of optimal regularization parameters. However, due to the limitation of artificial noise model and prior information, the generalization performance of single model is limited, making it difficult to apply to different types of noise pollution in real scene. What is more, it is challenging to fulfill the practical processing requirements of the massive high-dimensional hyperspectral remote sensing images.

2) Convolutional Neural Network (CNN) Based Approaches: With the development of deep learning technology, the CNN has recently seen widespread use in several low-level vision tasks of HSIs due to its excellent nonlinear fitting capability and automatic feature selection. In the latest research progress, many scholars are attempting to build an appropriate denoising network by training the network to learn the relationship between the model parameters and the noisy image implicitly. By doing this, effective noise reduction can be achieved without relying on manual constraints. Zhang et al. [28] proposed the 2-D image denoising architecture DnCNN to remove various noises in HSIs. They argued the learned filters can well extract the spatial structural information. Furthermore, utilizing residual learning based on DnCNN, Chang et al. [29] proposed an HSI-DeNet, which can remove many types of noises. There are also methods that include adding image gradient information to a network, such as the networks proposed by Maffei et al. [30] and Yuan et al. [31], which takes the spectrum data and image direction gradient information as the network input to remove noise. Recently, the attention mechanism plays a critical role in computer vision tasks. Many works have applied it to explore the correlation between the spatial and spectral properties in the field of HSI denoising. For example, Zou et al. [48] proposed an enhanced channel attention to make the network focus on features that are more conducive to spectral reconstruction. Wang et al. [49] applied the attention mechanism is used to select distinctive pixels in the feature maps for HSI denoising. Li et al. [50], [51] applied the vision transformer to capturing the nonlocal self-similarity of HSIs. On the other hand, given that deep learning is lack of interpretability, some scholars combine the model based and learning based models [52], [53], [54]. However, currently, these studies mainly focus on simple Gaussian noise situations, which are difficult to handle complex real noisy HSIs acquired by different sensors with varying numbers of bands.

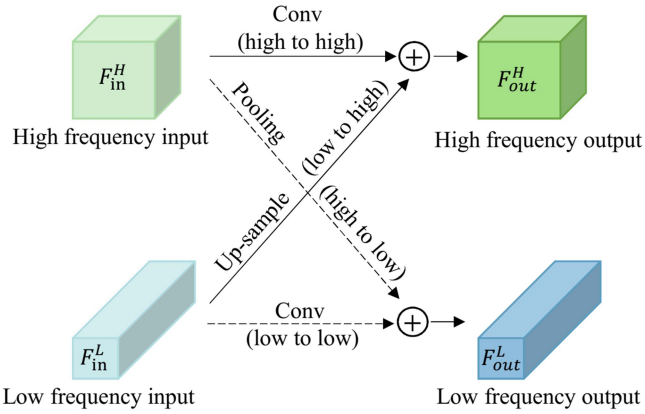

Fig. 1. Octave convolution kernel.

To sum it up, although many CNN- based methods have been developed for hyperspectral image denoising, most of these approaches rely heavily on a large amount of HIS data, resulting in low generalization and a significant amount of parameter redundancy. Therefore, fully mining the structural characteristics of real hyperspectral remote sensing image is an important task to improve the denoising effect of CNN-based approaches.

3) Octave-Based Approaches: Usually, a natural image can be defined as a discrete frequency signal, and the frequency distribution of noisy image y can be expressed as a combination of high-frequency information and low-frequency information, represented as $F(\mathrm{Y}) = \{F^H(\mathrm{Y}), F^L(\mathrm{Y})\}$. Based on the Octave convolution (Octave convolution) proposed by Chen et al. [37], we can separate the feature channels of an image by convolution, and get the high-and-low-frequency information of the image to obtain the frequency distribution of noise.

As shown in Fig. 1, in the Octave kernel, the ratio α represents the low-frequency proportion. The low-frequency features are represented by $\alpha \times$ c channels, whose spatial resolutions are decreased to 0.5 H $\times$ 0.5 W. The high-frequency features are represented by the $(1 - \alpha) \times$ c channels, whose spatial resolutions remain H $\times$ W.

In detail, because of the separation of the eigenvalues of the input and output, the Octave convolution weight W also needs to be separated, represented as$\{W^H, W^L\}$

$$W^H = \left[W^{H\to H}, W^{L\to H}\right]$$
$$W^L = \left[W^{H\to L}, W^{L\to L}\right]. \tag{1}$$

The different frequency feature vectors are fused. Specifically, $F^{H\to H}(\mathrm{Y})$, $F^{L\to L}(\mathrm{Y})$ means intra frequency forward propagation and $F^{L\to H}(\mathrm{Y})$, $F^{H\to L}(\mathrm{Y})$ means inter frequency forward propagation. The output characteristics of low and high frequencies are as follows:

$$\begin{aligned} F^H_{\text{out}} &= F^{H\to H}(\mathrm{Y}) + F^{L\to H}(\mathrm{Y}) \\ &= F_{\text{conv}}\left(F^H(\mathrm{Y}), W^{H\to H}\right) \end{aligned}$$

$$
\begin{aligned}
&+ \text{upsample}\left(F_{\text{conv}}\ \left(F^{L}\left(\mathrm{Y}\right),\ W^{L\to H}\right)\right)\\
F_{\text{out}}^{L} &= F^{H\to L}\left(\mathrm{Y}\right) + F^{L\to L}\left(\mathrm{Y}\right)\\
&= F_{\text{conv}}\ \left(\text{pool}\left(X^{H}, W^{H\to L}\right)\right) + F_{\text{conv}}\ \left(X^{L}, W^{L\to L}\right)
\end{aligned}
\tag{2}
$$

where upsample$(\cdot)$ represents the up-sampling operation using the nearest neighbor up-sampling method, pool$(\cdot)$ represents the down-sampling operation using the average pool method. Because of the ability to extract high-frequency features, Octave convolution has been applied in many HSI high-level tasks [55], [56], [57]. Following this line, we introduce this structure to our task for more effective HSIs denoising.

B. Motivations and Contributions

To overcome the shortcomings mentioned previously, this article proposes a denoising method based on the physical characteristics of hyperspectral noisy images to balance denoising performance and retain the noise-free component.

1) *Spectral-Feature:* HSI has a large number of spectral bands with strong correlation, which has an effect on denoising performance. However, noise randomly spread across different bands and lacks high spectral correlation. Therefore, how to capture the strong spectral correlation is a key issue for better denoising performance.
2) *Spatial-Feature*: In nature, most of the components in a picture are similar. Capturing the local and nonlocal spatial similarities benefit restoring the HSIs structural details. However, because of its random shape and distribution over different regions, noise lacks high spatial correlation. Therefore, it can be easily found that mining high correlation in the HSI spatial domain is of great significance for denoising.
3) *Frequency-Feature:* The high-frequency component is considered to contain more noise and texture details while the low-frequency component is considered to contain more content. This feature motivates us to focus the network on the high-frequency parts to better suppress noise and preserve the HSI content.

From above, the high correlation in the spectral domain and spatial domain, and also high frequency distribution are significant factors to be considered. Based on these feature priors, we propose UOANet to extract the frequency-spectral-spatial-feature for denoising the real noise of HSIs. Specifically, first, we introduce the parallel spatial-spectral attention mechanism to extract the high correlation information in spectral-spatial domain. Second, inspired by Chen's work [37], we introduce Octave convolution to separate the high-frequency and low-frequency information. This allows the network to focus on learning high-frequency noise information and minimizing the computation of solution space. Finally, we train the UOANet end-to-end to learn all of its parameters. The contributions of this article can generally be summed up as follows.

1) To better utilize the frequency-spectral-spatial-feature for HSIs denoising, we propose two noised image feature extraction modules, residual Octave convolution module (ResOct) and SSAT. A novel ResOct module is introduced in the encoding phase to extract high-frequency features, allowing the network to locate the noise information. Considering the spectral-spatial relationships between HSI pixels, in the decoding phase, an innovative spatial-spectral attention mechanism SSAT is proposed for noise feature learning, which fully captures the correlation information in feature maps.
2) An end-to-end denoising scheme is proposed based on the physical characteristics of hyperspectral noisy images, which considers the frequency of noise distribution, the spatial and spectral correlation of hyperspectral images. What is more, we use negative residual mapping to significantly reduce the mapping range, ensuring the generalization of the model.
3) Experimental results on both synthetic and real HSI datasets confirmed that our proposed model can achieve comparable or better performance compared with other state-of-the-art methods in the richness of image high-frequency details and model convergence.

The rest of this article is organized as follows. In Section II, we introduce the proposed network in detail. In Section III, we conduct a variety of experiments on synthetic and real HSI datasets. In Section IV, we prove the effectiveness of each module design. Finally, Section V concludes this article.

II. Proposed Model

An HSI is degraded by many factors during the imaging process. Therefore, it is necessary to improve the quality of hyperspectral imaging and increase the capacity for expression and information extraction. Image degradations caused by various mechanisms produce various types of noise. In this article, we will discuss additive and signal-independent noise (specifically, Gaussian noise, impulse noise, deadline noise, and stripe noise), which can be linearly modeled as

$$\mathrm{Y} = \mathrm{X} + \mathrm{N} \tag{3}$$

where $\mathrm{Y} \in R^{H\times W\times B}$ is the observed noisy HSI, $\mathrm{X} \in R^{H\times W\times B}$ is the clean HSI, $\mathrm{N} \in R^{H\times W\times B}$ is the addictive random noise. e H, W, B indicated the spatial height, spatial width, and the number of spectral bands, respectively.

Given a noisy HSI, our goal is to recover the clean HSI X from the observed noisy HSI Y. In this section, we introduce the overall network architecture of UOANet for HSI denoising, and then present the core building block in our network in detail.

A. Overall Network Architecture

The network takes Unet as the backbone and noisy HSIs as the input to predict clean HSIs. As shown in Fig. 2, UOANet uses Unet as the backbone, making better use of image context and location information than CNN, which is widely used in various low-level tasks [32], [33], [34], [35]. The network contains four encoding layers and three decoding layers. The left side of the network is the feature extraction network (encoder) to get the abstract semantic features. The right side of the network is the feature fusion network (decoder), which reconstructs the clean HSI image with the clean semantic features after denoising.

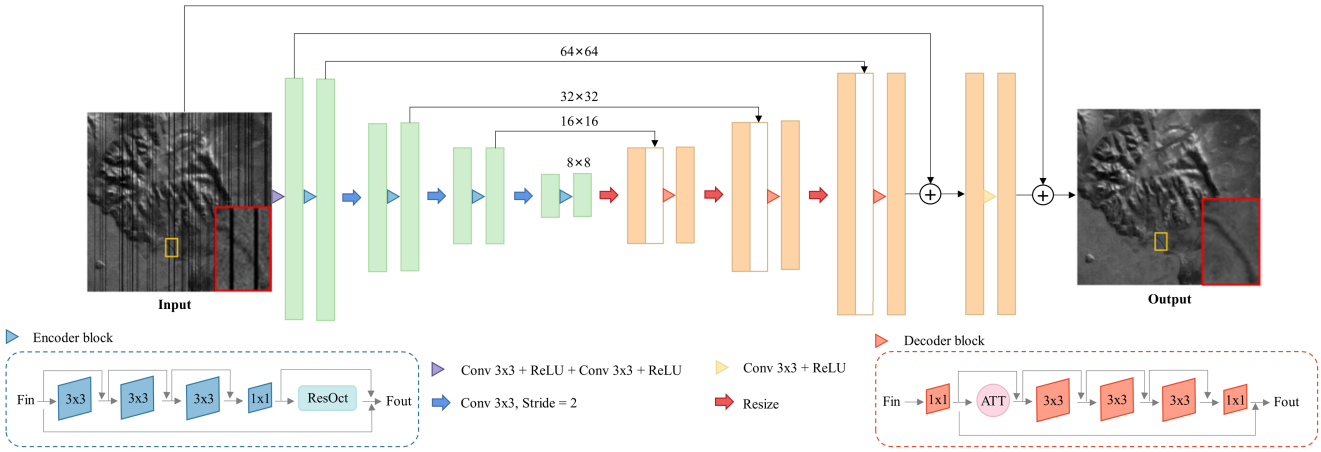

Fig. 2. Overall network structure of the UOANet.

