Attention-Based Octave Network for Hyperspectral Image Denoising

Ziwen Kan, Suhang Li, Mingzheng Hou, Leyuan Fang, *Senior Member, IEEE*, and Yi Zhang, *Senior Member, IEEE*

***Abstract*—Inevitable corruption and degeneration make the performance of subsequent high-level semantic tasks in hyperspectral images (HSIs) unsatisfactory. Despite that many denoising methods have been proposed, significant room for improvement still remains. To better suppress noise and preserve the HSI spatial–spectral structure, we propose an attention-based Octave dense network. A separable spectral feature extraction module is introduced to extract the spatial–spectral features consistent with the structure prior. The extracted features are fine-tuned by the attention module in both channel and spatial domains; then, several dense denoising blocks are elaborately employed to focus on noise feature learning; in order to focus on high-frequency features, which usually have more noise information, we introduce the Octave kernel to implement these blocks. Experiments based on simulated and real-world noisy images demonstrate that the proposed method outperforms the existing traditional and learning-based methods in both quantitative evaluations and visual effects, benefiting the subsequent classification task. In addition, the effectiveness of each module is proven by ablation experiments. Our source code is made available at: https://github.com/LbzSteven/AODN.**

***Index Terms*—Attention module, hyperspectral image (HSI) denoising, octave network, separable feature extraction.**

I. Introduction

HYPERSPECTRAL images (HSIs) have high spectral resolution and hundreds of channels, which allow them to have abundant information in both spectral and spatial domains. HSIs have been applied in numerous remote sensing applications, such as ground object classification [2], unmixing [3], and anomaly detection [4]. However, limited by the imaging condition, HSIs usually 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.

In recent decades, a number of methods have been proposed to obtain noise-free HSIs from noisy observations [5]–[24]. The preliminary denoising strategy is to simply apply natural image denoising methods [5]–[7] to process noisy HSIs band by band. However, this strategy ignores the spectral correlation between adjacent bands, which may cause spectral distortion in the results.

To explore spectral–spatial information and suppress noise, numerous methods, such as wavelet [8], nonlocal similarity [9]–[11], sparse representation [12], [13], and low-rank decomposition [14]–[17], [22]–[24], have been proposed. Despite the significant improvements obtained, considerable efforts for manual parameter tuning and computational time are required. Meanwhile, the noise level usually must be determined to achieve more accurate results. In summary, it is necessary to develop a more flexible and robust denoising method for HSIs.

Recently, with the rapid development of deep learning, HSIs have been extensively studied for computer vision, including image denoising. For example, DnCNN [25] learned the nonlinear features via residual connection. Mao *et al.* [26] presented a deep fully convolutional encoding–decoding framework for image restoration. However, due to the high spectral correlation among HSIs, directly applying these methods to hyperspectral denoising may not obtain satisfactory results [27]. In [27], a bandwise denoising network aided by adjacent bands was proposed to improve the denoising performance. Based on this work, in [28], nonlocal blocks and channel attention were introduced to capture the global features. In [29], a network with channel attention and residual connection was proposed. Dong *et al.* [30] proposed a 3-D denoising network equipped with separable convolution. Song *et al.* [31] proposed an unsupervised model using wavelet directional CycleGAN. These methods achieved good results, but significant room for improvement remains.

From the signal processing perspective, the high-frequency component is considered to contain more noise and fine details, and the low-frequency component contains more content. To the best of our knowledge, no existing denoising methods in convolutional neural network (CNN)-based HSI denoising decompose high-frequency features from original features and separately process both parts. In addition, different from a simple 3-D feature block that has equal dimensions, the spectral dimension

Manuscript received August 13, 2021; revised October 22, 2021 and November 10, 2021; accepted November 17, 2021. Date of publication November 22, 2021; date of current version January 20, 2022. This paper is an extension of the conference paper [1]. This work was supported in part by Sichuan Science and Technology Program under Grant 2021JDJQ0024, in part by the National Natural Science Fund of China under Grant 61922029, and in part by the Science and Technology Plan Project Fund of Hunan Province under Grant 2019RS2016. *(Corresponding authors: Leyuan Fang; Yi Zhang.)*

Ziwen Kan, Suhang Li, and Yi Zhang are with the College of Software Engineering, Sichuan University, Chengdu 610065, China, and also with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: kanziwensteven@gmail.com; lisuhang2000@gmail.com; yzhang@scu.edu.cn).

Mingzheng Hou is with the National Key Laboratory of Fundamental Science on Synthetic Vision, Sichuan University, Chengdu 610065, China, and also with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: houmingzheng@scu.edu.cn).

Leyuan Fang is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: leyuan_fang@hnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2021.3129622

contains the information of sampling bands and the spatial dimension contains the height and width information of terrestrial object, that is, structural difference exists between spectral and spatial domains and a traditional vanilla 3-D convolution kernel cannot explicitly represent this difference.

In this article, to overcome the shortcomings mentioned above, we propose an attention-based Octave dense network (AODN). We use multiscale separable convolution to extract information in adjacent bands, attention mechanism is introduced to guide feature learning in both spatial and channel domains, and we design a new denoising block combining the Octave convolution and dense connection to reduce the redundancy of low-frequency information. Our main contributions are summarized as follows.

1) A new feature extractor with multiscale separable convolution is proposed to explore the adjacent spatial–spectral information and reduce the model scale.
2) A novel dense denoising block aided by the Octave kernel and the attention mechanism is proposed to suppress noise. The Octave kernel is introduced to extract high-frequency features, enabling the network to locate the noise information. It is the first attempt to introduce the attention mechanism to guide the network to focus additional attention on meaningful feature maps for noise suppression in both spatial and channel domains.
3) Extensive experiments were conducted to demonstrate that the proposed network can effectively remove the noise in both simulated and real scenarios and achieve better qualitative and quantitative results than several state-of-the-art methods.

This paper is an extension of the conference paper [1]. Comparing with the conference version, this work has thefollowing improvements. Firstly, more systematic description about the theory part was presented. Secondly, we redesigned the experimental parts. More SOTAs were added into comparison and ablation studies were performed and presented.

The rest of this article is organized as follows. Section II describes the HSI degradation model and then introduces the existing HSI denoising methods. In Section III, the proposed model is elaborated. The simulated and real-data experimental results are presented in Section IV. Finally, Section V concludes this article.

II. Related Work

A. Hyperspectral Noise Degradation Model

An observed HSI is a 3-D data tensor $Y \in R^{M\times N\times B}$, in which M and N represent the spatial resolution and B denotes the spectral resolution. The HSI degradation model can be formulated as follows:

$$Y = X + D \tag{1}$$

where $X \in R^{M\times N\times B}$ is the noise-free data we attempt to recover and $D \in R^{M\times N\times B}$ represents the additive noise, such as Gaussian noise. Therefore, the denoising problem can be treated to reconstruct X from the noisy observation Y.

B. Traditional Hyperspectral Denoising Methods

Existing hyperspectral denoising methods can be roughly divided into three categories: band-by-band methods, transform domain methods, and model-optimization-based methods.

1) Band-by-Band Methods: Band-by-band methods directly apply natural image denoising methods to HSI denoising, such as block matching and 3-D filtering [5], weighted nuclear norm minimization [6], and expected patch log likelihood [7]. However, these methods only focus on the spatial information and fail to adopt the high correlations between the spectral bands, which usually lead to unsatisfactory results.

2) Transform Domain Methods: Transform domain methods try to obtain noise-free HSIs by transforming them with various basis functions. For example, Atkinson *et al.* [8] used discrete Fourier transform and wavelet-based estimation schemes for hyperspectral imagery; Rasti *et al.* proposed a denoising model using 3-D wavelets [18] and another method using first-order roughness penalty in a wavelet domain [19]. Othman and Qian [20] proposed a hybrid approach of spatial and spectral wavelet shrinkage on the derivative domain to suppress noise. However, these methods need to manually select the transform function, and the differences in the geometrical characteristics are ignored.

3) Model-Optimization-Based Methods: To make full use of the prior information of HSIs in both spatial and spectral domains, several methods have been proposed, such as total variation [21], nonlocal similarity [9]–[11], sparse representation [12], [13], and low-rank tensors [14]–[17], [22]–[24]. Specifically, Zhang [21] applied a cubic total variation model to denoise HSIs. Maggioni *et al.* [9] modified nonlocal block matching methods into a volumetric data pattern. Peng *et al.* [10] considered the nonlocal similarity over space and the global correlation across spectra using nonlocal tensor dictionary learning. He *et al.* [11] combined spatial nonlocal similarity and global spectral low-rank property (NGmeet). Lu *et al.* [12] improved noise-free estimation by utilizing spectral and spatial information via sparse representation. Li *et al.* [13] proposed a joint spectral–spatial distributed sparse representation for HSI denoising. Zhang *et al.* [14] proposed an HSI restoration method based on low-rank matrix recovery. Renard *et al.* [15] exploited low-rank matrix approximation to reduce spectral dimensionality. Wei *et al.* [16] proposed total-variation-regularized low-rank matrix factorization. Chang *et al.* [17] proposed hyper-Laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising (LLRT). Rasti *et al.* [22] proposed to use Stein's unbiased risk estimator to select all the parameter sparse and low-rank modeling (HyRes). Zhuang and Bioucas-Dias [23] presented a fast algorithm based on low-rank and sparse representation to suppress noise (FastHyDe). Zhang *et al.* [24] proposed a restoration model combining total variation regularization and nonlocal low-rank decomposition.

Despite the promising results achieved, laborious parameter adjustment prevents most of the methods' usage in practice, and there is still room for improvement. Therefore, it is urgent to develop a robust, flexible, and efficient architecture to denoise HSIs.

