Multiscale Adaptive Fusion Network for Hyperspectral Image Denoising

Haodong Pan, Feng Gao ⓘ, *Member, IEEE*, Junyu Dong ⓘ, *Member, IEEE*, and Qian Du ⓘ, *Fellow, IEEE*

Abstract—Removing the noise and improving the visual quality of hyperspectral images (HSIs) is challenging in academia and industry. Great efforts have been made to leverage local, global, or spectral context information for HSI denoising. However, existing methods still have limitations in feature interaction exploitation among multiple scales and rich spectral structure preservation. In view of this, we propose a novel solution to investigate the HSI denoising using a multiscale adaptive fusion network (MAFNet), which can learn the complex nonlinear mapping between clean and noisy HSI. Two key components contribute to improving the HSI denoising: A progressively multiscale information aggregation network and a coattention fusion module. Specifically, we first generate a set of multiscale images and feed them into a coarse-fusion network to exploit the contextual texture correlation. Thereafter, a fine fusion network is followed to exchange the information across the parallel multiscale subnetworks. Furthermore, we design a coattention fusion module to adaptively emphasize informative features from different scales, and thereby enhance the discriminative learning capability for denoising. Extensive experiments on synthetic and real HSI datasets demonstrate that the proposed MAFNet has achieved a better denoising performance than other state-of-the-art techniques.

Index Terms—Deep learning, fusion module, global contextual correlation, hyperspectral image (HSI) denoising, neural network.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) have developed rapidly with the maturity of remote sensing technology. HSIs have been extensively applied in land cover classification [1], [2], [3], [4], semantic segmentation [5], change detection [6], [7], [8], [9], [10], [11], oil spill monitoring [12], [13], [14], and geographic transport prediction [15]. In these applications, high-quality images are commonly desired. However, during the HSI acquisition, some noise corruptions are inevitable and degrade the visual quality considerably [16]. Hence, removing noise from the acquired HSI is a critical step for many remote sensing applications [17]. The task of highly efficient HSI denoising has recently captured numerous research attention.

The HSI denoising task aims to recover an underlying clean image I from a noise observed data I_N. The degradation model is commonly formulated as $I_N = I + I_N^*$. Here, I_N^* denotes the mixed noise. To solve the ill-posed inverse problem, many methods model the prior knowledge of the clean image I to constrain the solution space. Total variation [18], [19], [20], [21], sparsity-driven models [22], [23], [24], and low-rank representations [25], [26], [27], [28] are commonly used. The intrinsic structures in HSIs are modeled by these optimization-based models for noise removal.

During the past few years, the deep learning-based model for low-level vision tasks has demonstrated significant potential and performance improvement. It has been extensively employed to image restoration tasks, such as compression artifact reduction [29], image denoising [30], [31], and image super-resolution [32], [33]. In HSI denoising, the deep learning-based model has yielded excellent results recently. Deep convolutional neural networks (CNNs) are capable of exploiting rich feature representations from large-scale training data, instead of hand-crafted features, which are designed according to prior knowledge. Most existing CNN-based HSI denoising methods follow high-resolution feature processing [34], [35], [36]. These methods do not employ any downsampling operation; and hence, more accurate spatial details can be retained in the denoising results. However, the contextual information is inclined to get lost due to the limited receptive field. To effectively encode the contextual information, some researchers employ an encoder–decoder architecture [37], [38]. The input HSI is progressively mapped into a low-resolution representation and then gradually mapped to the original resolution. The broad context can be learnt in the low-resolution representation [39]. However, some fine spatial details are hard to be reserved. Such detailed information is hard to be recovered in the decoding stages.