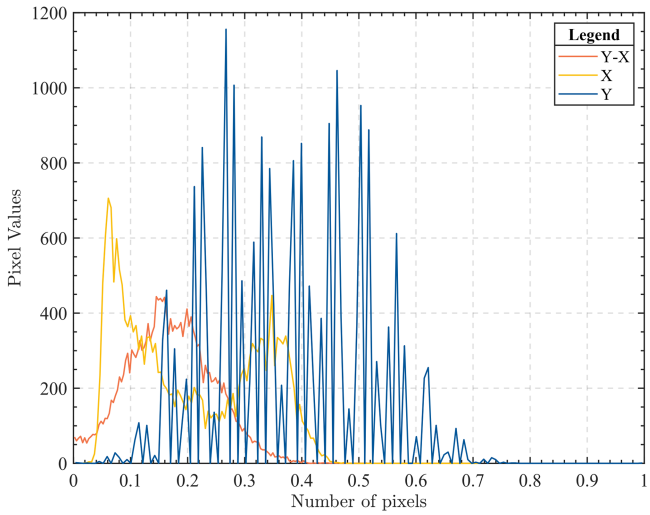

Fig. 3. Histogram distributions of Y,X,|Y-X|.

Specifically, we adopt the nearest neighbor interpolation for resize to help recover lost details during the encoder process. Symmetric skip connections are added in each layer to facilitate information transfer at various levels and promotes gradient back-propagation, which can help network training.

Compressing the mapping range is crucial for narrowing the solution space and enhancing network learning [36]. As shown in Fig. 3, taking Indian pines hyperspectral image as an example, noisy band as Y, clean band as X, we observe that when compared to the clean image Y, the residual of the rainy image Y-X has a significant range reduction in pixel values. This implies that the residual can be introduced into the network to aid with mapping learning. This skip connection can also directly propagate lossless information over the entire network, which is useful for estimating the final denoised image. In light of this idea, the mean squared error is employed as the loss function as follows:

$$\mathcal{L} = \sum_{\mathrm{i}} \left| \mathrm{H}\left(\mathrm{Y_i}\right) + \mathrm{Y_i} - \mathrm{X_i} \right|_{\mathrm{F}}^{2} \tag{4}$$

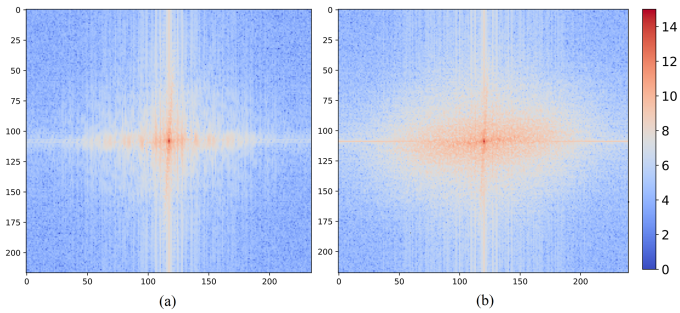

Fig. 4. Spectrum of noisy image pairs.

where a training group with N pairs $\{\mathrm{y_i}, \mathrm{s_i}, \mathrm{x_i}\}_{\mathrm{N}}$ of image data, $y_i$is the observed corrupted ith band data, and $x_i$is the corresponding ith noise-free data, $\mathrm{H}(\cdot)$ represents the UOANet.

B. Residual Octave Convolution Module

Inspired by Chen's work [37], we compare the noisy and clean HSI of the same scene to the spectrum in the frequency domain. As shown in Fig. 4, compared with the clean background spectrum, the spectrum of the noisy HSI diffuses energy from the high frequency part to the surroundings. Therefore, we can find that the noise frequency distribution F(N) mostly exist in high frequency information, and can be captured by the two-branch structure. Let $F^{H\to H}(\mathrm{Y})$ represents the noise frequency present in high-frequency component, and $F^{L\to H}(\mathrm{Y})$ represents the noise frequency present between high and low frequency component, the expression can be represented as

$$F\left(\mathrm{N}\right) = F^{H\to H}\ \left(\mathrm{Y}\right) + F^{L\to H}\left(\mathrm{Y}\right). \tag{5}$$

Bring (5) into (2)

$$\begin{aligned} F\left(\mathrm{N}\right) &= F_{\mathrm{out}}^{H} = F_{\mathrm{conv}}\ \left(F^{H}\left(\mathrm{Y}\right), W^{H\to H}\right) \\ &+ \ \mathrm{upsample}\ \left(F_{\mathrm{conv}}\ \left(F^{L}\left(\mathrm{Y}\right), W^{L\to H}\right)\right). \end{aligned} \tag{6}$$

As shown in Fig. 5(a), we introduce the ResOct in UOANet in encoding layers, enabling the network to locate the noise information. ResOct consists of residual blocks (RoctB) and

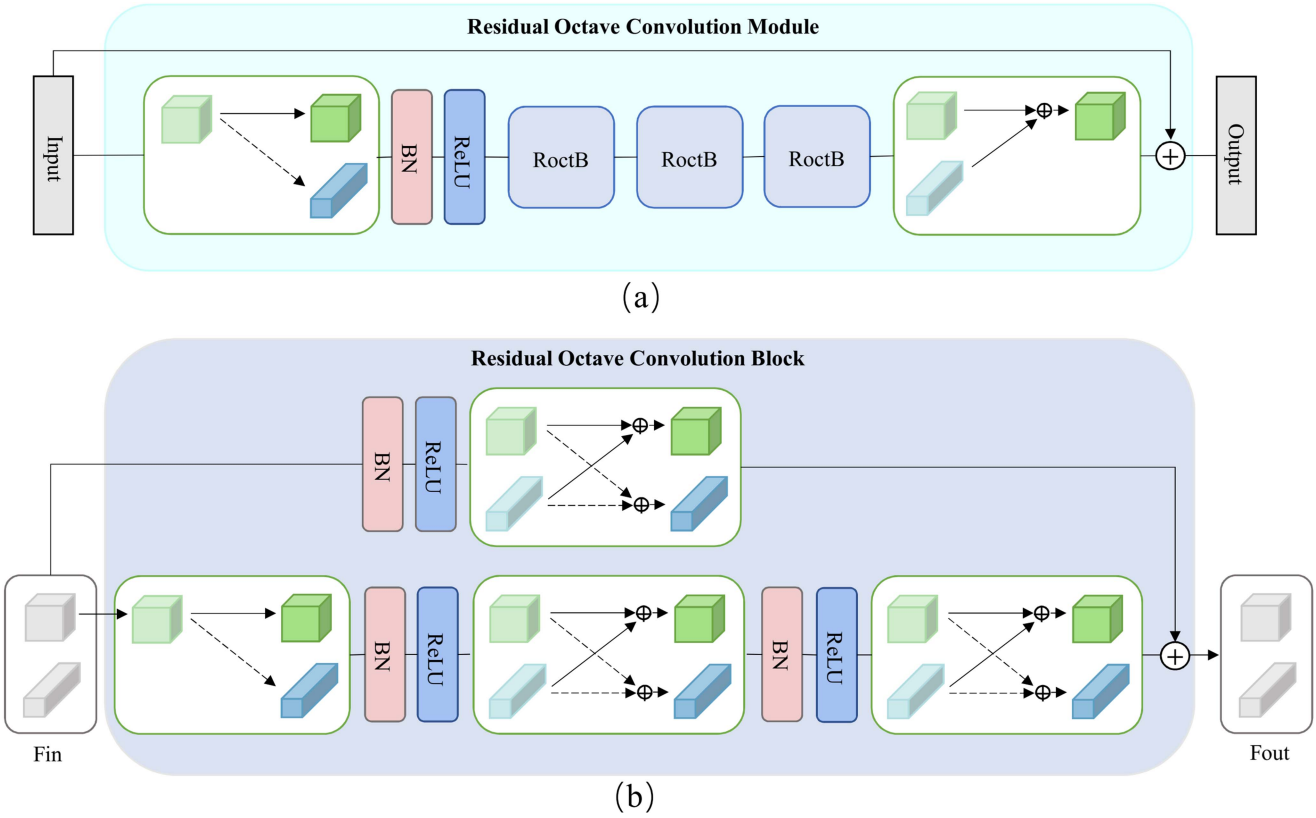

Fig. 5. Residual Octave convolution module (ResOct) and residual Octave convolution block (RoctB).

long-short jumping connections. First, the high and low frequency information is separated by an Octave convolution, and then activated by BN and Relu, respectively. After three series RoctB, the high and low frequency information is recombined by an inverse Octave convolution. Finally, the edge information of the original graph is fused to output through a long jump connection.

The ResOct structure combines Octave convolution with the residuals structure to enable cross-layer feature interaction, and extract deeper high-frequency feature semantic information while reducing the impact of low-frequency features. This structure preserves raw information and avoids the gradient, the network training process has some advantages. The structure optimizes the network training and effectively improves the denoising effect of hyperspectral image. The preservation of the raw information and avoidance of the gradient optimizes the network training process, effectively improving the effect of hyperspectral image denoising.

The structure of RoctB is shown in Fig. 5(b), the design of the jump link refers to the structure of RESNET50, when identity is mapped, instead of simply adding it, it passes through an Octave convolution of a BN layer and a RELU layer, then, the high and low frequency features extracted by Octave convolution are fused along the feature channel, which makes the model more easily converge and the training of the network more simple and efficient.

The operations of the ResOct are represented as

$$F_{\text{out}} = F_{in} + f_{\text{ResOct}}\left(F_{\text{in}}\right). \tag{7}$$

C. Parallel Spatial–Spectral Attention Module

As 3-D data, the hyperspectral image has characteristics of the spectral-spatial structure, global spectral correlation, and local/nonlocal spatial interactions. To model spatial and coherence spectral of the HSIs, attention mechanism is introduced for more detailed clean HSI restoration. First, we resize the encoded feature map, then splice it with the same-sized shallow encoded feature map along the channel, using a 1×1 convolution model the global context. Then, we design SSAT to adaptively recalibrate spatial, spectral, and channel characteristics. The SSAT, which adopts the ideas of CBMA [38] can weight the feature map to better align the reduction result with the physical properties of the HSIs.

The structure of SSAT is shown in Fig. 6, which consists of two parts: 1) Spatial attention module and 2) spectral attention module. Specially, since what is learned by later modules is affected by what has been processed by previous modules, regardless of the sequential sequence of spatial and channel attention [39], the model effect becomes unstable and it is impossible to ensure the correctness of effective promotion. Therefore, this article integrates spatial and spectral attention information simultaneously to avoid the interference of different

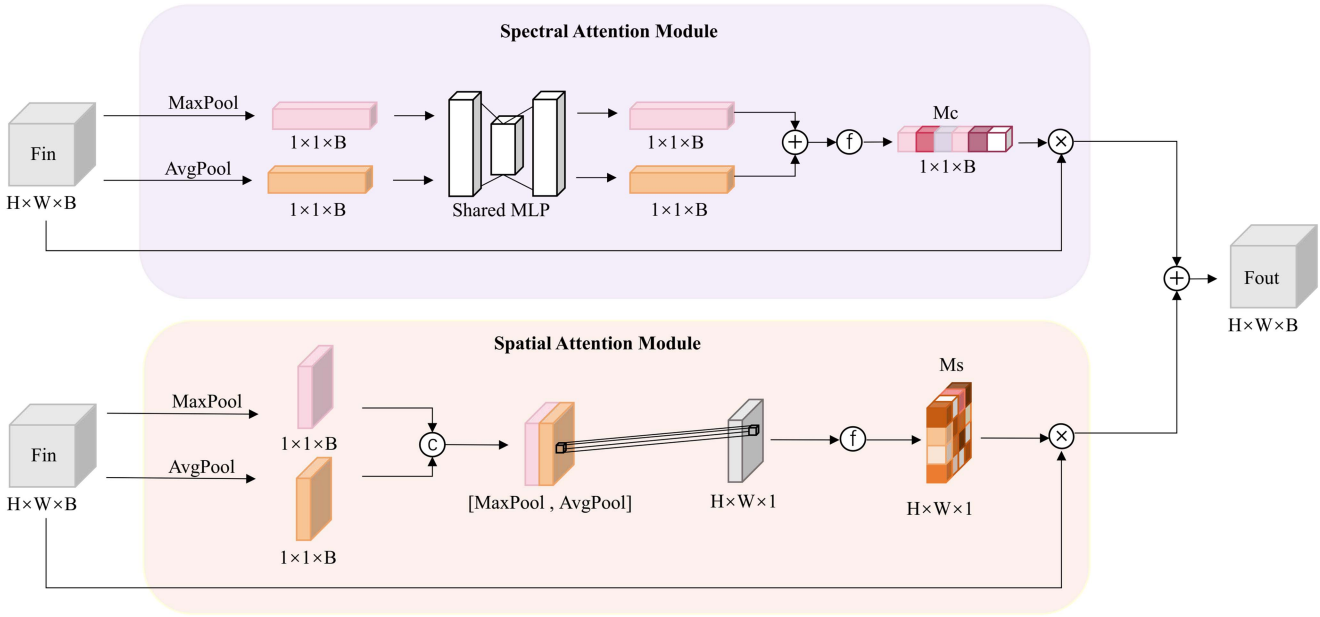

Fig. 6. Parallel spatial–spectral attention module (SSAT).

attention modules caused by serial structure. The operations of the ATT are represented as

$$\begin{aligned} F_1 &= M_c\ (F_{\text{in}}) \otimes F_{\text{in}} \\ F_2 &= M_s\ (F_{\text{in}}) \otimes F_{\text{in}} \\ F_{\text{out}} &= \lambda_1 F_1\ + \lambda_2 F_2 \end{aligned} \tag{8}$$

where $\lambda_1$and λ_2 are the weighting parameters.