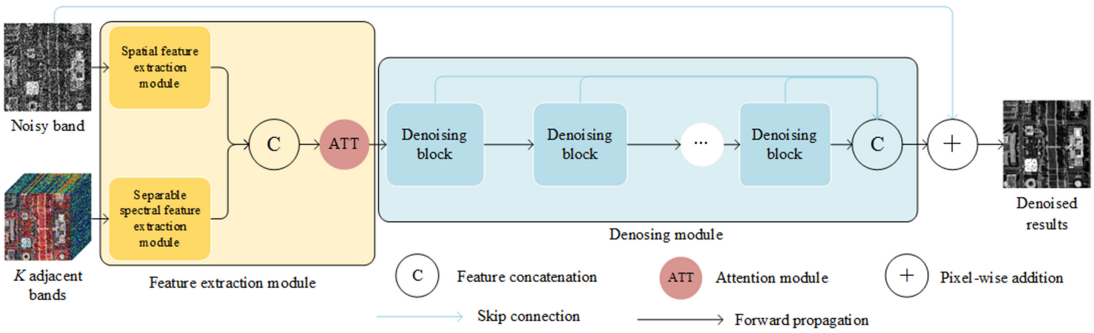


Fig. 1. Overall network structure of the AODN.

C. CNN-Based Hyperspectral Denoising Methods

In recent years, CNN-based methods have shown their potential in image denoising [25], [26], [32], [33]. Due to the properties of hyperspectral resolution and spectral correlation, directly applying these methods to HSI denoising cannot obtain satisfactory results. To preserve more spectral information and reduce computational cost, several attempts have been made [27]–[30], [34], [35]. In [27], an auxiliary adjacent spectral branch was introduced to guide band-by-band denoising, which achieved impressive results. SSGN [34] applied spatial–spectral gradients to address mixed noise. ENCAM [28] tried to learn global information with nonlocal blocks. ARDN [29] and ENCAM used a channel attention module to guide the feature tuning. Meanwhile, some methods restored the HSI data with a 3-D block, but the network structures must be modified once the number of bands changes. 3-D dilated kernels were introduced by Liu and Lee [35]. However, their model is unable to handle all of the data due to memory constraints and must randomly select some bands of HSIs to conduct training and experiments. The methods with vanilla 3-D kernels do not consider the difference between the spatial and spectral domains while extracting features from adjacent bands. Dong *et al.* [30] proposed a 3-D denoising architecture with separable kernels to reduce the computational costs. Additionally, no existing methods explicitly decompose high-frequency features that contain more noise and detailed information. Compared with these methods, the proposed method can fully extract spectral information with a separable convolution extraction model and focus additional attention on high-frequency features via Octave convolution.

III. Proposed Method

A. Network Architecture

The architecture of the proposed AODN is illustrated in Fig. 1. The input of AODN is composed of a noisy band with a size of $W \times H$ and auxiliary adjacent bands with sizes of $W \times H \times K$. $K/2$ bands in front of current band and $K/2$ bands behind current band, totally K bands, are selected as the adjacent bands. For the first $K/2$ or the last $K/2$ bands in the spectrum, we simply use the first K bands or the last K bands as their adjacent bands, respectively.

A spatial extraction module is introduced to learn spatial information from the noise band, and a separable spectral feature extraction module is introduced to learn spatial–spectral information from adjacent bands. This strategy is flexible since HSIs from different bands can be processed with a unified structure, which means we process each band one by one.

After the concatenation and attention module, six denoising blocks are cascaded to extract noise features. Several skip connections are used, which have been proven effective in solving the vanishing gradient problem in HSI denoising [26]. The residual φ_i can be defined as follows:

$$\varphi_i = y_i - x_i \tag{2}$$

where y_i is the observed corrupted ith band data and x_i is the corresponding ith noise-free data.

The mean squared error is employed as the loss function, which is formulated as a training group with N pairs $\{y_i, s_i, x_i\}_N$ of image data, y_i and s_i represent the ith noisy data and its adjacent spectral data, respectively, and x_i is the ground truth of the ith data. The loss function is as follows:

$$L_\theta = \frac{1}{N}\sum_{i=1}^{N} \|F(y_i, s_i, \theta) - \varphi_i\|_2^2 \tag{3}$$

where φ_i is the ground truth of the ith residual, N denotes the number of training samples, $F(\cdot)$ represents the AODN network, and θ represents the trainable parameter set.

B. Separable Spectral Feature Extraction Module

The separable feature extraction module is demonstrated in Fig. 2, which can learn spatial spectrum from two branches. Both branches apply multiscale kernels to extract features from multiple scales. The upper branch uses vanilla 2-D convolution to learn spatial information. The lower branch attempts to learn auxiliary spatial–spectral features from a 3-D input tensor. Generally, vanilla 3-D convolution is suited for extracting feature maps from 3-D input, but 3-D convolution will cause extra computational expense. In addition, there are structural differences between interspatial and spectral features, which are difficult to extract properly using a vanilla 3-D convolution.

As we mentioned above, spatial and spectral features are different in the structure prior and are not represented well by vanilla 3-D kernels. To tackle this problem, the spectral branch

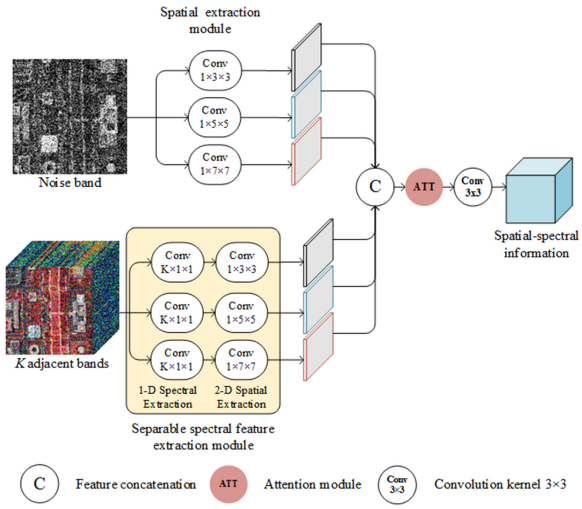


Fig. 2. Feature extraction module.

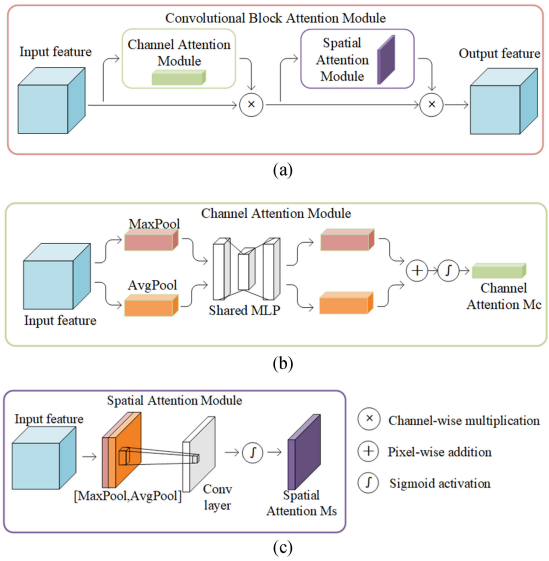


Fig. 3. Proposed attention module. (a) Overall framework of the attention module. (b) Channel attention module. (c) Spatial attention module.

applies 3-D separable convolution to learn spectral information. The proposed separable feature extraction module extracts two features separately, with 1-D kernels focusing on spectral information and 2-D kernels focusing on spatial information; thus, the spectral and spatial information are represented appropriately, which conforms to the structure prior. Additionally, the separated kernel reduces the total number of parameters, easing the network training. After that, the learned features are concatenated, processed by an attention module and fused by a convolution. The fused feature consists of fine extracted multiscale spatial–spectral information, which can be used in later denoising blocks.

C. Attention Module

The proposed attention module is shown in Fig. 3. As shown in Fig. 3(a), the attention module consists of two parts. The first channel module learns the channel attention map M_C to refine features in channel domain. Then, the spatial attention module learns the spatial attention map M_S to refine the features in spatial domain. The overall process can be described as follows:

$$F' = M_C(F) \otimes F$$
$$F'' = M_S(F') \otimes F' \tag{4}$$

where $\otimes$ denotes elementwise multiplication, F is the input feature, F' is the feature tuned by the channel attention module, and F'' is the output feature. This block has been proven effective in enhancing the channelwise and spatialwise feature representation ability of CNNs [36].

1) Channel Attention: As shown in Fig. 3(b), this module aggregates feature maps with both average pooling and max pooling along the spatial domain. The channel attention module forwards them through a shared multilayer perceptron (MLP) that has only one hidden layer. The hidden layer size is set to $C/16 \times 1 \times 1$, where C is the input channel size. The outputs of two branches are merged to obtain the channel attention map M_C. The process can be described as follows:

$$M_C(F) = \sigma(W_1(W_0(F^C_{\text{avg}})) + W_1(W_0(F^C_{\text{max}}))) \tag{5}$$

where σ is the Sigmoid activation function and W_1 and W_0 are the shared MLP parameters. $F^C_{\text{avg}} \in R^{1\times1\times B}$ and $F^C_{\text{max}} \in R^{1\times1\times B}$ are the features generated by average and max pooling operations in spatial domain, respectively.

2) Spatial Attention: As shown in Fig. 3(c), spatial attention focuses on the interspatial domain. First, average pooling and max pooling operations along the channel axis are employed to generate the descriptors: $F^S_{\text{avg}} \in R^{M\times N\times1}$ and $F^S_{\text{max}} \in R^{M\times N\times1}$. Two descriptors are concatenated and fed into a vanilla convolution. The spatial attention process can be described as follows:

$$M_S(F) = \sigma(f[F^S_{\text{avg}}; F^S_{\text{max}}]) \tag{6}$$

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

Since our proposed attention module is dedicated to fine-tuning the learned features from both spectral and spatial information and the spatial feature and separable spectral feature extraction modules only extract the corresponding feature, we only apply the attention module after the feature concatenation and each denoising block.

D. Octave Dense Denoising Block

In natural images, information can be decomposed into different frequencies. Higher frequencies usually contain fine details and noise information that are indispensable for image restoration, while lower frequencies usually contain global structural information and are redundant in most cases [37]. Chen *et al.* [37] proposed an Octave convolution block to reduce redundancy in the feature map and focus on the high-frequency parts, which usually contain noise information. The Octave convolution can restore "slowly" by varying features at a lower resolution while reducing both memory and computational costs.