It is crucial to encode the contextual information and preserve the spatial details simultaneously for robust HSI denoising. However, it is a nontrivial task due to the following challenges. 1) *Tradeoff between the spectral–spatial detail preservation and contextual information modeling*. HSI denoising creates pixel-to-pixel correspondence from the observed data with complex noise to the clean image, and it is essential to preserve the detailed spectra and texture via high-resolution feature processing networks. However, the contextual information is hard to be encoded while preserving the spatial details by existing models. Hence, how to preserve the detailed spectra and texture whiles
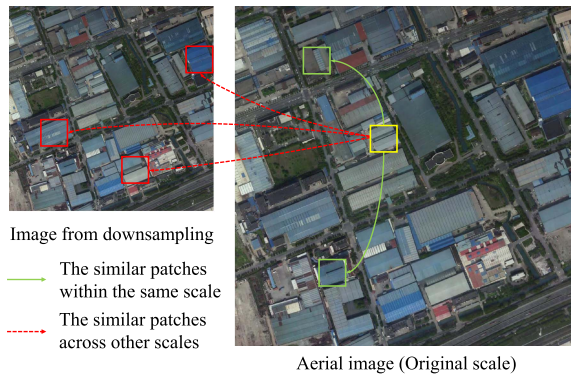
Fig. 1. Illustration of the multiscale representation of high-resolution remote sensing image. Specially, there are similar patches in different scales, both in the same scale (highlighted in green) and across different scales (highlighted in red), can make contribution as additional information to reconstruct the target area (highlighted in blue).

encoding the contextual information effectively is of great significance. 2) *Multiscale information aggregation*. Image contents from multiple scales encode complementary information for feature representation, as shown in Fig. 1. Existing multiscale models rarely exchange information across different scales flexibly. The correlations among different scales have not been fully exploited. Therefore, how to aggregate multiscale information into a unified framework is a tricky task. Wang et al. [40] proposed a deep high-resolution representation network (HRNet) for visual recognition. It maintains high-resolution representations in the whole network and repeatedly exchanges information among multiresolution features. It has been widely used for human pose estimation [41], semantic segmentation [42], [43], [44], and multispectral image classification [45]. If such framework could be introduced into HSI restoration, the denoising performance could be further improved.

To solve the aforementioned issues, we proposed a multiscale adaptive fusion network (MAFNet) for HSI denoising. The framework of MAFNet is illustrated in Fig. 2. Specifically, we first generate a set of multiscale images and feed them into a coarse-fusion network to exploit the contextual texture correlation. Meanwhile, a fine fusion network is followed to exchange information across the parallel multiscale subnetworks. Furthermore, we design a coattention fusion module to adaptively emphasize informative features from different scales, and thereby enhance the discriminative learning capability for denoising. We next adopt a reconstruction loss together with a global gradient regularization to optimize the network. We conduct extensive experiments on five publicly available datasets. The experimental results show that the proposed MAFNet outperforms several state-of-the-art baselines. Our MAFNet differs from HRNet in two respects: First, in the coarse-fusion network, the information interaction direction is only from the low-resolution features to the high-resolution features. In this stage, we aim to increase the receptive field to capture more content. Second, in multiscale feature fusion, HRNet transforms features from different resolutions to the same size and then concatenates them together as the fusion output. The proposed MAFNet use adaptive instance (AIN) and coattention mechanism for adaptive feature fusion.

The contributions of this work can be summarized as follows.

1) We propose a novel HSI denoising model MAFNet, which progressively fuses multiscale information. Hence, the global contextual information modeling and spatial detail preservation can be achieved simultaneously.
2) We present a coattention fusion module to dynamically select the useful features from each scale subnetwork, enhancing the discriminative learning capability. Thereby, multiscale information is adaptively aggregated, and the correlations among different scales are concurrently enhanced.
3) Extensive experiments are conducted on two benchmark datasets, which demonstrates the rationality and effectiveness of the proposed MAFNet. Meanwhile, we have released our codes to benefit the remote sensing image restoration community.

The rest of this article is organized as follows. In Section II, we review closely related HSI denoising methods. The details of MAFNet are described in Section III. Experiments on several datasets on HSI datasets are presented in Section IV. Finally, Section V concludes this article.