1) Spatial Attention Module: Spatial attention focuses on the interspatial domain. First, average pooling and max pooling operations along the channel axis are used to generate the descriptors: $\mathrm{F}^{\mathrm{S}}_{\mathrm{avg}} \in R^{M\times N\times 1}$ and $\mathrm{F}^{\mathrm{S}}_{\mathrm{max}} \in R^{M\times N\times 1}$. Two descriptors are concatenated and fed into a vanilla convolution. The spatial attention process can be represented as

$$\mathrm{M_S}\ (\mathrm{F_{in}}) = \sigma\left(\mathrm{f}\left[\mathrm{F}^{\mathrm{S}}_{\mathrm{avg}}; \mathrm{F}^{\mathrm{S}}_{\mathrm{max}}\right]\right) \tag{9}$$

where σ is the Sigmoid activation function and f is a 2-D convolution with a 7×7 kernel.

2) Spectral Attention Module: This module combines feature maps along the spatial domain using average pooling and maximum pooling. They are forwarded by the channel attention module using a single hidden layer shared multilayer perceptron (MLP). The channel attention map M_C is created by merging the results of two branches. The process can be represented as

$$\mathrm{M_C}\ (\mathrm{F_{in}}) = \sigma\left(\mathrm{W_1}\left(\mathrm{W_0}\left(\mathrm{F}^{\mathrm{C}}_{\mathrm{avg}}\right)\right) + \mathrm{W_1}\left(\mathrm{W_0}\left(\mathrm{F}^{\mathrm{C}}_{\mathrm{max}}\right)\right)\right) \tag{10}$$

where σ is the Sigmoid activation function and W1 and W0 are the shared MLP parameters. $\mathrm{F}^{\mathrm{C}}_{\mathrm{avg}} \in R^{1\times 1\times B}$ and $\mathrm{F}^{\mathrm{C}}_{\mathrm{max}} \in R^{1\times 1\times B}$ are the features generated by average and max pooling operations in spatial domain, respectively.

III. Experiment Results

A. Training Experimental Settings

1) Training Data Set: We conduct several experiments using data from ICVL hyperspectral dataset [40], which comprise 201 images at a resolution of 1392 × 1300 over 31 spectral bands. We use 100 images for training, 1 image for validation, while 40 images are for testing. The training set is expanded by cropping the photos into 64 × 64 × 31 patch pairs. Each image is standardized into [0, [1] before conducting experiments.

2) Noise Setting: Figures that are composed of only black lines and shapes. These figures should have no shades or halftones of gray, only black and white.

Case 1: Non-i.i.d. Gaussian noise. Entries in all bands are contaminated by zero-mean Gaussian noise with different intensities, which are randomly chosen from 30 to 70.

Case 2: Gaussian + Stripe noise. As mentioned in case 1, each band is corrupted by non-i.i.d Gaussian noise. Besides, ten bands in the ICVL data set are randomly selected to add stripe noise, and the number of stripes in each band is 5% to 15% of columns.

Case 3: Gaussian + Deadline noise. Each band is contaminated by non-i.i.d Gaussian noise, as mentioned in case 1. Then, ten bands in the ICVL data set are chosen randomly to add deadline noise. The number of deadlines in each band is 5% to 15% of columns.

Case 4: Gaussian + Impulse noise. All bands are corrupted by Gaussian noise as mentioned in case 1. Then, ten bands in ICVL data set are randomly selected to add impulse noise with different intensities, and the percentage of impluse ranges from 10% to 70%.

Case 5: Mixture noise. First, all bands are corrupted by Gaussian noise as previously mentioned. Then each band

is randomly corrupted by at least one kind of the other four noise mentioned previously.

3) Competing Methods: In Gaussian noise case, we compare with several representative traditional methods including filtering-based approaches (BM4D [9]), dictionary learning approach (TDL [20]), and tensor-based approaches (ITSReg [26], LLRT [27]).

In complex noise case and real-world noise case, the competing traditional baselines include filtering-based approaches (BM4D [9], BWBM3D [10]), and low-rank matrix recovery approaches (LRMR [44], LRTV [45]). For DL approaches, we compare our model with DnCNN [28], HSI-SDeCNN [30], HSID-CNN [31], QRNN3D [47], SST [50], SERT [51], MAC-Net [52], and T3SC [53].

4) Evaluation Indexes: In order to evaluate the denoising performance of simulated experiments in both the spatial domain and spectral domain, three quantitative criteria are introduced as follows. Smaller SAM and larger PSNR and SSIM imply better denoising.

Mean peak signal-to-noise ratio (MPSNR) [41]

$$\mathrm{MPSNR} = \frac{1}{B}\sum_{k=1}^{B} 10\lg\frac{M\times N\times A^2}{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(x\left(i,j,k\right)-y\left(i,j,k\right)\right)^2} \tag{11}$$

where M, N, and B represent the HSIs width, height, and number of bands, respectively. A is the maximum value of all the gray values. $y(i,j,k)$ represents the original clean image, while $x(i,j,k)$ represents the approximated image.

Mean structural similarity index (MSSIM) [42]

$$\mathrm{MSSIM} = \frac{1}{\mathrm{B}}\sum_{\mathrm{i}=1}^{\mathrm{B}}\frac{\left(2\mu_{\mathrm{x_i}}\mu_{\mathrm{y_i}}+\mathrm{C}_1\right)\left(2\sigma_{\mathrm{x_iy_i}}+\mathrm{C}_2\right)}{\left(\mu_{\mathrm{x_i}}^2+\mu_{\mathrm{y_i}}^2+\mathrm{C}_1\right)\left(\sigma_{\mathrm{x_i}}^2+\sigma_{\mathrm{y_i}}^2+\mathrm{C}_2\right)} \tag{12}$$

where $\mu_{\mathrm{x_i}}$ and $\mu_{\mathrm{y_i}}$, stand for the mean values of the ith estimated and original clean image, respectively, $\sigma_{\mathrm{x_i}}^2$ and $\sigma_{\mathrm{y_i}}^2$ are the variances, $\sigma_{\mathrm{x_iy_i}}$ is the covariance, and C_1 and C_2 are constants that prevent the denominator from being 0. The mean value of SSIM of each band is adopted to assess the structural similarity of the whole.

Mean spectral angle mapper (MSAM) [43]

$$\mathrm{MSAM} = \frac{1}{\mathrm{MN}}\cos^{-1}\left[\frac{\sum_{\mathrm{i}=1}^{\mathrm{MN}}\mathrm{t_i}\mathrm{p_i}}{\sqrt{\left(\sum_{\mathrm{i}=1}^{\mathrm{MN}}\mathrm{t_i^2}\right)\left(\sum_{\mathrm{i}=1}^{\mathrm{MN}}\mathrm{p_i^2}\right)}}\right] \tag{13}$$

where $\mathrm{t_i}$ denotes the estimated spectrum and $\mathrm{p_i}$ is the original spectrum. This metric is adopted to assess the spectral fidelity of denoising algorithms.

5) Implementation Detail: We adopted the incremental learning method to stabilize and accelerate the training, which also avoids the network converging to a poor local minimum.

Hyperparameter values were empirically set to make network learning fast yet stable. Small batch size (i.e., 16) is used to accelerate training at first stage, while large batch size (i.e., 64) is adopted to stabilize training when tackling harder cases (e.g., complex noise case). The overview of our training procedures is shown in Table I, with detailed hyperparameter setting. We used the Adam algorithm as the optimizer.

All experiments were performed on a PC with an Intel(R) Xeon(R) Gold 5218R CPU, and an NVIDIA 2080Ti GPU. A quantitative and qualitative analysis has been conducted for both simulated and real data.

B. Experiment on ICVL HSIs

1) Testing Data Set: We design two different scenarios to verify and evaluate the denoising performance of UOANet.

2) Results of ICVL HSIs: In Gaussian Noise case, Table II shows the index values of MPSNR, MSSIM, and MSAM after the proposed algorithm and seven other contrast algorithms are denoised. As can be seen from the table, our UOANet algorithm can achieve the best or the second best index in most bands, because our method fully considers the spatial-spectral correlation of noisy HSI. In addition, UOANet uses octave to preserve low-frequency information and denoise high-frequency information by convolution. It can be easily observed from Fig. 7, our method can better remove the noise and retain the details. At the same time, in addition, the PSNR values for each band in Fig. 7 are shown in Fig. 9(a) and (b), from which it can be observed that our method achieves a higher PSNR in almost all bands compared with other methods.

In complex noise case, denoising quantization results are shown In Table III. From Table III, we can see that our method achieves significantly better denoising results than some of the most advanced methods, such as LRMR, LRTV, because these methods are based on low-rank matrices, and some structural information is lost in the process of denoising. Compared with two methods based on depth learning (DNCNN and HSID-CNN), our method can explore spectral-spatial information and suppress noise thanks to the SSAT attention module. From Fig. 8, we can observe that although the competing methods LRTV, LRMR, and HSID-CNN and can obtain cleaner denoising results, the denoising image still contains some noise or structural information that is not well preserved. In contrast, our method can not only remove the complex noise well, but also preserve the structure and details better, so as to obtain better visual reconstruction results. Furthermore, we show a PSNR value for each band in Fig. 9(c) and (d), from which we can observe that our method can achieve a higher PSNR in almost all bands compared with other competing methods. And the spectral curves of pixels (130,74 s) in Case 5 are plotted in Fig. 10. It shows that compared to other methods, we are also closer to the ground truth value. UOANet can reconstruct HSIs with higher quality in both spatial and spectral domains.

C. Experiment on Remote Sensing Images With Simulated Noise

1) Testing Data Set: The main motivation of proposing UOANet was to improve the generalization ability of the model, besides experiments with close-range images, such as ICVL

TABLE I
OVERVIEW OF OUR INCREMENTAL TRAIN POLICY

Stage	1	2	3		
Noise model	Gaussian noise with known σ	Gaussian noise with unknown σ	Complex noise		
Epoch	0~100	100~200	200~250	250~281	281~300
Learning rate	10-3	10-3	10-3	5×10-4	10-4
Batch size	16	16	64		

TABLE II
QUANTITATIVE RESULTS OF DIFFERENT METHODS IN GAUSSIAN NOISE CASE ON ICVL DATASET

Index	Noisy	BM4D [9]	TDL [20]	ITSReg [26]	LLRT [27]	DnCNN [28]	HSID-CNN [31]	Ours
σ=30								
MPSNR	18.588	38.996	41.133	41.780	42.378	30.03	37.534	**43.643**
MSSIM	0.093	0.941	0.962	0.966	0.970	0.895	0.931	**0.984**
MSAM	0.864	0.121	0.063	0.073	**0.059**	0.176	0.143	0.080
σ=50								
MPSNR	14.151	36.231	38.801	39.303	39.352	30.671	35.982	**41.144**
MSSIM	0.036	0.902	0.940	0.948	0.947	0.879	0.898	**0.982**
MSAM	1.043	0.162	0.078	0.079	0.082	0.201	0.172	0.080
σ=70								
MPSNR	11.229	34.424	37.176	37.614	37.699	31.799	33.659	**38.330**
MSSIM	0.018	0.865	0.918	0.934	0.932	0.797	0.849	**0.971**
MSAM	1.149	0.196	0.096	0.086	0.096	0.235	0.200	**0.085**
blind								
MPSNR	15.302	37.650	40.249	40.636	40.832	31.272	36.541	**40.975**
MSSIM	0.046	0.916	0.950	0.955	0.957	0.851	0.903	**0.981**
MSAM	1.089	0.171	0.082	0.082	0.081	0.261	0.189	**0.081**

Bold represents the best performance.