This characteristic has made it feasible for Octave networks to be applied in HSI high-level tasks [38]–[40]. Meanwhile, Octave convolution can focus on high-frequency features and learn more

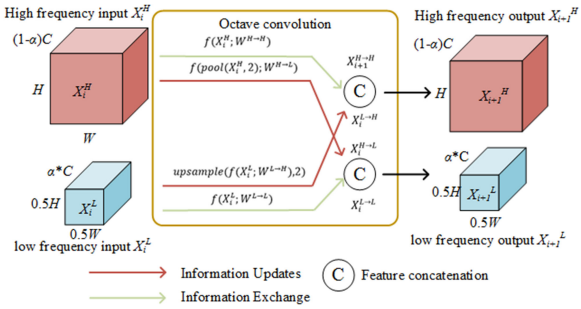


Fig. 4. Octave convolution kernel.

noisy information. Based on this consideration, to the best of our knowledge, the model proposed in our article is the first attempt to apply an Octave network to the HSI denoising task. As shown in Fig. 4, Octave convolution can not only process both frequency tensors, but also enable efficient interfrequency communication. The red lines represent intrafrequency channel information updating, and the green lines represent interfrequency channel information exchange. In the Octave kernel, the ratio α represents the low-frequency proportion. The $\alpha \times c$ channels are the low-frequency features, whose spatial resolutions are reduced to $0.5H \times 0.5$ W, and the $(1-\alpha) \times c$ channels are the high-frequency features, whose spatial resolutions remain $H \times W$. Octave convolution has the exact same parameters as vanilla convolution, but it can reduce the computational and memory costs by reducing the low-frequency features.

Specifically, assuming that the input and the output of an Octave block are $X_i = \{X_i^H, X_i^L\}$ and $X_{i+1} = \{X_{i+1}^H, X_{i+1}^L\}$, respectively, H represents the high-frequency group and L represents the low-frequency group. In the Octave network, $X_{i+1}^H = X_i^{H\to H} + X_i^{L\to H}$ and $X_{i+1}^L = X_i^{L\to L} + X_i^{H\to L}$, where $X_i^{A\to B}$ represents a convolutional update from group A to group B. Specifically, $X_i^{H\to H}, X_i^{L\to L}$ means intrafrequency forward propagation and $X_i^{L\to H}$, $X_i^{H\to L}$ means interfrequency forward propagation. Additionally, the Octave kernel can be split into two components $W = [W^H, W^L]$. Each component can be separated into intra- and interfrequency parts, which are defined as $W^H = [W^{H\to H}, W^{L\to H}]$ and $W^L = [W^{L\to L}, W^{H\to L}]$, respectively. Thus, $X_{i+1}^H$and X_{i+1}^L can be computed as follows:

$$\begin{aligned} X_{i+1}^H &= X_i^{H\to H} + X_i^{L\to H} \\ &= \sum (W^H)^T X_i \\ &= \sum (W^{H\to H})^T X_i^H + \text{upsample}\left(\sum (W^{L\to H})^T X_i^L\right) \\ X_{i+1}^L &= X_i^{H\to L} + X_i^{L\to L} \\ &= \sum (W^L)^T X_i \\ &= \sum (W^{H\to L})^T \text{pool}(X_i^H) + \sum (W^{L\to L})^T X_i^L \end{aligned} \tag{7}$$

where T represents the transpose operation, pool$(\cdot)$ represents the average pooling operation, and upsample$(\cdot)$ represents the upsampling operation. The denoising block is shown in Fig. 5, in which we use an Octave dense structure. For the high- or low-frequency channel alone, the output of each layer has the same number of channels, and the input of the channel is the concatenation of all the outputs from the previous layers. The Octave kernel is applied to update information in the same channel and exchange information between channels. The Octave dense structure leverages features more effectively to make the parameter transfer frequently. After the dense structure, an attention module is applied to refine the feature map. The block also applies residual connection to avoid vanishing gradients.

IV. Experimental Results

A. Datasets and Implementation Details

1) Datasets: Three datasets were used for the simulated and real-data experiments, including the Washington DC Mall dataset, the Pavia University (PU) dataset, and the Indian Pines (IP) dataset. The pixel levels of these images were normalized to [0,1].

The Washington DC Mall dataset was collected by the Hyperspectral Digital Imagery Collection Experiment, with a spatial resolution of 1208 × 303 and 191 bands. The images were cropped into two parts: one with a size of 1080 × 303 for training and the other with a size of 200 × 200 for testing.

The IP dataset was collected by the Airborne Visible Infrared Imaging Spectrometer with a spatial resolution of 145 × 145 and 220 bands.

The PU dataset was acquired by the Reflective Optics System Imaging Spectrometer with a spatial resolution of 610 × 340 and 103 bands. The IP and PU datasets contain real noise and were employed for the real-data experiments.

During the training phase, the images in the training set were cropped into patches with a size of 40 × 40, and the stride was set to 40. The simulated noisy data were generated by adding Gaussian noise. Data augmentation was performed, which includes resizing scale [0.5, 1, 1.5, 2] times, flipping along the horizontal and vertical directions and rotating 0°, 90°, 180°, and 270°.

2) Parameter Setting and Network Training: PyTorch was adopted to implement the proposed model. The number of adjacent bands, K, was set to 64, and the low-frequency ratio α was set to 0.2. Adam [41] was employed as the optimizer with momentum parameters of 0.9, 0.999, and 10^{-8}. The Kaiming initialization method in [42] was introduced to initialize the parameters, and the learning rate was initially set to 10^{-5}. The training process took 100 epochs. We train our network on a PC with an i9-10900X CPU and an NVIDIA 2080Ti with 11-GB memory.

3) Comparative Methods and Quantitative Indices: Several state-of-the-art methods were compared, including four popular similarity or low-rank-based methods BM4D [9], LLRT [17], HyRes [22], and FastHyDe [23], and deep-learning-based methods HSIDCNN [27] and ENCAM [28].

To quantitatively measure the denoising performance, three widely used indices were chosen, including the peak signal-to-noise ratio (PSNR), structural similarity index measurement (SSIM) [43], and spectral angle map (SAM) [44]. Usually, higher PSNR and SSIM mean higher reconstruction achievement, and lower SAM means a smaller difference between the spectral structures of ground truth and the denoised result.

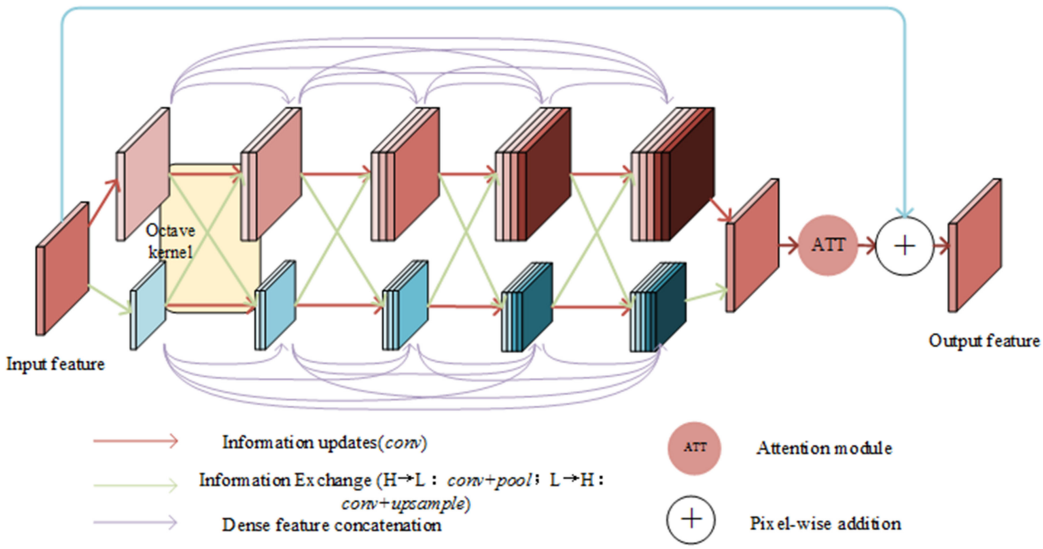


Fig. 5. Proposed Octave dense denoising block.

TABLE I
EFFECT OF DIFFERENT ADJACENT BANDS K IN THE PROPOSED AODN

Methods	PSNR	SSIM	SAM	#FLOPs
K=0	22.9637	0.7739	0.1995	**1.328E+09**
K=16	29.8123	0.9585	0.1011	1.452E+09
K=32	30.1309	0.9629	0.0937	1.454E+09
K=48	30.8069	0.9687	0.0890	1.455E+09
K=64	**31.1509**	**0.9705**	**0.0825**	1.457E+09
K=80	30.9028	0.9693	0.0843	1.458E+09
K=96	30.6813	0.9682	0.0873	1.460E+09
K=112	30.6186	0.9652	0.0877	1.462E+09
K=128	30.1039	0.9596	0.0908	1.463E+09

TABLE II
EFFECT OF DIFFERENT LOW-FREQUENCY RATIO α IN THE PROPOSED AODN

Methods	PSNR	SSIM	SAM	#FLOPs
α=0.1	31.0012	0.9690	0.0837	1.593E+09
α=0.2	**31.1509**	**0.9705**	**0.0825**	1.457E+09
α=0.3	30.7802	0.9677	0.0905	1.327E+09
α=0.4	30.9086	0.9695	0.0836	1.204E+09
α=0.5	30.7514	0.9678	0.0879	**1.086E+09**

TABLE III
EFFECT OF EACH COMPONENT IN THE PROPOSED AODN

Methods	Att	Oct	Sep	PSNR	SSIM	SAM	#FLOPs
no_module				28.9005	0.9549	0.1083	1.842E+09
no_oct_att			✓	29.2141	0.9580	0.1043	1.731E+09
no_oct_sep	✓			30.4597	0.9664	0.0925	1.845E+09
no_att_sep		✓		30.4836	0.9658	0.0928	1.564E+09
no_sep	✓	✓		30.8130	0.9681	0.0845	1.568E+09
no_att		✓	✓	30.5846	0.9666	0.0904	1.453E+09
no_oct	✓		✓	30.5079	0.9672	0.0906	1.735E+09
AODN	✓	✓	✓	**31.1509**	**0.9705**	**0.0825**	1.457E+09
no_cha_att		✓	✓	30.6759	0.9679	0.0900	1.455E+09
no_spa_att		✓	✓	30.6017	0.9675	0.0910	1.457E+09
no_upper	✓	✓	✓	21.0839	0.8422	0.2938	1.385E+09
no_adjacent	✓	✓		22.9937	0.7739	0.1995	**1.328E+09**

B. Hyperparameter Experiments

In this section, we discuss the effects of hyperparameters in the proposed network architecture. There are two important hyperparameters in AODN: the number of adjacent bands K and the low-frequency ratio α in Octave kernels. A random noise level ("$\sigma_n = \text{rand}(100)$") was to perform the test. PSNR, SSIM, and SAM are introduced to evaluate the denoising performance, and FLOPs were employed to evaluate the model complexity. The quantitative results listed in the following tables are the average of the results of ten repeated experiments. The best results are marked in bold.