II. RELATED WORK

The HSI denoising is an essential step to improve the image quality before interpretation. To date, a great number of methods have been proposed to reduce the noise in HSIs. In this article, the existing methods are classified into three categories and introduced, respectively.

A. Filter-Based Methods for HSI Denoising

In the beginning, the filter operator is generally used for HSI denoising, and it aims to separate the clean image from noisy signals by nonlocal means filter or Fourier transform. Othman et al. [46] presented a wavelet shrinkage method for HSI denoising. The method benefits from the feature dissimilarity between the spectral and spatial domains. Zelinski et al. [47] proposed a method based on wavelet decomposition and sparse approximation for HSI denoising, thereby exploiting the correlation between bands and higher quality band information. Maggioni et al. [48] proposed BM4D for volumetric data denoising, which is an extension of the BM3D filter. It embeds the grouping and collaborative filtering paradigms, thus integrating spatial and frequency domain filtering. Letexier and Bourennane [49] used the multidimensional Wiener filter for HSI denoising. Quadtree decomposition is also utilized to keep local characteristics. These filter-based methods are sensitive to the transform function and, therefore, can hardly remove the mixed noise in HSIs.

B. Model-Based Methods for HSI Denoising

The model-based method is the most popular representation tool for HSI denoising. Total variation [18], [19], [20], [21], sparsity-driven models [22], [23], [24], and low-rank representations [25], [26], [27], [28] are commonly employed to establish an optimization model for HSI denoising. For instance, Yuan et al. [18] presented an adaptive total variation model for HSI denoising, which simultaneously models the spectral and spatial
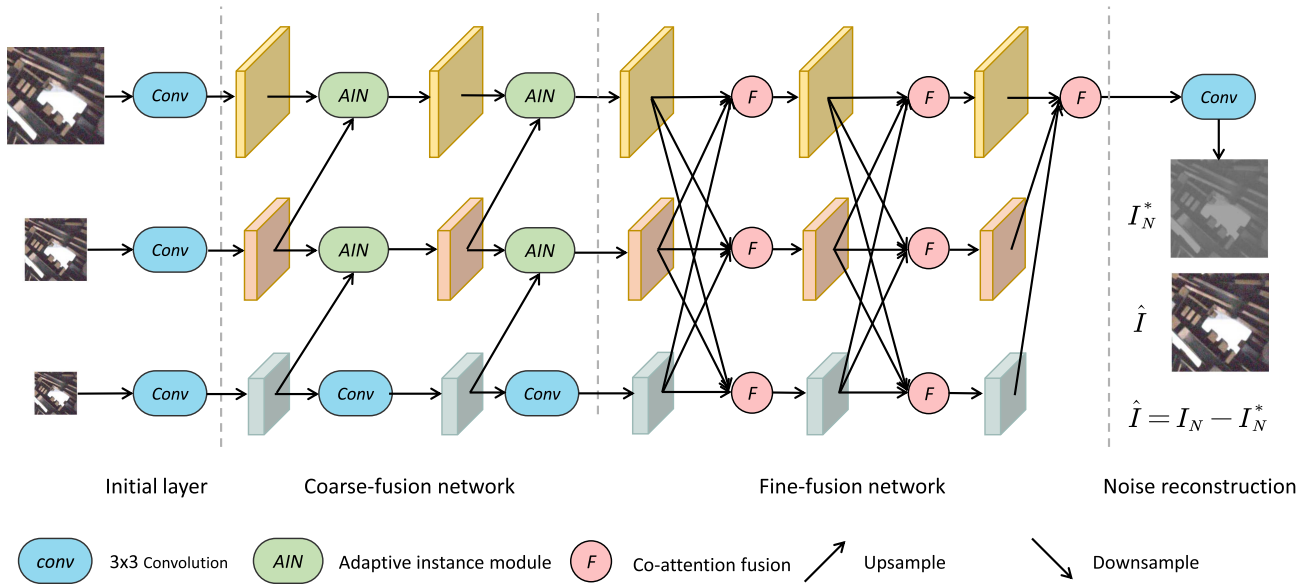
Fig. 2. Framework of the MAFNet. MAFNet is composed with initial layer, coarse-fusion network, fine-fusion network, and noise reconstruction. Initial layer obtains the feature representation from the HSI image, then coarse-fusion network realizes the information transfer from the low-scale network to the high-scale network, then fine-fusion network fully integrates the contextual global information, and finally we reconstruct noise to get denoising images. For upsampling, we use transposed convolution to get the hight-resolution representation of the feature.