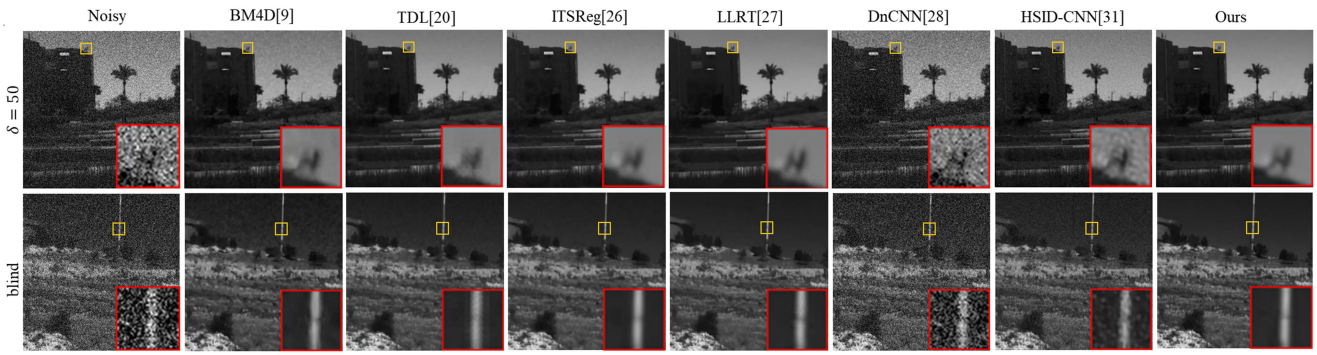

Fig. 7. Simulated Gaussian noise removal results on ICVL dataset.

HSIs. We also ran all the competing methods on remote sensing images, such as SDG images acquired by the SDGSAT-1 satellite. Compared with ICVL HSIs, SDG images has a higher spectral resolution 456 × 444, but much lower spatial resolution, which contains with a spatial resolution of 30 m per pixel. Therefore, experiments on real data of different satellites and loads were conducted to verify the generalization ability. We design three different scenarios to verify and evaluate the denoising performance of UOANet.

2) Results of SDG Images: In blind Gaussian noise case, from Fig. 11, we can observe that although BM4D, TDL, ITSReg, and LLRT can obtain cleaner denoising results, the structural information of the denoised image is not well preserved, resulting in oversmoothing. However, DNCNN and HSID-CNN still contain more noise, because they are changed to the scene, which is different from the training set, which shows that the generalization ability of the model is weak. Our method performs better in detail maintenance, noise removal, and model generalization.

In mixture complex noise case, from Fig. 12, we can observe that although LRMR and LRTV can obtain cleaner denoising results for dead-line noise, the LRMR denoising image still

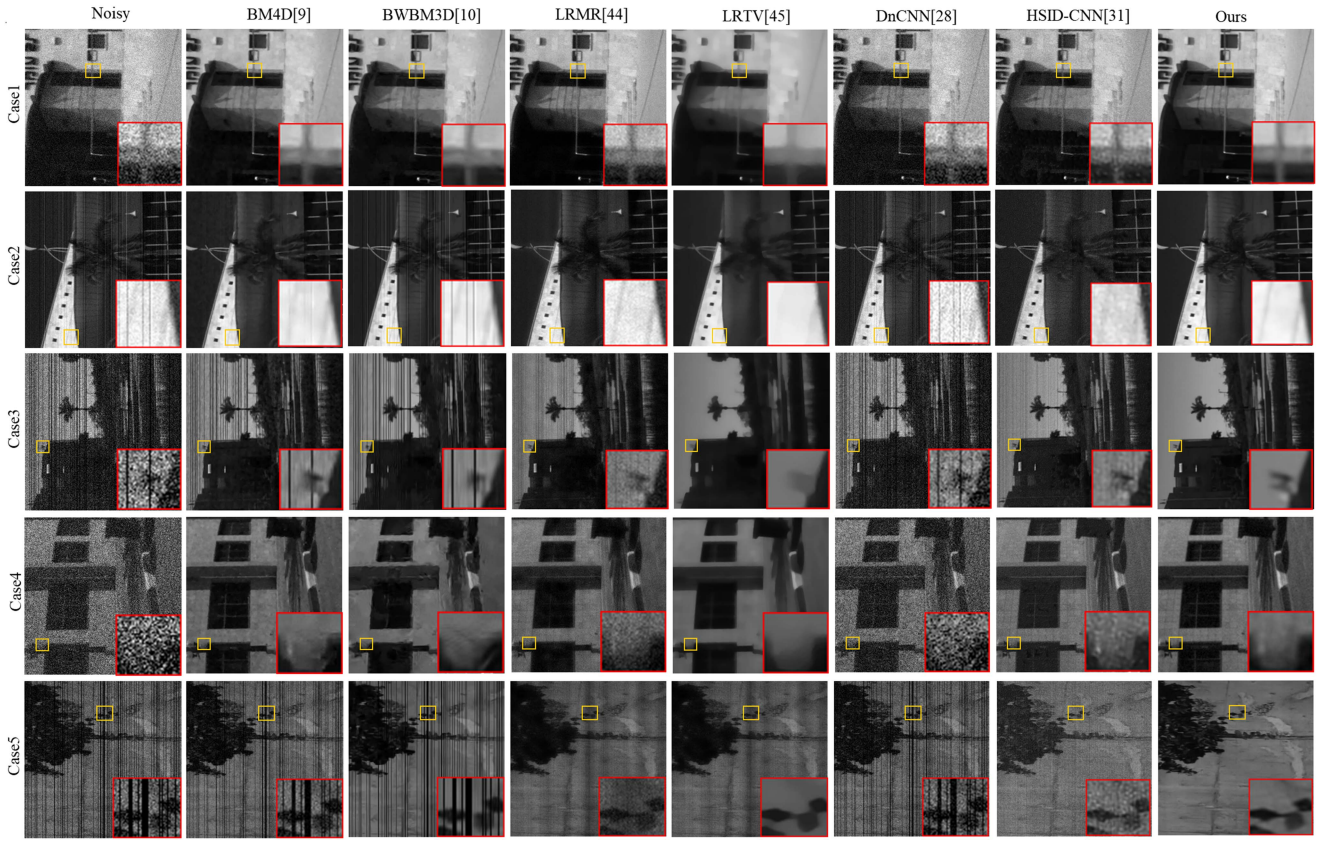

Fig. 8. Simulated complex noise removal results on ICVL dataset.

TABLE III
QUANTITATIVE RESULTS OF DIFFERENT METHODS IN COMPLEX NOISE CASE ON ICVL DATASET

Index	Noisy	BM4D [9]	BWBM3D [10]	LRMR [44]	LRTV [45]	DnCNN [28]	HSID-CNN [31]	Ours
Case1								
MPSNR	17.714	37.391	37.038	29.189	35.227	29.234	32.901	**40.585**
MSSIM	0.131	0.896	0.905	0.524	0.921	0.786	0.857	**0.984**
MSAM	1.100	0.203	0.210	0.604	0.097	0.293	0.231	**0.072**
Case2								
MPSNR	18.066	36.817	35.176	29.310	35.557	29.767	30.439	**40.212**
MSSIM	0.140	0.880	0.861	0.535	0.924	0.801	0.892	**0.983**
MSAM	1.088	0.221	0.279	0.606	0.094	0.309	0.319	**0.076**
Case3								
MPSNR	18.322	35.382	34.863	28.948	34.642	29.023	31.464	**40.476**
MSSIM	0.149	0.869	0.867	0.529	0.912	0.759	0.85	**0.985**
MSAM	1.093	0.227	0.252	0.62	0.132	0.310	0.345	**0.077**
Case4								
MPSNR	14.406	26.275	28.919	24.981	33.229	21.791	25.312	**34.300**
MSSIM	0.083	0.502	0.658	0.425	0.879	0.481	0.501	**0.906**
MSAM	1.064	0.699	0.834	0.714	0.358	0.753	0.703	**0.200**
Case5								
MPSNR	14.029	22.679	23.461	24.459	31.633	20.9877	24.431	**35.617**
MSSIM	0.081	0.394	0.543	0.432	0.842	0.3528	0.448	**0.941**
MSAM	1.066	0.812	0.890	0.748	0.481	0.9786	0.723	**0.165**

Bold represents the best performance.

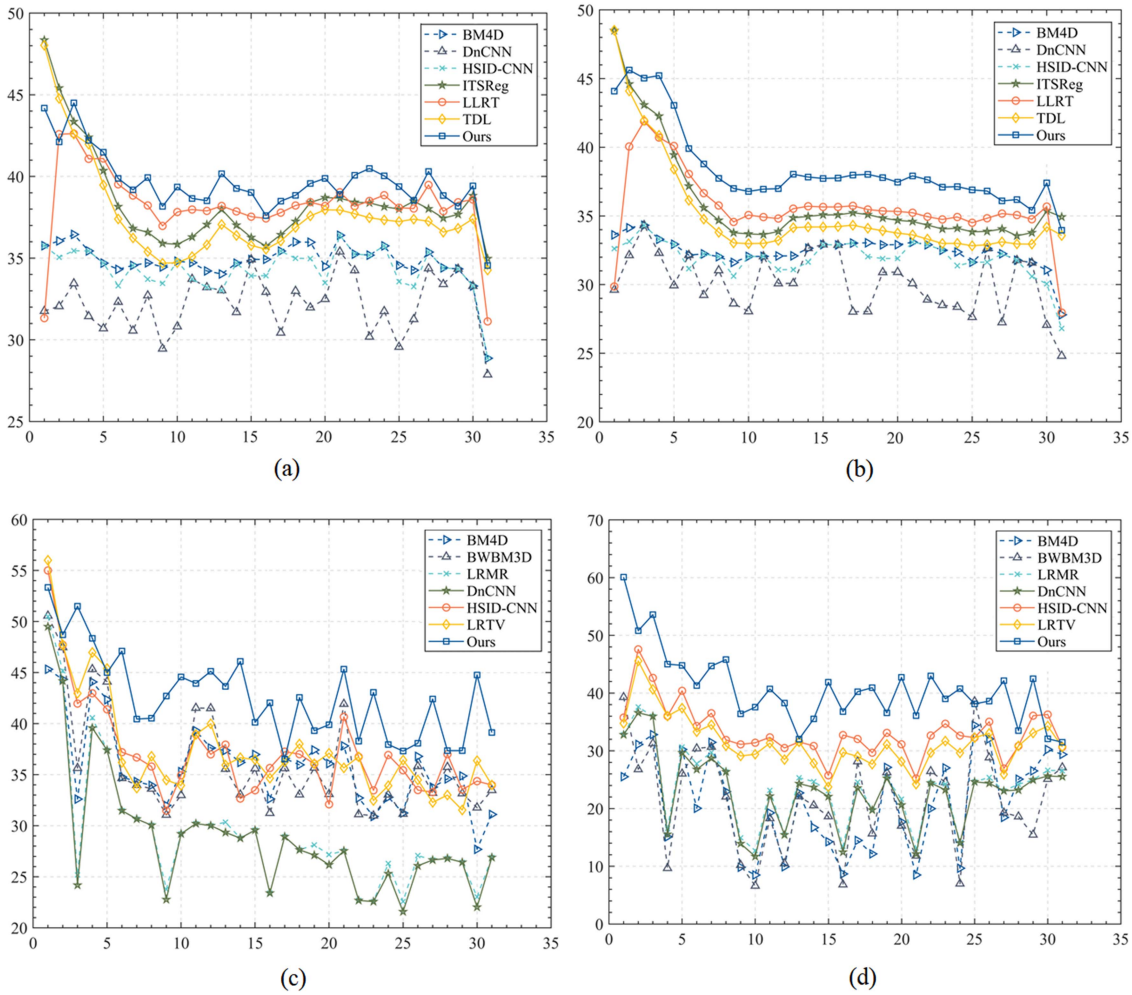

Fig. 9. (a), (b) PSNR values of each band corresponding to Gaussian noise removal results in Fig. 7, and (c), (d) PSNR values of each band corresponding to complex noise removal results in Fig. 8. The ordinate represents the PSNR value, and the abscissa represents the band number.

retains some Gauss Speckle noise, the structure information of LRTV is not well preserved, so it is too smooth. Benefiting from the ability to integrate spatial context information and interchannel dependencies, our method can preserve the structure details better while finely remove the complex noise.