1) Number of Adjacent Bands K: As we mentioned above, the number of adjacent bands K affects the amount of auxiliary spectral information. A larger K can include more auxiliary spectral information from more bands but raises the computational cost and has a negative impact on the model flexibility. As shown in Table I, when the network had no adjacent band, it requires the least computational resources but performed the worst. As K increases, the FLOPs raise slowly since the computational cost is mainly determined by the denoising module rather than the feature extraction module. As K increases from zero, the quantitative results improve dramatically. The gains become small and tend to be saturated when $K > 64$. Then, as K continues to grow, the performance declines. The possible reason lies in that too many spectral bands make it hard for the network to focus on the current noisy band.

2) Low-Frequency Ratio α: As we mentioned above, the ratio α represents the proportion of the low-frequency feature in an Octave architecture. A larger α could reduce more spatial redundancy in "slow-vary" features and accelerate the training process. When α increases, the model complexity drops significantly, which shows the effectiveness of the Octave kernel. However, larger α also causes more information loss. The balance between high and low frequencies is essential for Octave networks. As shown in Table II, when $\alpha = 0.2$, the network achieves the best results. The network with smaller α achieves less satisfactory results because the redundancy of low frequency

TABLE IV
QUANTITATIVE EVALUATION OF THE DENOISING RESULTS OF THE SIMULATED EXPERIMENTS

σ_n	Index	BM4D	LLRT	HyRes	FastHyDe	HSIDCNN	ENCAM	AODN
25	MPSNR	31.0411 ± 0.0057	33.1928 ± 0.0127	34.2129 ± 0.0114	33.6563 ± 0.0051	33.5952 ± 0.0099	34.3035 ± 0.0116	**34.6831 ± 0.0076**
	MSSIM	0.9683 ± 0.0001	0.9775 ± 0.0001	0.9838 ± 0.0000	0.9814 ± 0.0000	0.9821 ± 0.0001	0.9851 ± 0.0000	**0.9863 ± 0.0000**
	MSAM	0.09089 ± 0.0001	0.0781± 0.0002	0.0678±0.0002	0.0734± 0.0002	0.0680 ± 0.0001	0.0609 ± 0.0001	**0.0584 ± 0.0001**
50	MPSNR	26.6731 ± 0.0100	29.6450 ± 0.0124	30.3021±0.0142	30.9673 ± 0.0073	29.1337 ± 0.0114	30.1075 ± 0.0136	**31.1677 ± 0.0147**
	MSSIM	0.9202 ± 0.0002	0.9487 ± 0.0003	0.9642 ± 0.0001	0.9711 ± 0.0002	0.9544 ± 0.0002	0.9640 ± 0.0002	**0.9718 ± 0.0001**
	MSAM	0.0963 ± 0.0002	0.1032 ± 0.0003	0.0963 ± 0.0003	0.0889 ± 0.0002	0.1013 ± 0.0002	0.0872 ± 0.0002	**0.0777 ± 0.0002**
75	MPSNR	24.1909 ± 0.0068	26.5560 ± 0.0175	27.9773 ± 0.0137	28.4354 ± 0.0131	26.7553 ± 0.0125	27.5770 ± 0.0066	**28.7288 ± 0.0147**
	MSSIM	0.8659 ± 0.0004	0.9288 ± 0.0002	0.9416±0.0003	0.9508 ± 0.0002	0.9257 ± 0.0003	0.9399 ± 0.0002	**0.9532 ± 0.0002**
	MSAM	0.1563 ± 0.0002	0.1213 ± 0.0005	0.1198 ± 0.0006	0.1065 ± 0.0003	0.1244 ± 0.0103	0.1113 ± 0.0003	**0.0959 ± 0.0002**
100	MPSNR	22.5099 ± 0.0077	25.0147 ± 0.0081	26.3669 ± 0.0063	26.5706 ± 0.0140	24.9785 ± 0.0139	25.8177 ± 0.0143	**27.1317 ± 0.0278**
	MSSIM	0.8107 ± 0.0005	0.9066 ± 0.0005	0.9200 ± 0.0004	0.9325 ± 0.0003	0.8915 ± 0.0005	0.9138 ± 0.0003	**0.9354 ± 0.0004**
	MSAM	0.1777 ± 0.0004	0.1396 ± 0.0002	0.1336 ± 0.0007	0.1340 ± 0.0003	0.1408 ± 0.0003	0.1277 ± 0.0003	**0.1102 ± 0.0003**
rand 100	MPSNR	24.8244 ± 0.2022	24.6332 ± 0.2185	30.2293 ± 0.4672	30.431 ± 0.5249	29.1854 ± 0.2304	30.9009 ± 0.2933	**31.1509 ± 0.2478**
	MSSIM	0.8875 ± 0.0037	0.8986 ± 0.0067	0.9659 ± 0.0026	0.9598 ± 0.0128	0.9527 ± 0.0018	0.9671 ± 0.0016	**0.9705 ± 0.0014**
	MSAM	0.1482 ± 0.0021	0.1401 ± 0.0064	0.1115 ± 0.0028	0.1253 ± 0.0036	0.1065 ± 0.0024	0.0895 ± 0.0023	**0.0825 ± 0.0023**
rand *+stripe*	MPSNR	24.4279 ± 0.1232	24.1863 ± 0.1661	29.7561 ± 0.5231	28.9707 ± 0.3915	28.6328 ± 0.2141	29.8859 ± 0.2205	**30.3573 ± 0.2093**
	MSSIM	0.8766 ± 0.0025	0.8857 ± 0.0051	0.9668 ± 0.0013	0.9510 ± 0.0149	0.9472 ± 0.0020	0.9615 ± 0.0015	**0.9682 ± 0.0013**
	MSAM	0.1393 ± 0.0023	0.1489 ± 0.0050	0.1190 ± 0.0035	0.1249 ± 0.0030	0.1154 ± 0.0027	0.0953 ± 0.0020	**0.0904 ± 0.0020**
rand *+deadline*	MPSNR	24.6272 ± 0.2322	24.0456 ± 0.2475	29.4789 ± 0.5607	29.1474 ± 0.5193	28.0792 ± 0.2165	29.5383 ± 0.2558	**30.8666 ± 0.2579**
	MSSIM	0.8732 ± 0.0045	0.8824 ± 0.0065	0.9648 ± 0.0023	0.9489 ± 0.0211	0.9430 ± 0.0021	0.9606 ± 0.0019	**0.9681 ± 0.0014**
	MSAM	0.1432 ± 0.0050	0.1532 ± 0.0071	0.1089 ± 0.0034	0.1107 ± 0.0047	0.1250 ± 0.0025	0.0975 ± 0.0019	**0.0886 ± 0.0022**

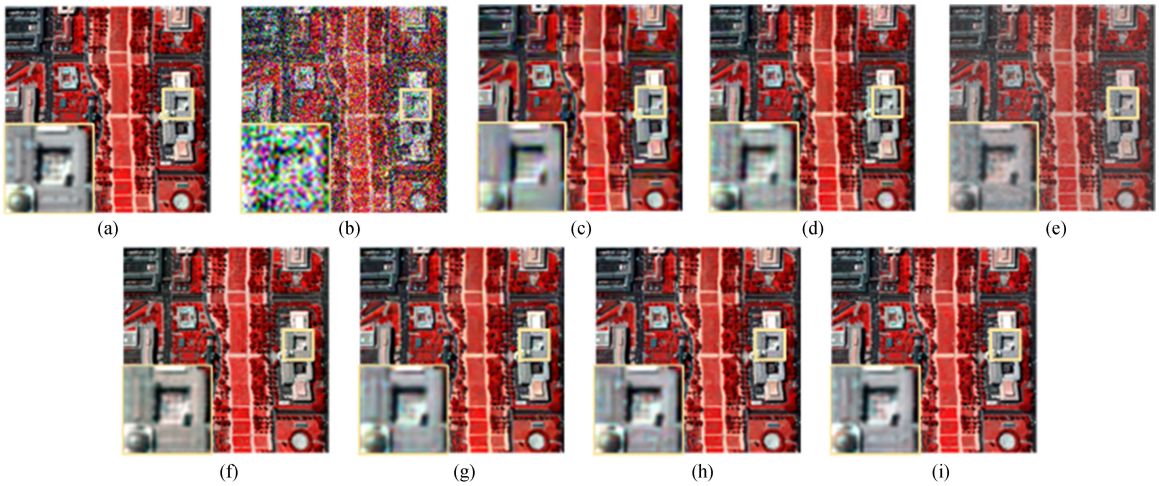


Fig. 6. Results and magnified results for the Washington DC Mall image with $\sigma_n = 100$ in Case 1. (a) Pseudocolor noise-free image with bands (57, 27, 17). (b) Noisy image. (c) BM4D. (d) LLRT. (e) HyRes. (f) FastHyDe. (g) HSIDCNN. (h) ENCAM. (i) AODN.