noise distribution. Zhang et al. [20] proposed an HSI denoising method based on nonlocal low-rank tensor decomposition, in which the nonlocal similarity between the data cubes is captured to build a clean image. Zhuang et al. [50] proposed a denoising method that exploited the low-rank structure of the HSI data, utilizes the low-rank and self-similar characteristics contained in HSI for sparse and compact representation. Cao et al. [24] presented a subspace-based nonlocal low-rank method for HSI denoising. These methods achieve a satisfying performance due to the comprehensive consideration of the image's prior information.

C. CNN-Based Methods for HSI Denoising

Recently, research on natural image restoration has been dominated by the deep CNN in recent years [29], [30], [31], [32], [51], [52], [53]. Chang et al. [34] introduce the deep CNN model for HSI denoising, uses residual learning, dilated convolution, and multichannel filtering to enhance the ability to express spectral features. Liu et al. [35] presented a 3-D atrous convolution method for HSI denoising. Atrous convolution was employed to enlarge the receptive fields. Zhang et al. [54] employed the gradient learning strategy to capture the intrinsic and deep features of HSI. Cao et al. [55] proposed two global reasoning modules to exploit the contextual information along the channel and spatial dimensions, respectively. Both modules are combined in dense CNNs to exploit rich feature representations. Lin et al. [56] built a CNN-constrained nonnegative matrix factorization model for HSI denoising, realizes the optimization of noisy images through three stages: update of the spectral matrix, update of the abundance matrix, and estimation of the sparse noise. Wei et al. [57] designed QRNN3D, the model adopts the encoder–decoder structure, realizes the use of spatial

pixel correlation information and spectral global information by building a three-dimensional recurrent unit, and uses the special structure of alternating directions to eliminate unreasonable causal dependencies. Capable of flexible processing of HSIs.

Leveraging the powerful linear modeling capability of the deep CNN, these methods achieved promising performance on HSI denoising. However, existing CNN-based methods rarely build feature communications between cascaded multiscale layers; thus, the correlated noise information across different scales is not fully exploited.

III. METHODOLOGY

In this section, we present the MAFNet for HSI denoising, which exploits the inherent correlation of noise across multiple scales. As illustrated in Fig. 2, MAFNet consists of the following four parts: initial layer, coarse-fusion network, fine-fusion network, and noise reconstruction. Four parts work together to estimate the noise image I_N^*. The noise-free data is generated by subtracting I_N^* from the observation data Y. The details of each part of the network are presented in the following.