In particular, we used the average of all the pixels in each band to evaluate the effect of denoising. Figs. 13 and 14 show the longitudinal averages of the SDG images of Scene 1 and Scene2 before and after denoising, respectively. It can be seen from the graph that the curve of the original graph of noise has a sharp fluctuation, which indicates that the image contains banded noise. Compared with other contrast algorithms, the curve of our algorithm is smoother, which shows that our algorithm can better remove noise.

D. Experiment on HSIs With Real-World Noise

1) Testing Data Set: In this article, we evaluate our model on remotely sensed hyperspectral datasets, including EO-01 data and Indian Pines data. All of them have been used for real HSI denoising experiments [44], [45], [46]. EO-01 data are captured through the Hyperion sensor with size $400 \times 1000 \times 242$ and are mainly degraded by stripe, deadline, and Gaussian mixed noise. For simplicity, we select EO-01 sub image with a size of $240 \times 240 \times 31$. The Indian Pines are captured through the AVIRIS with size contains $145 \times 145 \times 220$ with a spatial resolution of 20 m per pixel, and some bands are seriously polluted by the atmosphere and water, as well as degraded by stripe, deadline, and Gaussian mixed noise, making it difficult to remove this noise.

2) Results of Real Noisy Images: For EO-01 data, it can be observed in Fig. 15 that scene was affected by striping noise and deadlines. It can be seen from the results that the visual effect of BM4D, BWBMD processing is not good: only a small amount of stripe noise is slightly suppressed and many obvious stripes remain. The LRMR, HSID-CNN, and DnCNN method generally remove the stripes, but a few stripes are not removed locally. After LRTV, 3DQRNN, MACNet, T3SC, SST, and SERT processing, some of the stripe interference can be removed, but the restored image is excessively smoothed due to the missing of texture information. Among these methods, our method produces the best denoising results for the restored

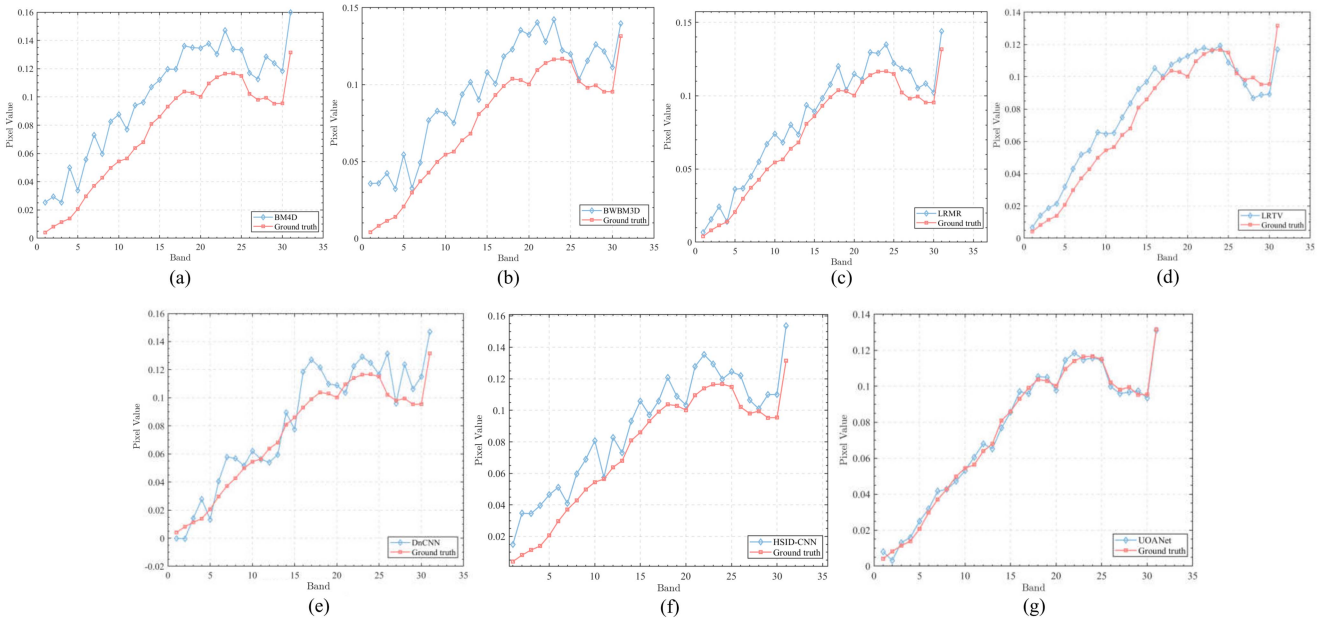

Fig. 10. Spectral reflectance of pixel (130, 74) for each band in Case5. The ordinate represents the reflectance value, and the abscissa represents the band number. (a) BM4D. (b) BWBM3D. (c) LRMR. (d) LRTV. (e) DnCNN. (f) HSID-CNN. (g) Ours.

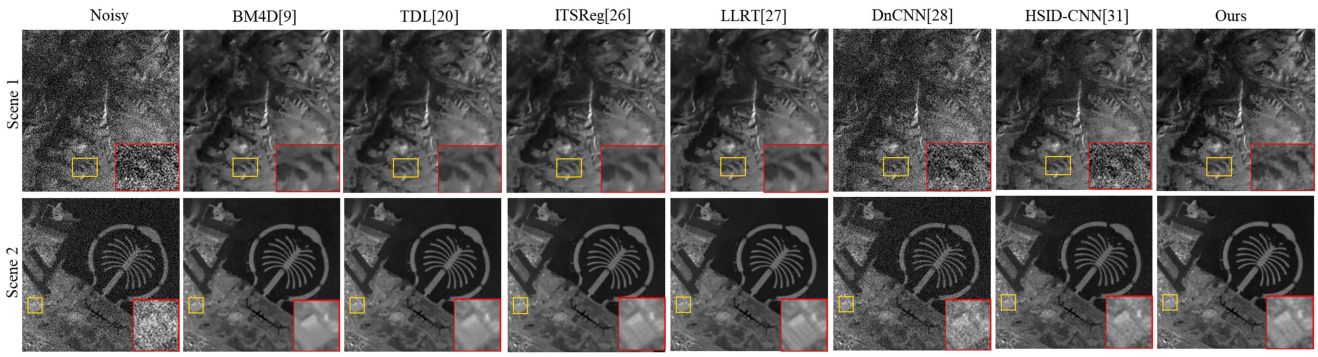

Fig. 11. Simulated noise removal results at image in blind Gaussian noise on SDG dataset.

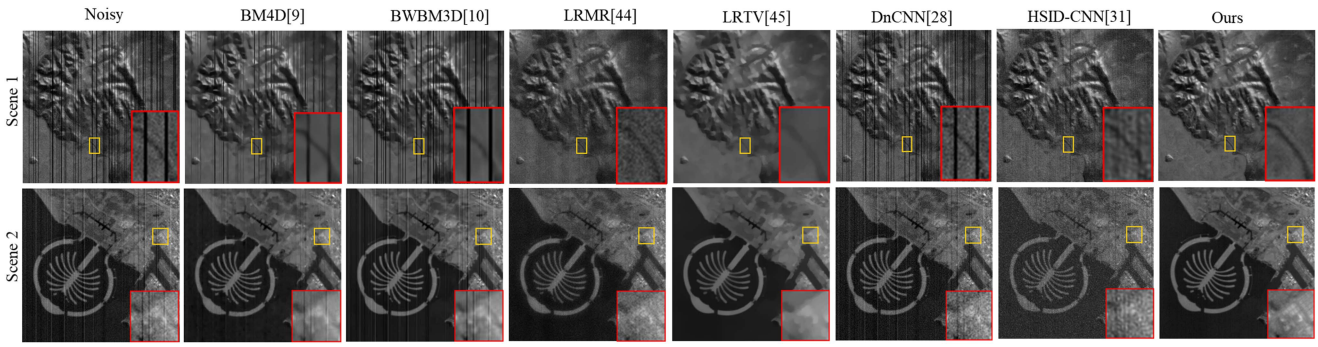

Fig. 12. Simulated noise removal results at image in mixture complex noise on SDG dataset.

image retains the original structure features. Fig. 15 depicts the spectrum of denoising and noise HSI at position (152,82). As can be observed in Fig. 16, DNCNN, HSIDCNN, LRTV, 3DQRNN and UOANet provide optimal denoising, but UOANet has the best spectral fidelity.

For Indian Pines data, it can be observed in Fig. 17 that terrible atmosphere and water absorption obstruct the view to the real scenario, severely degrading the quality of images. The Gaussian denoising methods, such as BM4D, BWBM3D, and DnCNN cannot accurately estimate the underlying clean image

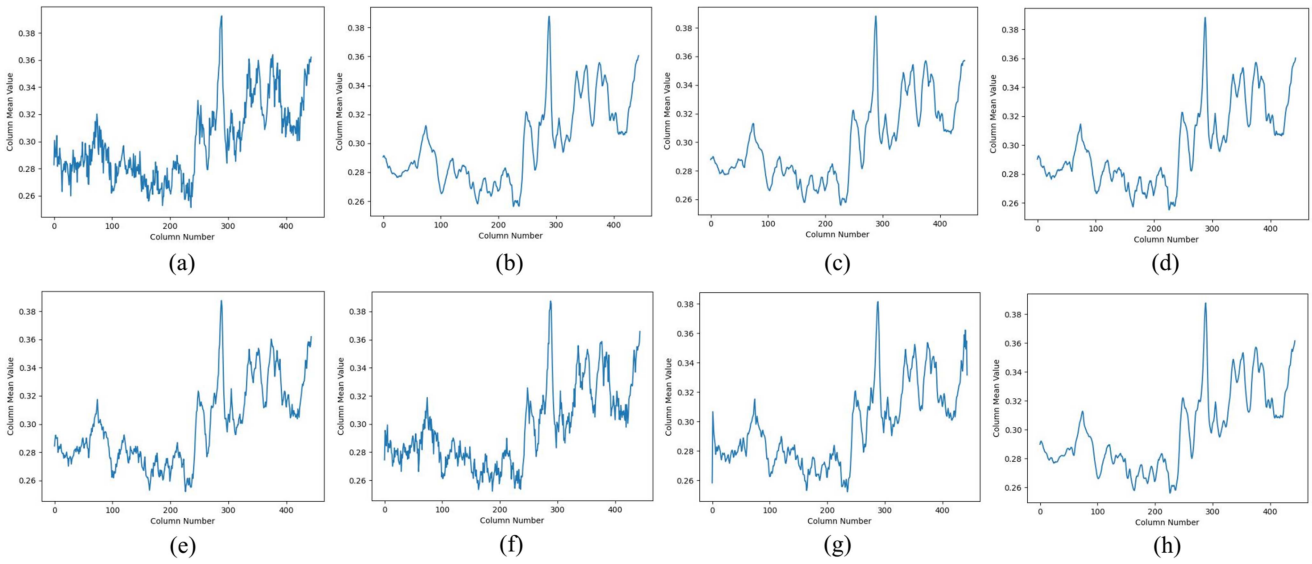

Fig. 13. Column mean values at 2th band of Fig. 11 Scene 1 image. The ordinate represents the column mean value, and the abscissa represents the column number. (a) Noisy. (b) BM4D. (c) TDL. (d) ITSReg. (e) LLRT. (f) DnCNN. (g) HSID-CNN. (h) Ours.