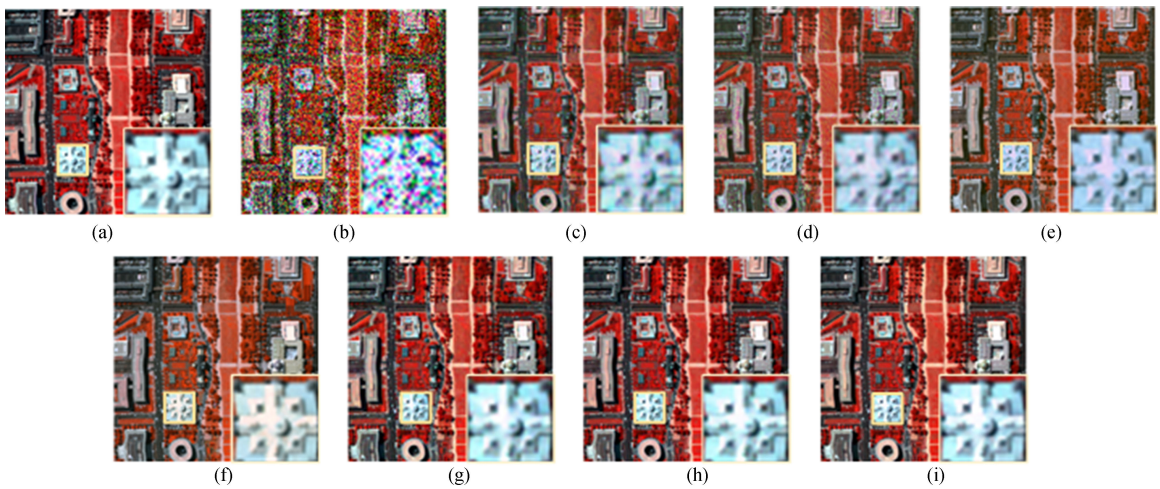


Fig. 7. Results for the Washington DC Mall image with $\sigma_n = \text{rand}(100)$ in Case 2. (a) Pseudocolor noise-free image with bands (57, 27, 17). (b) Noisy image. (c) BM4D. (d) LLRT. (e) HyRes. (f) FastHyDe. (g) HSIDCNN. (h) ENCAM. (i) AODN.

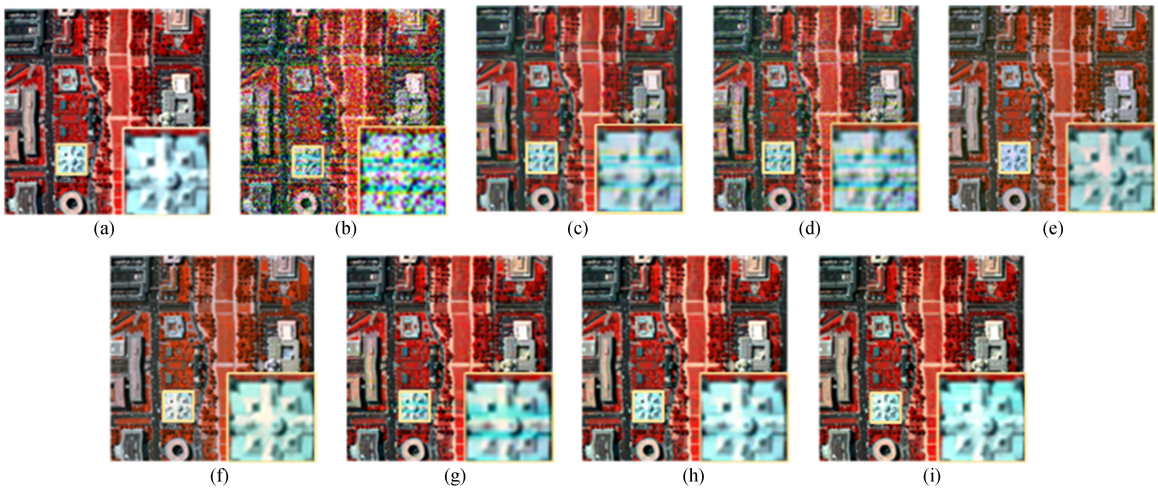


Fig. 8. Results and magnified results for the Washington DC Mall image in Case 3. (a) Pseudocolor noise-free image with bands (57, 27, 17). (b) Noisy image. (c) BM4D. (d) LLRT. (e) HyRes. (f) FastHyDe. (g) HSIDCNN. (h) ENCAM. (i) AODN.

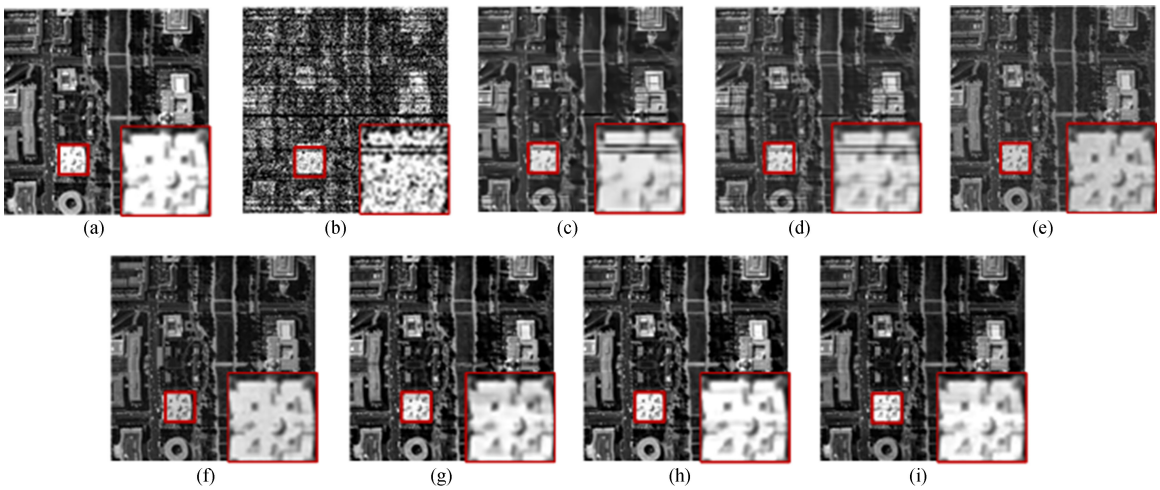


Fig. 9. Results and magnified results for the Washington DC Mall image in Case 4. (a) Real noise-free image band 2. (b) Noisy image. (c) BM4D. (d) LLRT. (e) HyRes. (f) FastHyDe. (g) HSIDCNN. (h) ENCAM. (i) AODN.

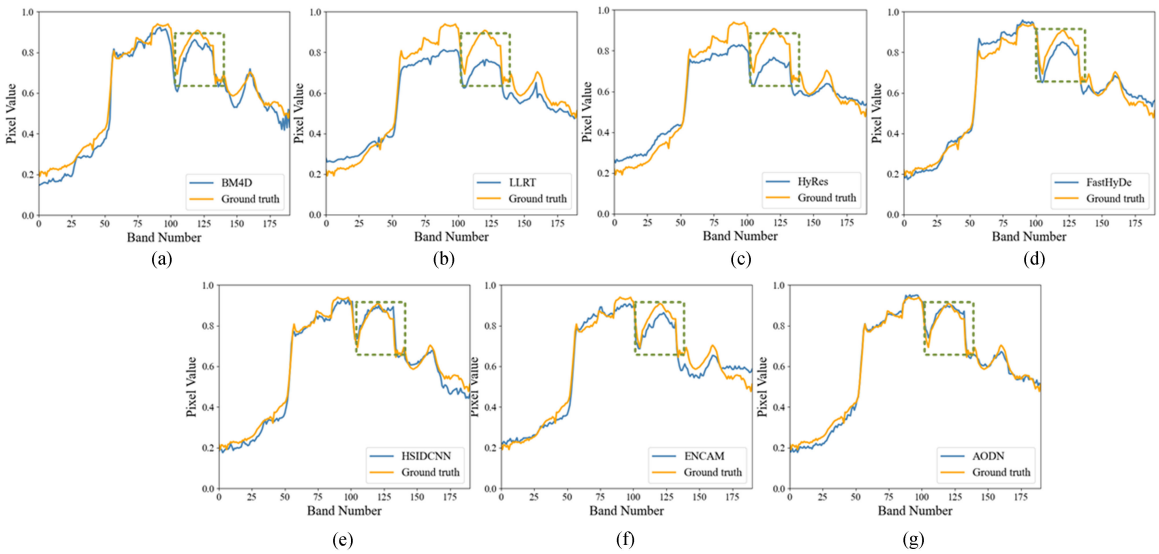
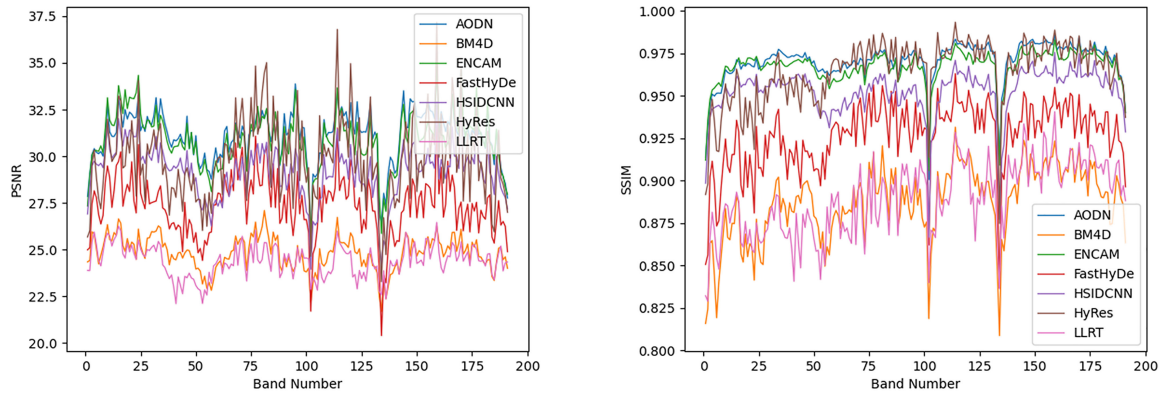


Fig. 10. Spectra of pixel (87,112) in the restoration results with $\sigma_n = 100$ in Case 1. (a) BM4D. (b) LLRT. (c) HyRes. (d) FastHyDe (e) HSIDCNN. (f) ENCAM. (g) AODN.