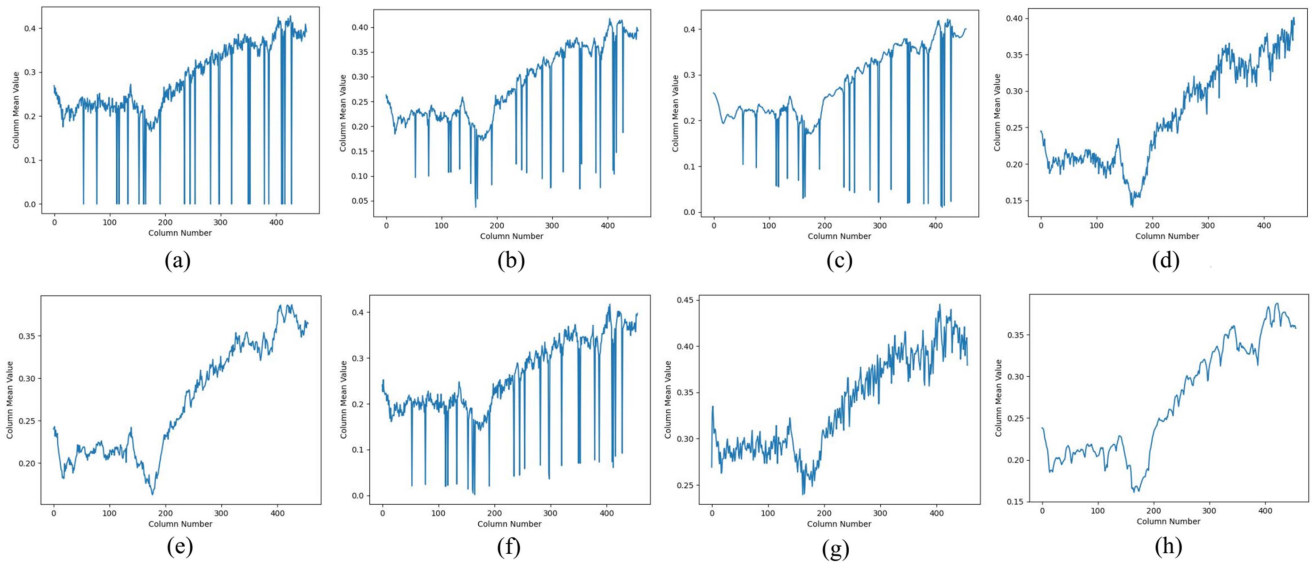

Fig. 14. Column mean values at 1th band of Fig. 13 Scene 1 image. The ordinate represents the column mean value, and the abscissa represents the column number. (a) Noisy. (b) BM4D. (c) BWBM3D. (d) LRMR. (e) LRTV. (f) DnCNN. (g) HSID-CNN. (h) Ours.

due to the non-Gaussian noise structure. The LRMR, LRTV, HSID-CNN, HSI-SDeCNN, MACNet, T3SC, SST, and SERT method generally removes the noises, but a few noises are not removed locally. Our method successfully tackles this unknown noise, and produces sharper and clearer result than others. Specifically, to comprehensively compare the denoising effect, we also show false color images of these constructed results of the Indian Pines (band 144, 154, and 164) in Fig. 18. It can be easily seen that other competing methods still exists much dense noise in the restored bands, while our proposed method can almost remove the most complex noise. The spectral reflectance of the pixels (103, 64) is plotted in Fig. 19 , and it can be seen that all methods provide very similar spectra in real visual perception. But our method is more complete in preserving the curve details of the spectrum and achieves the clear restored bands.

IV. Discussion

A. *Effectiveness of the ResOct and SSAT Module*

In this section, we examine the effectiveness of ResOct and SSAT on the denoising performance. Table V presents the denoising results of different module settings. Meanwhile, the sensitivity of attention module weighting parameters λ_1 and λ_2 are discussed. Fig. 20 presents the visualizations of the SSAT modules.

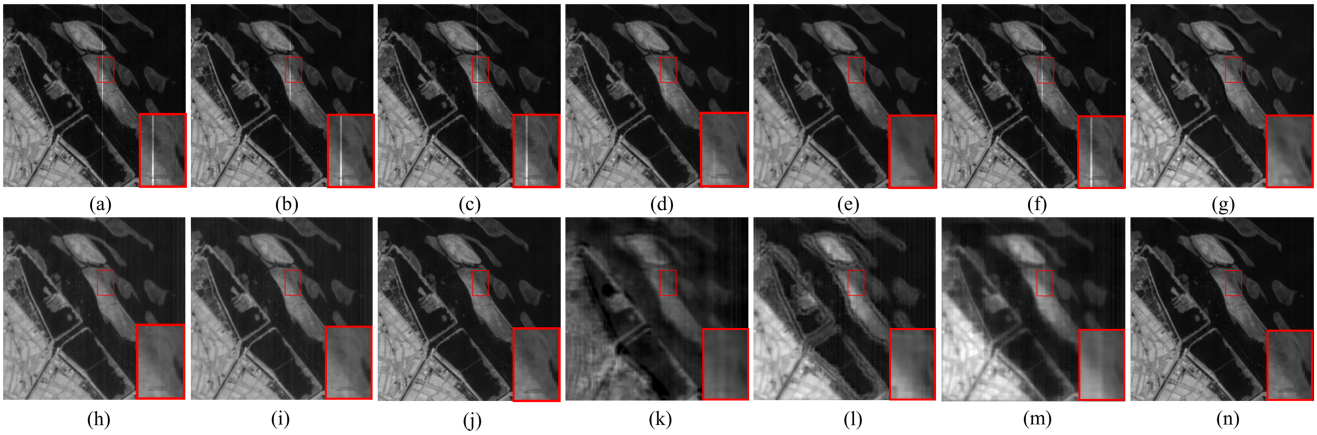

Fig. 15. Real noise removal results of image on EO-01 dataset. (a) Noisy. (b) BM4D. (c) BWBM3D. (d) LRMR. (e) LRTV. (f) DnCNN. (g) HSID-CNN. (h) HSI-SDeCNN. (i) 3DQRNN. (j) MACNet. (k) T3SC. (l) SST. (m) SERT. (n) Ours.

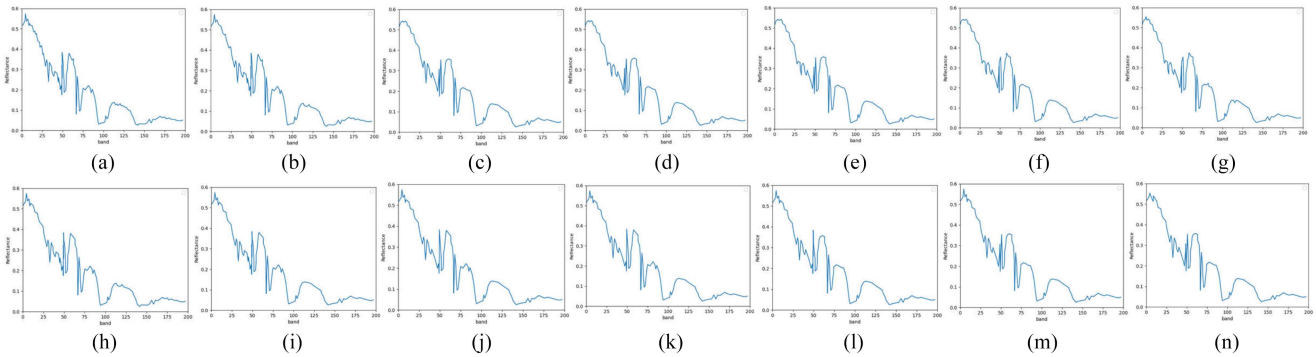

Fig. 16. Spectral reflectance of pixel (152, 82) for each band on EO-01 dataset image. The ordinate represents the reflectance value, and the abscissa represents the band number. (a) Noisy. (b) BM4D. (c) BWBM3D. (d) LRMR. (e) LRTV. (f) DnCNN. (g) HSID-CNN. (h) HSI-SDeCNN. (i) 3DQRNN. (j) MACNet. (k) T3SC. (l) SST. (m) SERT. (n) Ours.

TABLE IV
DENOISING RESULTS OF DIFFERENT MODULE SETTINGS

Model	MPSNR	MSSIM	MSAM
no_SSAT_ResOct	38.451	0.942	0.099
no_SSAT	39.543	0.967	0.098
no_ResOct	39.631	0.969	0.094
UOANet-II	39.435	0.968	0.097
UOANet-III (λ_1=0.25, λ_2=0.75)	38.939	0.959	0.097
UOANet-IV (λ_1=0.75, λ_2=0.25)	39.442	0.971	0.089
UOANet (λ_1=0.5, λ_2=0.5)	**40.975**	**0.981**	**0.081**

Bold represents the best performance.

TABLE V
DENOISING RESULTS OF DIFFERENT α SETTINGS

Model	MPSNR	MSSIM	MSAM	Flops(#)
α=0.1	39.651	0.970	0.093	2.455E+08
α=0.2	**40.975**	**0.981**	**0.081**	2.397E+08
α=0.3	38.965	0.963	0.094	2.211E+08
α=0.4	38.876	0.962	0.093	2.006 E+08
α=0.5	38.757	0.960	0.094	**1.887E+08**

Bold represents the best performance.

From Table IV, we can observe that ResOct encourages to focus on capture the noise feature in high-frequency and preserve image detailed structures in low-frequency, and thereby improves denoising performance with a gain of 1.0924 in terms of PSNR and 0.0251 in terms of SSIM, verifying its effectiveness. Similarly, PSNR, SSIM, improves from 38.4511 to 39.6315, 0.9422 to 0.9691, SAM reduced from 0.0992 to 0.0943, respectively, by adding the SSAT. This is because the SSAT can help capture the global spatial coherence and cross-channel correlation. Furthermore, by adding CBAM (namely UOANet-II) to the model, we can obtain a certain promotion. However, it can be seen that UOANet (λ_1= 0.5, λ_2= 0.5) achieves better index evaluation results, proving that parallel fusion of spatial and spectral attention information improving the denoising performance significantly.

To prove the effectiveness of SSAT module in exploring the spectral relationship and spatial relationship among feature, we

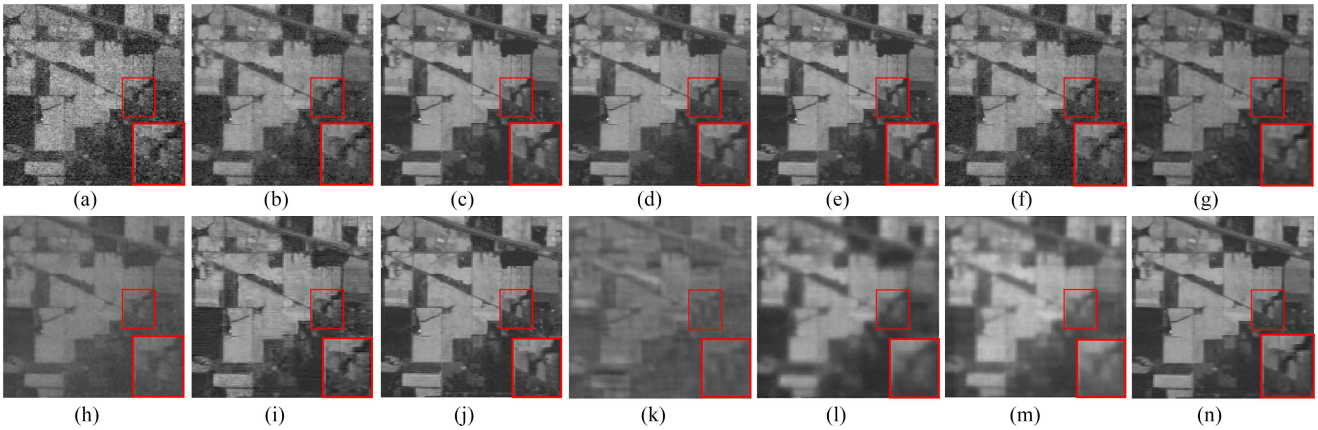

Fig. 17. Real noise removal results of image on Indian Pines dataset. (a) Noisy. (b) BM4D. (c) BWBM3D. (d) LRMR. (e) LRTV. (f) DnCNN. (g) HSID-CNN. (h) HSI-SDeCNN. (i) 3DQRNN. (j) MACNet. (k) T3SC. (l) SST. (m) SERT. (n) Ours.

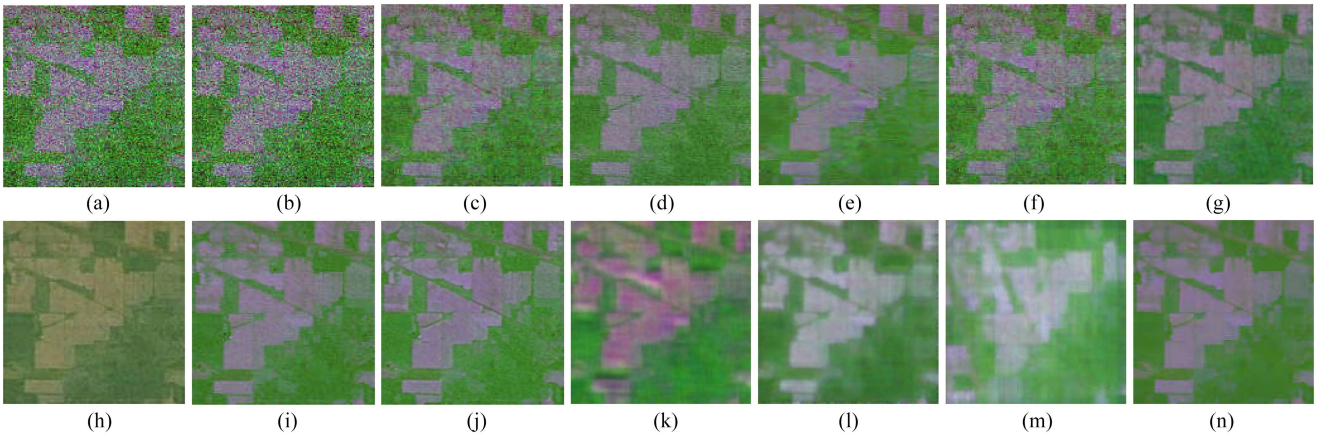

Fig. 18. False color images of the denoised results on band 144, 154, and 164 of Indian Pines dataset. (a) Noisy. (b) BM4D. (c) BWBM3D. (d) LRMR. (e) LRTV. (f) DnCNN. (g) HSID-CNN. (h) HSI-SDeCNN. (i) 3DQRNN. (j) MACNet. (k) T3SC. (l) SST. (m) SERT. (n) Ours.

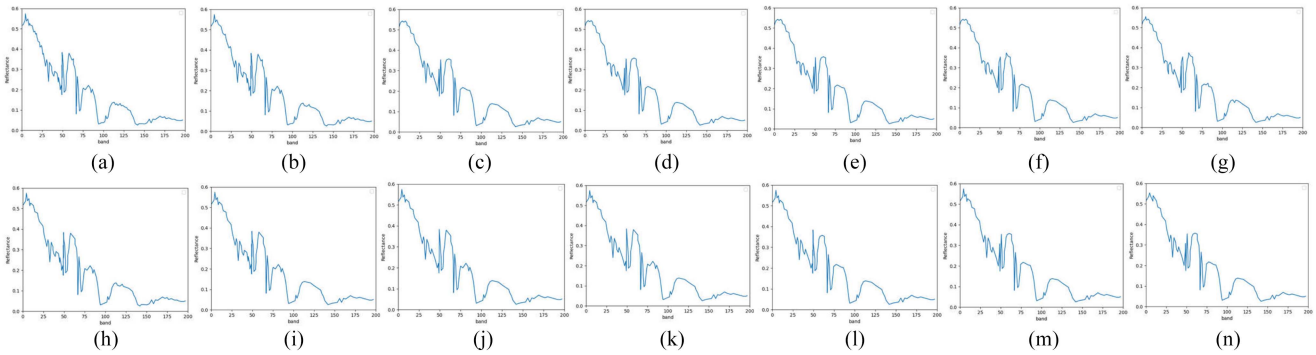

Fig. 19. Spectral reflectance of pixel (103, 64) across each band on Indian Pines dataset image. The ordinate represents the reflectance value, and the abscissa represents the band number. (a) Noisy. (b) BM4D. (c) BWBM3D. (d) LRMR. (e) LRTV. (f) DnCNN. (g) HSID-CNN. (h) HSI-SDeCNN. (i) 3DQRNN. (j) MACNet. (k) T3SC. (l) SST. (m) SERT. (n) Ours.

show feature maps learned by the SSAT module in Fig. 20. From Fig. 20(b), it can be easily seen that the features with a strong correlation to spectral information will have large response. For example, the whole sky appears red. It proves that the SSAT module is able to capture the spectral interrelationship along the channel dimension. From Fig. 20(c), it can easily observe that features with similar information will have a high reaction. For instance, the edges of the two cars appears red. It proves that the SSAT module is able to explore the spatial relationship among pixels.

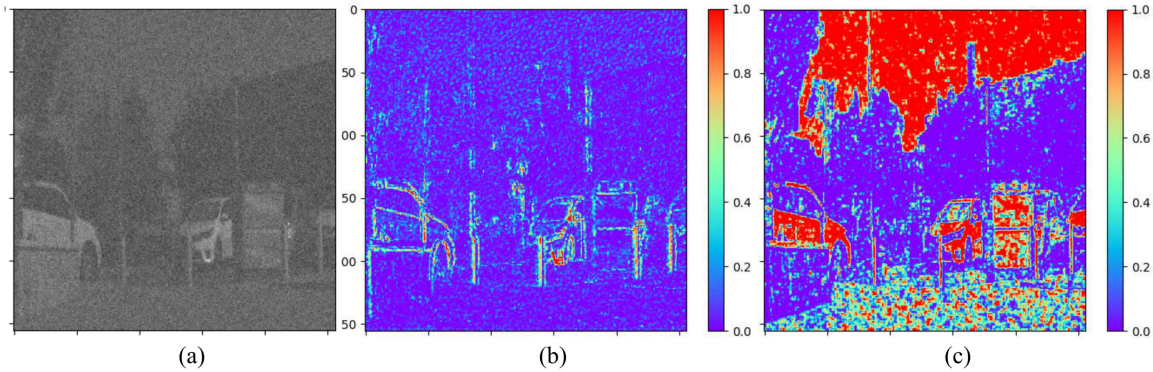

Fig. 20. Visualizations of the SSAT modules. (a) Input noisy image. (b) Feature map learned by the spectral attention module. (c) Feature map learned by the spectral attention module.

TABLE VI
PARAMETERS OF DIFFERENT CNN-BASED DENOISING METHODS

Model	Parameters (#)
DnCNN[28]	555137
HSID-CNN[31]	398400
HSI-SDeCNN[30]	1892100
QRNN3D[47]	860453
SST[50]	2145800
SERT[51]	1932400
MACNet[52]	435620
T3SC[53]	834100
UOANet	**237413**

Bold represents the best performance.

TABLE VII
TIME COSTS OF DIFFERENT DENOISING METHODS IN BLIND CASE AND MIXTURE CASE

Model	MPSNR (dB)	Time (s)
Blind Gaussian Noise		
UOANet	**40.975**	**0.94**
BM4D[9]	37.650	170.7
TDL[20]	40.249	28.7
ITSReg[26]	40.636	580.06
LLRT[27]	40.832	1813.4
DnCNN[28]	31.272	2.85
HSID-CNN[31]	36.541	1.51
HSI-SDeCNN[30]	37.513	1.31
QRNN3D[47]	35.664	1.4
Mixture Complex Noise		
UOANet	**35.617**	**0.93**
BM4D[9]	22.679	169.49
BWBM3D[10]	23.461	1.771
LRMR[44]	24.459	13.213
LRTV[45]	31.633	220.26
HSID-CNN[31]	24.431	1.32
HSI-SDeCNN[30]	26.743	1.3
QRNN3D[47]	28.399	1.41

Bold represents the best performance.

B. Sensitivity Analysis of α

By adjusting the α value, the parameters of Octave convolution can be changed to save network parameters and computing resources. The larger α, the greater the proportion of low-frequency features selected. Although this can reduce the complexity of the network, but may lead to the loss of high-frequency features. Therefore, when using Octave convolution, we need to select the appropriate a value by experiment.

As shown in Table V, as α increases, the complexity of the model decreases. When α is too large, the effect of image denoising will be degraded by overcompression of spatial information. When α is too small, low-frequency redundancy makes it difficult for the network to pay attention to high-frequency features, and the network gets poor results. As can be observed, the performance of the proposed UOANet is best when $\alpha = 0.2$.

C. Running Time Assessment

In this section, we compared training and testing time for different algorithm. For training time cost, we compared the parameters for different CNN -based methods. AS Table VI show, with the advantage of ResOct module, our method significantly reducing the number of parameters required. For testing time cost, we compared the average running time required for different methods of noise removal in blind Gaussian noise and mixture complex noise. AS Table VII shows, with the benefits of GPUs and end-to-end structures, deep-learning-based methods exhibit less runtime than the traditional methods. Our method performs best and requires the least amount of processing time.

V. CONCLUSION

Although many denoising methods have been suggested, most of them are unable to fully exploit the physical properties of hyperspectral noise images. In this article, we propose two key modules ResOct and SSAT in light of HSIs frequency distribution and spatial-spectral correlation. Based on these two modules, we improve the Unet network, and propose an HSI denoising network UOANet, which combines octave and attention mechanism. ResOct is embedded in the up-sampling process of UNet network, and uses the down-sampling low-frequency features to map the frequency features of noise, to remove the spatial redundancy, and to improve the network speed. SSAT is embedded in the down-sampling process, and the attention

mechanism performs both global average and global maximum mixing pooling on the channel and spatial dimensions, which can provide more effective global and local details for the network during the decoding process. Simultaneously, the residual modules are fused in U-Net network to avoid the problems of gradient disappearance, and further enhance the ability of denoising.

Finally, we compare the denoising results, efficiency and visual effects of different methods on ICVL, SDG, EO, and Indian Pines, it has been demonstrated that the proposed method is superior to both model-based and depth-learning-based methods in subjective visual effects and objective quantitative measurements.

Acknowledgment

The authors would like to thank the SDG BIGDATA Center and the National Space Science Center, Beijing, China, for providing them with data.

References

[1] L. Li, L. Jiang, J. Zhang, S. Wang, and F. Chen, "A complete YOLO-based ship detection method for thermal infrared remote sensing images under complex backgrounds," *Remote Sens.*, vol. 14, no. 7, Mar. 2022, Art. no. 1534, doi: 10.3390/rs14071534.

[2] T. Alipour-Fard, M. E. Paoletti, J. M. Haut, H. Arefi, J. Plaza, and A. Plaza, "Multibranch selective Kernel networks for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1089–1093, Jun. 2021, doi: 10.1109/lgrs.2020.2990971.

[3] X. Xue, H. Zhang, B. Fang, Z. Bai, and Y. Li, "Grafting transformer on automatically designed convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jun. 2022, Art. no. 5531116, doi: 10.1109/tgrs.2022.3180685.

[4] X. Liang, Y. Zhang, and J. Zhang, "Water retrieval embedded attention network with multiscale receptive fields for hyperspectral image refined classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2022, Art. no. 5509022, doi: 10.1109/tgrs.2021.3091985.

[5] X. Liang, Y. Zhang, and J. Zhang, "Attention symbiotic neural network for hyperspectral image refined classification based on relative water content retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5998–6016, Jul. 2021, doi: 10.1109/tgrs.2020.3013285.

[6] F. Xiong, J. Zhou, and Y. Qian, "Material based object tracking in hyperspectral videos," *IEEE Trans. Image Process.*, vol. 29, pp. 3719–3733, Jan. 2020, doi: 10.1109/tip.2020.2965302.