(a) PSNRs of band result in Case 2

(b) SSIMs of band result in Case 2

Fig. 11. PSNR and SSIM values of the different denoising methods in each band of the simulated experiment with noise level $\sigma_n = \text{rand}(100)$. (a) PSNR values of band result in Case 2. (b) SSIM values of band result in Case 2.

TABLE V
NUMBER OF TRAINING AND TESTING SAMPLES OF THE IP DATASET

No	class	train	test	sum
1	Alfalfa	5	41	46
2	Corn-notill	143	1285	1428
3	Corn-mintill	83	747	830
4	Corn	24	213	237
5	Grass-pasture	49	434	483
6	Grass-trees	73	657	730
7	Grass-pasture-mowed	3	25	28
8	Hay-windrowed	48	430	478
9	Oats	2	18	20
10	Soybean-notill	98	874	972
11	Soybean-mintill	246	2209	2455
12	Soybean-clean	60	533	593
13	Wheat	21	184	205
14	Woods	127	1138	1265
15	Buildings-Grass-Trees-Drives	39	347	386
16	Stone-Steel-Towers	10	83	93

TABLE VI
NUMBER OF TRAINING AND TESTING SAMPLES OF THE PU DATASET

No	class	train	test	sum
1	Asphalt	664	5967	6631
2	Meadows	1865	16784	18649
3	Gravel	210	1889	2099
4	Trees	307	2757	3064
5	Painted metal sheets	135	1210	1345
6	Bare Soil	503	4526	5029
7	Bitumen	133	1197	1330
8	Self-Blocking Bricks	369	3313	3682
9	Shadows	95	852	947

remains in the feature maps, making it hard for the network to focus on the high-frequency features. The results with larger α are worse, which may lie in that when the network reduces the redundancy of low frequency, it also loses more spatial information.

C. Ablation Experiments

In this experiment, we discuss the effectiveness of the AODN structure. There are three main blocks, including the separable

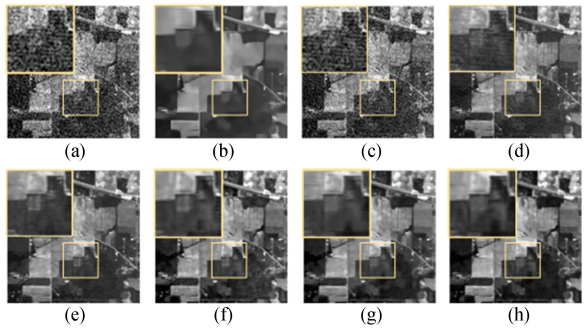


Fig. 12. Results for the IP image. (a) Real image band 2. (b) BM4D. (c) LLRT. (d) HyRes. (e) FastHyDe. (f) HSIDCNN. (g) ENCAM. (h) AODN.

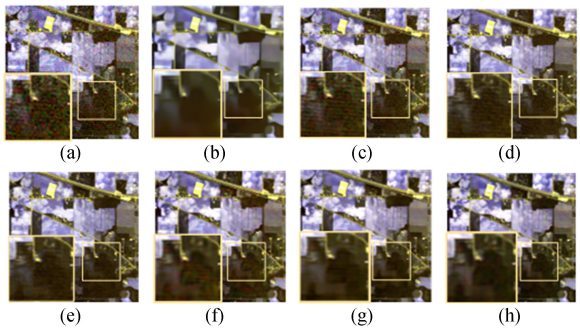


Fig. 13. Results for the IP image. (a) Pseudocolor noisy image with bands (2, 3, 203). (b) BM4D. (c) LLRT. (d) HyRes. (e) FastHyDe. (f) HSIDCNN. (g) ENCAM. (h) AODN.

TABLE VII
CLASSIFICATION ACCURACY RESULTS FOR THE IP DATASET

Index	Noise	BM4D	LLRT	HyRes	FastHyDe	HSIDCNN	ENCAM	AODN
OA	73.93%	88.33%	77.37%	87.94%	85.96%	95.15%	93.22%	**95.66%**
Kappa	0.6995	0.8668	0.7415	0.8628	0.8397	0.9448	0.9227	**0.9506**

TABLE VIII
CLASSIFICATION ACCURACY RESULTS FOR THE PU DATASET

Index	Noise	BM4D	LLRT	HyRes	FastHyDe	HSIDCNN	ENCAM	AODN
OA	91.15%	92.47%	92.33%	91.63%	91.71%	95.20%	96.34%	**97.04%**
Kappa	0.8820	0.8999	0.8977	0.8884	0.8896	0.9363	0.9515	**0.9609**

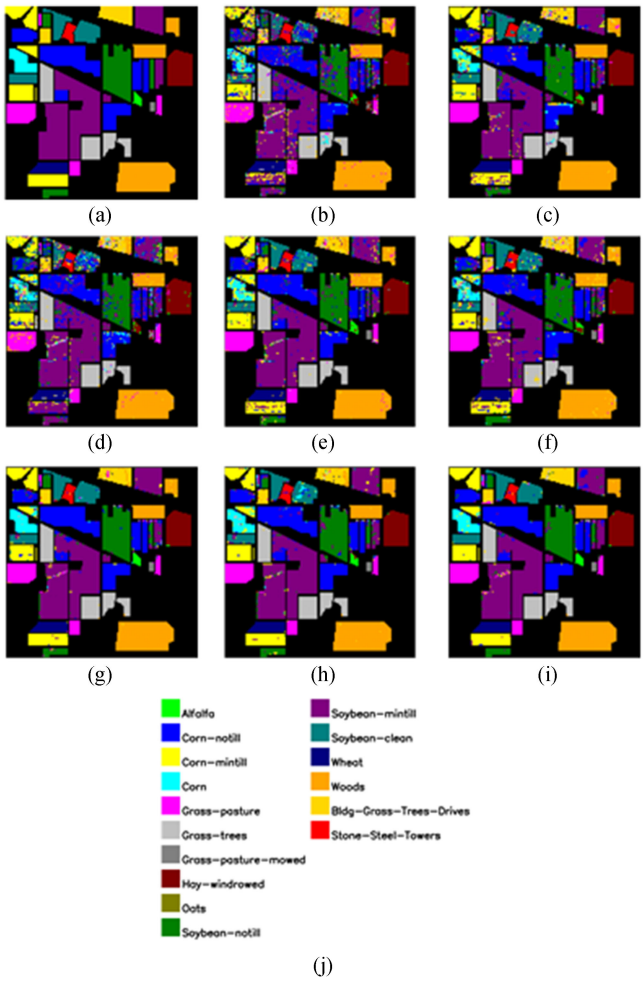


Fig. 14. Classification results for the IP image using SVM before and after denoising. (a) Ground truth. (b) Noisy image. (c) BM4D. (d) LLRT. (e) HyRes. (f) FastHyDe. (g) HSIDCNN. (h) ENCAM. (i) AODN.

feature extraction module, the attention module, and the Octave denoising network. To show the effect of these modules, 11 variants of AODN were evaluated. These variants remove some specific components of the module in the proposed architecture, as shown in Table III. For examples, "no_module" means the model does not have any proposed modules and "no_oct_att" means the model does not contain the Octave kernel and the attention module. The specific modules contained in the variant are marked in the second column of Table III. Especially, "no_cha_att" means no channel attention module in the model, and "no_spa_att" means no spatial attention module in the model. "no_upper" means no upper spatial branch in the model, and "$no\_adjacent$" means no spectral branch in the model. We used a random noise level ("$\sigma_n = \text{rand}(100)$")) to conduct the test. PSNR, SSIM, and SAM were introduced to evaluate the denoising performance, and FLOPs were adopted to evaluate the model complexity. The quantitative results are shown with the mean of ten repeated experiments. The best scores are highlighted in bold.

As shown in Table III, all three modules contribute to the denoising performance. The separable kernel in feature extraction conforms to the HSI structure prior. The attention module can fine-tune feature learning with elementwise multiplication. The Octave module improves denoising results by guiding the architecture to focus on high-frequency features. Without any of the proposed blocks, the quantitative performance declines. If the channel attention or the spatial attention part is removed, the quantitative performance declines obviously.

D. Simulated-Data Experiments

In the simulated experiment, Gaussian noise and mixed noise were simulated according to the following four cases.

1) *Case 1 (Fixed noise level):* In each band, the noise intensities are the same. The noise level σ_n was set from 25 to 100 in sequence, as listed in Table IV.
2) *Case 2 (Unknown noise level):* In different bands, the noise intensities are different. The noise level of each band was generated according to a random probability distribution ("$\sigma_n = \text{rand}(100)$"), as listed in Table IV.
3) *Case 3 (Mixed Gaussian noise and stripes):* All bands in the HSIs were corrupted by Gaussian noise, and some of the bands were corrupted by stripe noise. The strength of Gaussian noise equals to that in Case 2. In our experiments, ten bands of the original data were injected with simulated stripe noise. The number of stripes in each band was set to 5–15% rows.
4) *Case 4 (Mixed Gaussian noise and deadlines):* All bands in the HSIs were corrupted by Gaussian noise, and some of the bands were corrupted by deadlines. The strength of Gaussian noise equals to that in Case 2. In our experiments, ten bands of the original data were injected with simulated deadlines. The number of deadlines in each band was set to 5–15% rows.

In Table IV, averages and standard deviations of each metric for the ten repeated experiments are listed, and the best scores are highlighted in bold. To visualize the denoising results, the cases of noise levels $\sigma_n = 100$ and $\sigma_n = \text{rand}(100)$ are shown in Figs. 6 and 7, and specific regions are enlarged. The cases of mixed noise with stripes and deadlines are shown in Figs. 8 and 9. The spectral curves of pixels (87,112) in Case 1 are plotted in Fig. 10. The results of PSNR and SSIM in different bands in Case 2 are depicted in Fig. 11.

It can be observed that in Table IV, the proposed AODN outperforms all other methods, obtaining notable improvements on all the metrics, which shows that AODN can reconstruct HSIs with higher quality in both spatial and spectral domains.

In fixed-noise-level experiments, all methods achieve good performance, but the proposed AODN outperforms the others. BM4D generates an oversmoothed result. As shown in the enlarged region in Fig. 8, LLRT smooths some edge details, and HyRes suppresses noise well in Case 1 but produces unexpected artifacts. FastHyDe not only removes noise well but also blurs some details. HSIDCNN and ENCAM, which benefit from the merits of supervised learning, achieve promising results. The proposed AODN results in better visualization than the two learning-based methods, indicating the progressiveness of the proposed architecture.

The spectral curve of pixels (87,112) in Case 1 is shown in Fig. 9. BM4D, HyRes, FastHyDe, and ENCAM perform well in noise suppression, but the bands 100–140 have worse

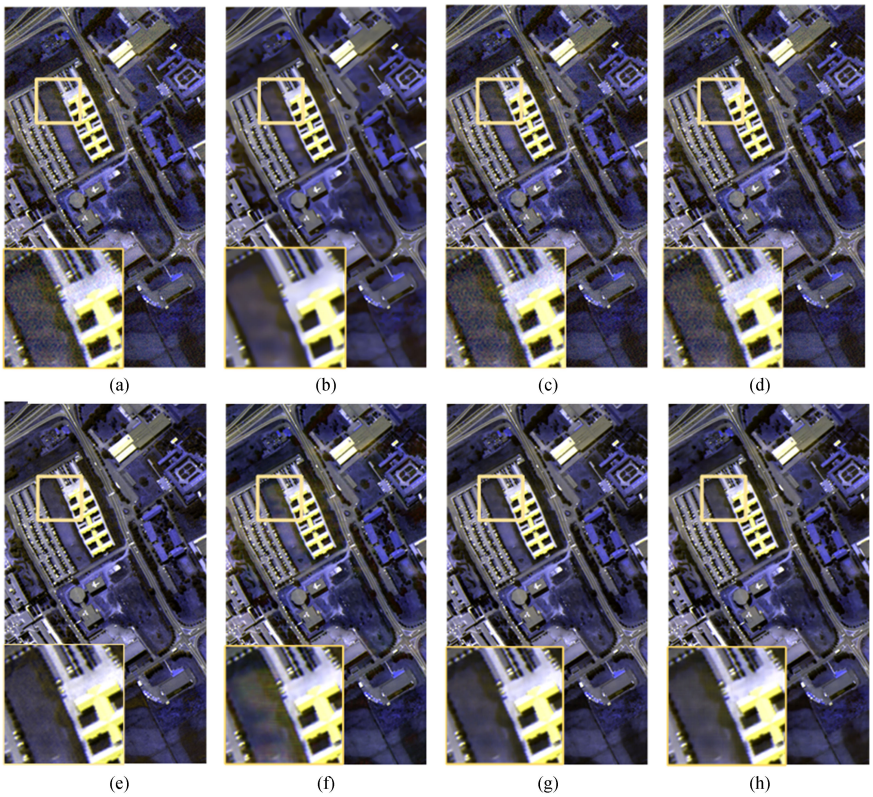


Fig. 15. Results for the PU image. (a) Real image band 2. (b) BM4D. (c) LLRT. (d) HyRes. (e) FastHyDe. (f) HSIDCNN. (g) ENCAM. (h) AODN.

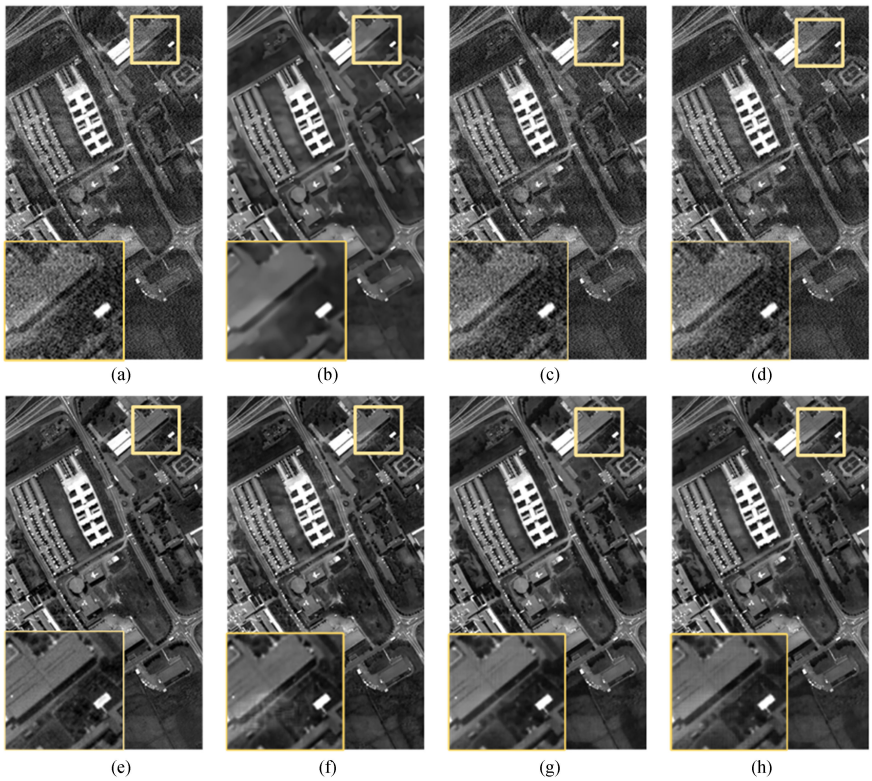


Fig. 16. Results for the PU image. (a) Pseudocolor image with bands (2, 3, 57). (b) BM4D. (c) LLRT. (d) HyRes. (e) FastHyDe. (f) HSIDCNN. (g) ENCAM. (h) AODN.

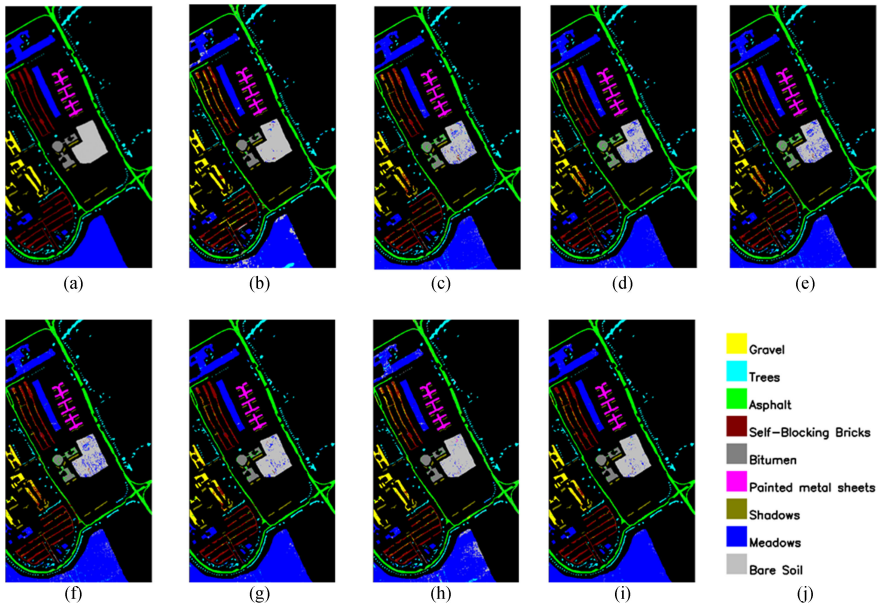


Fig. 17. Classification results for the PU image using SVM before and after denoising. (a) Ground truth. (b) Noisy image. (c) BM4D. (d) LLRT. (e) HyRes. (f) FastHyDe. (g) HSIDCNN. (h) ENCAM. (i) AODN.

results than other bands. LLRT cannot remove noise in the spectral domain well, and some noise still remains in bands 1–50 and 100–140. HSIDCNN performs well at suppressing noise and preserving spectral structural information. However, AODN continues to outperform all other methods, resulting in higher fidelity quality among all bands.

In unknown-noise-level experiments, as shown in Table IV, AODN continues to achieve the best quantitative results among all methods. For visual inspection, in Fig. 10, BM4D suppresses noise relatively well but generates a blurred effect. In the results of LLRT and HyRes, some noise still remains. FastHyDe, HSIDCNN, and ENCAM perform well in blind situations. AODN outperforms all other methods in both edge and detail preservation.

Fig. 11 shows the average PSNR and SSIM values band by band of the repeated experiments of Case 2. When addressing blind denoising problems, BM4D and LLRT cannot suppress noise on heavily corrupted bands. FastHyDe, HSIDCNN, and ENCAM can achieve promising PSNR and SSIM on bands with low noise levels and are relatively better on heavy noise bands. HyRes performs quite well in some bands, but in the front bands, the results are not satisfactory. AODN outperforms all the other methods except HyRes in nearly every band. Compared with HyRes, AODN still shows stable denoing performance.

In Cases 3 and 4, the Gaussian noise is mixed with stripes or deadlines; BM4D, LLRT, and HSIDCNN cannot eliminate stripe noise or deadlines well as shown in the enlarged region in Figs. 8 and 9. Some noise still remains in the result of HyRes. FastHyDe, ENCAM, and AODN perform well in this case.

E. Real-Data Experiments

To validate the flexibility and robustness of AODN, we conducted two real data experiments on the IP and PU datasets. Both datasets are contaminated by real noise in several bands. We applied the model trained on the Washington DC Mall dataset with random noise to both sets. To quantitatively evaluate the results, we utilize the SVM classifier to perform supervised classification with denoised data. The training sets include 10% of the test samples randomly generated from each class. The numbers of training data and testing data for each class on IP and PU datasets are listed in Tables V and VI, respectively. We implemented SVM with Sklearn, and its hyperparameters are set in a set ($c = [0.1, 1, 10, 100, 1000]$, gamma $= [0.1, 1, 10, 100, 1000]$). SVM iterates all the parameter combinations and preserves the best result.

1) IP Dataset: The first few bands and several other bands of the IP HSI are seriously degraded by Gaussian and impulse noise [45]. Figs. 12 and 13 show the results of different methods, which represent band number 2 and the pseudocolor result with combined bands (2, 3, 203), respectively. LLRT and HyRes cannot suppress noise well. BM4D produces oversmoothed and residual strip noise. FastHyDe, HSIDCNN, and ENCAM succeed in denoising, but some information is lost. AODN reconstructs the image with fine details.