[7] L. Hongmei, H. Lin, Z. Ruiqiang, L. Lei, W. Diangang, and L. Jiazhou, "Object tracking in video sequence based on Kalman filter," in *Proc. Int. Conf. Comput. Eng. Intell. Control*, 2020, pp. 106–110, doi: 10.1109/iceic51584.2020.00029.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007, doi: 10.1109/tip.2007.901238.

[9] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Jan. 2013, doi: 10.1109/tip.2012.2210725.

[10] A. Heo, J.-H. Lee, E.-J. Choi, W.-C. Choi, S. H. Kim, and D.-J. Park, "Noise reduction of hyperspectral images using a joint bilateral filter with fused images," in *Proc. Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imagery XVII*, May 2011, pp. 1–10, vol. 10.1117, doi: 10.1117/12.884359.

[11] T. Hu, N. Liu, W. Li, R. Tao, F. Zhang, and P. Scheunders, "Destriping hyperspectral imagery by adaptive anisotropic total variation and truncated nuclear norm," in *Proc. 11th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens.*, 2021, pp. 1–4, doi: 10.1109/whispers52202.2021.9484033.

[12] W. He, H. Zhang, H. Shen, and L. Zhang, "Hyperspectral image denoising using local low-rank matrix recovery and global spatial–Spectral total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 713–729, Mar. 2018, doi: 10.1109/jstars.2018.2800701.

[13] J. Peng, H. Wang, X. Cao, X. Liu, X. Rui, and D. Meng, "Fast noise removal in hyperspectral images via representative coefficient total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5546017, doi: 10.1109/tgrs.2022.3229012.

[14] H. Wu, K. Zhang, S. Wu, M. Zhang, and S. Shi, "Hyperspectral superresolution reconstruction via decomposition of low-rank and sparse tensor," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 8943–8957, Oct. 2022, doi: 10.1109/jstars.2022.3214653.

[15] H. V. Nguyen, M. O. Ulfarsson, J. Sigurdsson, and J. R. Sveinsson, "Deep sparse and low-rank prior for hyperspectral image denoising," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2022, pp. 1217–1220, doi: 10.1109/igarss46834.2022.9884071.

[16] Y. Chen, W. He, X.-L. Zhao, T.-Z. Huang, J. Zeng, and H. Lin, "Exploring nonlocal group sparsity under transform learning for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5537518, doi: 10.1109/tgrs.2022.3202359.

[17] C. Deng, L. Li, Z. He, J. Li, and Y. Zhu, "Monte Carlo non-local means method for hyperspectral image denoising," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2018, pp. 4772–4775, doi: 10.1109/igarss.2018.8519108.

[18] W. He et al., "Non-local meets global: An integrated paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 4, pp. 2089–2107, Apr. 2022, doi: 10.1109/tpami.2020.3027563.

[19] H. Zeng, X. Xie, W. Kong, S. Cui, and J. Ning, "Hyperspectral image denoising via combined non-local self-similarity and local low-rank regularization," *IEEE Access*, vol. 8, pp. 50190–50208, 2020, doi: 10.1109/access.2020.2979809.

[20] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, and B. Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2949–2956, doi: 10.1109/cvpr.2014.377.

[21] X. Gong, W. Chen, and J. Chen, "A low-rank tensor dictionary learning method for hyperspectral image denoising," *IEEE Trans. Signal Process.*, vol. 68, pp. 1168–1180, Feb. 2020, doi: 10.1109/tsp.2020.2971441.

[22] S. Xiaorui and W. Lingda, "Denoising of hyperspectral images based on principal component analysis and adaptive sparse coding," in *Proc. 10th IAPR Workshop Pattern Recognit. Remote Sens.*, 2018, pp. 1–6, doi: 10.1109/prrs.2018.8486272.

[23] W.-H. Wu, T.-Z. Huang, X.-L. Zhao, J.-L. Wang, and Y.-B. Zheng, "Hyperspectral image denoising via tensor low-rank prior and unsupervised deep spatial–spectral prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5545514, doi: 10.1109/tgrs.2022.3228927.

[24] S. Xu, X. Cao, J. Peng, Q. Ke, C. Ma, and D. Meng, "Hyperspectral image denoising by asymmetric noise modeling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2022, Art. no. 5545214, doi: 10.1109/tgrs.2022.3227735.

[25] X. Zhou, Y. Zhang, J. Liu, J. Zhao, and Y. Hu, "Hyperspectral image restoration using 3-D hybrid higher degree total variation regularized nonconvex local low-rank tensor recovery," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Oct. 2022, Art. no. 6014805, doi: 10.1109/lgrs.2022.3217581.

[26] Q. Xie et al., "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 1692–1700, doi: 10.1109/cvpr.2016.187.

[27] Y. Chang, L. Yan, and S. Zhong, "Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 5901–5909, doi: 10.1109/cvpr.2017.625.

[28] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017, doi: 10.1109/tip.2017.2662206.

[29] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2019, doi: 10.1109/tgrs.2018.2859203.

[30] A. Maffei, J. M. Haut, M. E. Paoletti, J. Plaza, L. Bruzzone, and A. Plaza, "A single model CNN for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2516–2529, Apr. 2020, doi: 10.1109/tgrs.2019.2952062.

[31] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–Spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Feb. 2019, doi: 10.1109/TGRS.2018.2865197.

[32] J. Li, R. Cui, B. Li, Y. Li, S. Mei, and Q. Du, "Dual 1D-2D spatial-spectral CNN for hyperspectral image super-resolution," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 3113–3116, doi: 10.1109/igarss.2019.8898352.

[33] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral image super-resolution based on multi-scale wavelet 3D convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 2770–2773, doi: 10.1109/igarss.2019.8898813.
[34] V. K. Ha, J. Ren, Z. Wang, G. Sun, H. Zhao, and S. Marshall, "Multiscale spatial fusion and regularization induced unsupervised auxiliary task CNN model for deep super-resolution of hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4583–4598, May 2022, doi: 10.1109/jstars.2022.3176969.
[35] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4590–4603, Jul. 2020, doi: 10.1109/tgrs.2020.2964288.
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778, doi: 10.1109/cvpr.2016.90.
[37] Y. Chen et al., "Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 3434–3443, doi: 10.1109/iccv.2019.00353.
[38] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.
[39] J. Park, "BAM: Bottleneck attention module," Jul. 2018, *arXix:1807.06514*.
[40] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 19–34, doi: 10.1007/978-3-319-46478-7_2.
[41] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, 2008, Art. no. 800, doi: 10.1049/el:20080522.
[42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/tip.2003.819861.
[43] R. H. Yuhas, A. F. H. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the Spectral Angle Mapper (SAM) algorithm," in *Proc. JPL, Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, Jun. 1992, vol. 1, pp. 1–3.
[44] C. Wang et al., "Translution-SNet: A semisupervised hyperspectral image stripe noise removal based on transformer and CNN," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jul. 2022, Art. no. 5533114, doi: 10.1109/tgrs.2022.3182745.
[45] C. Wang et al., "Hyperspectral image stripe removal network with cross-frequency feature interaction," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Dec. 2021, Art. no. 5521515, doi: 10.1109/tgrs.2021.3138740.
[46] X. Cao, X. Fu, C. Xu, and D. Meng, "Deep spatial-spectral global reasoning network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Apr. 2022, Art. no. 5504714, doi: 10.1109/tgrs.2021.3069241.
[47] K. Wei, Y. Fu, and H. Huang, "3-D quasi-recurrent neural network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 363–375, Jan. 2021, doi: 10.1109/tnnls.2020.2978756.
[48] C. Zou, C. Zhang, M. Wei, and C. Zou, "Enhanced channel attention network with cross-layer feature fusion for spectral reconstruction in the presence of Gaussian noise," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 9497–9508, Nov. 2022, doi: 10.1109/jstars.2022.3218820.
[49] J. Wang, F. Gao, J. Dong, Q. Du, and H.-C. Li, "Change detection from synthetic aperture radar images via dual path denoising network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2667–2680, 2022, doi: 10.1109/jstars.2022.3159619.
[50] M. Li, Y. Fu, and Y. Zhang, "Spatial-spectral transformer for hyperspectral image denoising," in *Proc. IEEE Conf. Artif. Intell. AAAI*, 2023, pp. 1–10.
[51] M. Li, J. Liu, Y. Fu, Y. Zhang, and D. Dou, "Spectral enhanced rectangle transformer for hyperspectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 5805–5814.
[52] F. Xiong, J. Zhou, Q. Zhao, J. Lu, and Y. Qian, "MAC-net: Model-aided nonlocal neural network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5519414, doi: 10.1109/TGRS.2021.3131878.
[53] T. Bodrito, A. Zouaoui, J. Chanussot, and J. Mairal, "A trainable spectral-spatial sparse coding model for hyperspectral image restoration," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 5430–5442, 2021. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/2b515e2bdd63b7f034269ad747c93a42-Paper.pdf
[54] F. Xiong, J. Zhou, S. Tao, J. Lu, J. Zhou, and Y. Qian, "SMDS-net: Model guided spectral-spatial network for hyperspectral image denoising," *IEEE Trans. Image Process.*, vol. 31, pp. 5469–5483, 2022, doi: 10.1109/tip.2022.3196826.
[55] T. Zhan et al., "A novel cross-scale octave network for hyperspectral and multispectral image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5545816, doi: 10.1109/tgrs.2022.3229086.
[56] C. Pu, H. Huang, X. Shi, and T. Wang, "Semisupervised spatial-spectral feature extraction with attention mechanism for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6011905, doi: 10.1109/lgrs.2022.3193304.
[57] Z. Kan, S. Li, M. Hou, L. Fang, and Y. Zhang, "Attention-based octave network for hyperspectral image denoising," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 1089–1102, 2022, doi: 10.1109/jstars.2021.3129622.

Siqi Wang received the B.S. degree in computer science and technology from Sichuan University, Chengdu, China, in 2021. She is currently working toward the M.A. degree in physical electronics with Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, and with the University of Chinese Academy of Sciences, Beijing, China.

Her current research interests include intelligent infrared detection and infrared image processing.

Liyuan Li received the B.S. degree in optoelectronics information science and engineering from Dalian University of Technology, Dalian, China, in 2018. She is currently working toward the Ph.D. degree in physical electronics with Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China.

Her current research interests include dim and small targets detection of IR through machine learning.

Xiaoyan Li received the B.S. degree in mechanism design, manufacturing, and automatization from Northwest A&F University, Xi'an, China, in 2016. He is currently working toward the Ph.D. degree in electronic circuit and system with Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Beijing, China, and with the University of Chinese Academy of Sciences, Beijing, China.

His current research interests include on-orbit accurate navigation and geometric calibration of remote sensing satellites.

Jingwen Zhang received the B.S. degree in communication engineering from Shanghai Maritime University, Shanghai, China, in 2020. She is currently working the Ph.D. degree in electronic circuit and system with Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China.

Her current research interests include intelligent infrared detection, infrared imaging system, and infrared image processing.

Lixing Zhao received the B.S. degree in intelligence science and technology from the Nanjing University of Science and Technology, Nanjing, China, in 2021. She is currently working toward the M.A. degree in electronic information with the Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou, China.

Her research interest includes GCPs extraction for heterologous remote sensing images through deep learning.

Xiaofeng Su received the B.S. degree in information engineering from Xiamen University, Xiamen, China, in 2010 and the Ph.D. degree in electronic circuit and system from Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Beijing, China, in 2015.

He is mainly engaged in the development of space infrared remote sensing payload, infrared focal plane detector information acquisition, and researches of image processing technology, infrared dim small target detection technology, and other relative aspects.

Fansheng Chen (Senior Member, IEEE) received the B.S degree in optoelectronic information engineering from Shandong University, Jinan, China, in 2002 and the Ph.D. degree in physical electronics from Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, in 2007.

Since 2013, he has been a Professor with the Shanghai Institute of Technical Physics, Chinese Academy of Sciences. He is committed to the research and development of the space infrared staring detection instruments, the high spatial and temporal resolution photoelectric payloads, and the application of infrared multispectral information acquisition technology in artificial intelligence, target recognition, and other relative aspects. His research interests include the design of spatial high resolution remote sensing and detection payloads, high-speed and low noise information acquisition technology, and infrared dim small target detection technology.