In SVM classification experiments, 16 classes are employed to evaluate the classification accuracy. The OA and kappa coefficient are given in Table VII. The original data SVM result and denoised SVM results of the IP dataset are shown in Fig. 14. Due to the residual noise, the results of LLRT and HyRes fail to obtain satisfactory classification results. Although BM4D and FastHyDe produce relatively better results, their classification results suffer from detail loss. HSIDCNN and ENCAM succeed in improving classification performance, but AODN obtains the best results.

In summary, the proposed method obtains the highest OA and kappa values of 95.66% and 0.9506, respectively. The visualization of classification results suggests that AODN has the best denoising results in structure preservation.

TABLE IX
AVERAGE RUNTIME IN THE SIMULATED DATASET WITH DIFFERENT METHODS

Index	BM4D	LLRT	HyRes	FastHyDe	HSIDCNN	ENCAM	AODN
Time(s)	272.760	1487.997	1.7047	**0.9543**	1.185	24.863	5.732

2) PU Dataset: The noise is mainly concentrated in the first band of PU [25]. Figs. 15 and 16 show the denoising results of different methods, which represent band 2 and the pseudocolor result with combined bands (97, 2, 3), respectively. As shown in Figs. 15 and 16, all the other methods fail to achieve either noise suppression or edge preservation, while AODN succeeds in recovering details from noisy observations.

In SVM classification experiments, nine classes are employed to test the classification accuracy. The OA and kappa coefficient are given in Table VIII. The original data SVM result and denoised SVM results of the PU dataset are shown in Fig. 17. Again, the proposed method result has the highest OA and kappa index values of 92.30% and 0.8993, respectively. Denoising results also indicate that AODN reconstructs HSIs with better edge and detail information.

F. Runtime Discussion

In this section, we discuss the efficiency of the proposed denoising methods. Table IX presents the average runtime in the simulated experiments. Deep-learning-based methods exhibit less runtime than the traditional methods, such as BM4D and LLRT, with the benefits of GPUs and end-to-end structures. However, low-rank methods, such as HyRes and FastHyDe, obtain relatively low computational costs and, thus, have shorter runtime. Among deep learning methods, AODN is relatively slower than HSIDCNN due to its deeper network and attention module, which leads to higher computational costs. The relatively low runtime complexity and high-quality results indicate that AODN is a cost-effective end-to-end architecture.

V. CONCLUSION

In this article, we present an AODN for HSI denoising. The proposed separative convolution feature extraction model can extract spatial–spectral features, which are fit to the HSI data structure prior. The attention module guides feature tuning in both spectral and spatial domains. The Octave kernel reduces low-frequency redundancy, which makes the architecture focus on high-frequency noise features and reduces the computational cost. We conducted hyperparameter experiments to choose parameters and ablation experiments to demonstrate the impacts of each proposed module. Simulated and real-world experiments indicate that AODN outperforms several state-of-the-art methods in both quantitative and qualitative aspects. Finally, the runtime is discussed, which shows that our network is cost-effective.

REFERENCES

[1] Z. Kan, S. Li, and Y. Zhang, "Attention based octave dense network for hyperspectral image denoising," in *Proc. IEEE Int. Conf. Big Data Artif. Intell.*, 2021, pp. 230–235.

[2] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," in *Proc. IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 72021, no. 6, pp. 2094–2107, Jun. 2014.

[3] X. Lu, H. Wu, Y. Yuan, P. Yan, and X. Li, "Manifold regularized sparse NMF for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2815–2826, Mar. 2013.

[4] R. Zhao, B. Du, and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1227–1234, Apr. 2014.

[5] 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.

[6] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 2862–2869.

[7] D. Zoran and Y. Weiss, "From learning models of natural image patches to whole image restoration," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 479–486.

[8] I. Atkinson, F. Kamalabadi, and D. L. Jones, "Wavelet-based hyperspectral image estimation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2003, vol. 2, pp. 743–745.

[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.

[10] 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.

[11] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 6868–6877.

[12] T. Lu, S. Li, L. Fang, Y. Ma, and J. A. Benediktsson, "Spectral-spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, Jan. 2016.

[13] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Noise removal from hyperspectral image with joint spectral-spatial distributed sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5425–5439, Sep. 2016.

[14] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[15] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[16] H. Wei, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2015.

[17] 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. 4260-.4268.

[18] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using 3D wavelets," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2012, pp. 1349–1352.

[19] B. Rasti, J. R. Sveinsson, M. O. Ulfarsson, and J. A. Benediktsson, "Hyperspectral image denoising using first order spectral roughness penalty in wavelet domain," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2458–2467, Jun. 2014.

[20] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.

[21] H. Zhang, "Hyperspectral image denoising with cubic total variation model," *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 7, pp. 95–98, 2012.

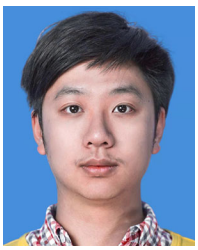
[22] B. Rasti, M. O. Ulfarsson, and P. Ghamisi, "Automatic hyperspectral image restoration using sparse and low-rank modeling," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 12, pp. 2335–2339, Dec. 2017.

[23] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 730–742, Mar. 2018.

[24] H. Zhang, L. Liu, W. He, and L. Zhang, "Hyperspectral image denoising with total variation regularization and nonlocal low-rank tensor decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3071–3084, May 2020.

[25] 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.

[26] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, vol. 29, pp. 2802–2810.
[27] 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.
[28] H. Ma, G. Liu, and Y. Yuan, "Enhanced non-local cascading network with attention mechanism for hyperspectral image denoising," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2020, pp. 2448–2452.
[29] Y. Zhao, D. Zhai, J. Jiang, and X. Liu, "ADRN: Attention-based deep residual network for hyperspectral image denoising," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 2668–2672.
[30] W. Dong, H. Wang, F. Wu, G. Shi, and X. Li, "Deep spatial-spectral representation learning for hyperspectral image denoising," *IEEE Trans. Comput. Imag.*, vol. 5, no. 4, pp. 635–648, Dec. 2019.
[31] J. Song, J.-H. Jeong, D.-S. Park, H.-H. Kim, D.-C. Seo, and J. C. Ye, "Unsupervised denoising for satellite imagery using wavelet directional CycleGAN," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6823–6839, Aug. 2021.
[32] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu, "Residual non-local attention networks for image restoration," 2019, *arXiv:1903.10082*.
[33] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2021.
[34] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial-spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, Oct. 2019.
[35] W. Liu and J. Lee, "A 3-D atrous convolution neural network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5701–5715, Aug. 2019.
[36] 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.
[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. 3435–3444.
[38] Q. Xu, Y. Xiao, D. Wang, and B. Luo, "CSA-MSO3DCNN: Multiscale octave 3D CNN with channel and spatial attention for hyperspectral image classification," *Remote Sens.*, vol. 12, no. 1, 2020, Art. no. 188.
[39] X. Tang *et al.*, "Hyperspectral image classification based on 3-D octave convolution with spatial-spectral attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2430–2447, Mar. 2021.
[40] Q. Xu, D. Wang, and B. Luo, "Faster multiscale capsule network with octave convolution for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 2, pp. 361–365, Feb. 2021.
[41] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[42] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.
[43] 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.
[44] R. H. Yuhas, J. W. Boardman, and A. F. Goetz, "Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques," in *Proc. 4th Annu. JPL Airborne Geosci. Workshop*, 1993, pp. 205–208.
[45] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, and G. He, "Denoising of hyperspectral images using nonconvex low rank matrix approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5366–5380, Sep. 2017.



Ziwen Kan is currently working toward the bachelor's degree in software engineering with Sichuan University, Chengdu, China.

His research interests include machine/deep learning, computer vision, and hyperspectral image processing.

Suhang Li is currently working toward the bachelor's degree in software engineering with Sichuan University, Chengdu, China. He is also currently working toward the Graduate degree with the Graduate School of Information Production and Systems, Waseda University, Tokyo, Japan.

His research interests include machine/deep learning, computer vision, image denoising, and salient object detection.

Mingzheng Hou received the B.S. and M.S. degrees from the College of Computer Science, in 2010 and 2013, respectively, Sichuan University, Chengdu, China, where she is currently working toward the Ph.D. degree.

Her main research interests include computer vision, image super-resolution, and target detection.

Leyuan Fang (Senior Member, IEEE) received the Ph.D. degree from the College of Electrical and Information Engineering, Hunan University, Changsha, China, in 2015.

From September 2011 to September 2012, he was a visiting Ph.D. student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. From August 2016 to September 2017, he was a Postdoctoral Researcher with the Department of Biomedical Engineering, Duke University. He is currently a Professor with the College of Electrical and Information Engineering, Hunan University, and an Adjunct Researcher with Peng Cheng Laboratory, Shenzhen, China. His research interests include sparse representation and multiresolution analysis in remote sensing and medical image processing.

Dr. Fang was a recipient of one Second-Grade National Award at the Nature and Science Progress of China in 2019. He is an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *Neurocomputing*.

Yi Zhang (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the College of Computer Science, Sichuan University, Chengdu, China, in 2005, 2008, and 2012, respectively.

From 2014 to 2015, he was a Postdoctoral Researcher with the Department of Biomedical Engineering, Rensselaer Polytechnic Institute, Troy, NY, USA. He is currently a Full Professor with the College of Computer Science, Sichuan University, where he is the Director of the Deep Imaging Group. He authored more than 70 papers in the field of image processing. These papers were published in several leading journals, including IEEE TRANSACTIONS ON MEDICAL IMAGING, IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING, MEDICAL IMAGE ANALYSIS, *European Radiology*, and *Optics Express*, and reported by the Institute of Physics and during the Lindau Nobel Laureate Meeting. His research interests include medical imaging, compressive sensing, and deep learning.

Dr. Zhang was a recipient of major funding from the National Key R&D Program of China, the National Natural Science Foundation of China, and the Science and Technology Support Project of Sichuan Province, China. He is a Guest Editor for *International Journal of Biomedical Imaging* and *Sensing and Imaging* and an Associate Editor for IEEE ACCESS.