A. Coarse-Fusion Network

For a given input HSI, the proposed model first downsamples the input image into 1/2 and 1/4 scales by using Gaussian kernels. Shallow features are extracted by multiple parallel convolutions, as illustrated in Fig. 2 (the initial layer). Next, the coarse-fusion network extracts deep features and fuses the multiscale information through several parallel AIN modules. The motivations for designing the coarse-fusion network are twofold: 1) The multiscale structure presents a solution to increase the receptive field to capture more content. 2) The AIN module can transfer the basic structures from the low-resolution feature maps to
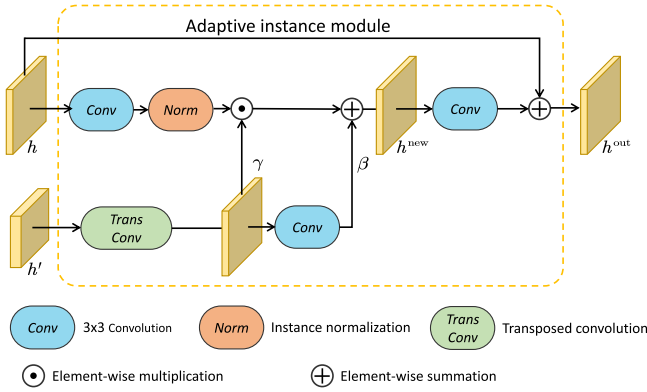
Fig. 3. Illustration of the proposed AIN module.

Fig. 4. Illustration of multiscale features aggregation from left to right, respectively.

the high-resolution ones. We choose AIN normalization [58] to build the AIN module due to its efficiency and compact representation. Fig. 3 shows the architecture of the AIN module.

The AIN normalization affine transforms the normalized feature map $h \in \mathbb{R}^{H \times W \times C}$ by taking an input $h' \in \mathbb{R}^{\frac{H}{2} \times \frac{W}{2} \times 2\,C}$. Here, H and W denote the height and width of the feature map, respectively. C is the number of channels. It should be noted that h' is the feature from the downscale. Specifically, the AIN normalization takes the current feature h and the downscale feature h' as input. First, we convert h' to the size of $H \times W \times C$ by transposed convolution to be consistent with the same dimension as h. Afterwards, for the purpose of using the contextual semantic information contained in the downscale features, we obtain the affine transform parameters from the transformed h' for each pixel (shift β and scale γ). Every feature map is pixel-wise affine transformed and channel-wise normalized, as illustrated in Fig. 3. The updated value in the feature map at position (i, j, c) can be formally represented as

$$h_{i,j,c}^{\text{new}} = \gamma_{i,j,c} \left(\frac{h_{i,j,c} - \mu_c}{\sigma_c} \right) + \beta_{i,j,c} \qquad (1)$$

where μ_c denotes the mean of h in channel c, σ_c denotes the standard deviation of features in channel c. To be more specific, μ_c is computed as

$$\mu_c = \frac{1}{HW} \sum_{i}^{H} \sum_{j}^{W} h_{i,j,c}. \qquad (2)$$

σ_c is computed as

$$\sigma_c^2 = \frac{1}{HW} \sum_{i}^{H} \sum_{j}^{W} (h_{i,j,c} - \mu_c)^2. \qquad (3)$$

It should be noted that $\gamma_{i,j,c}$ and $\beta_{i,j,c}$ are generated pixel-wisely from h'. Therefore, the images with spatially variant noise can be handled adaptively. Finally, a convolutional layer is applied on h^{new}, and residual connection is used to better transfer feature information.

B. Fine-Fusion Network

The outputs of the coarse-fusion network are fed into the fine-fusion network to refine the information from multiple
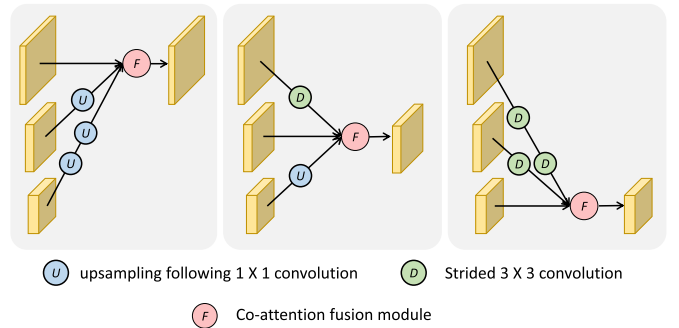
scales. It is well known in cognitive science that in the primate visual cortex, the local receptive fields of neurons are of different sizes. Hence, the capability of collecting multiscale information should be taken into account in deep networks.

Inspired by HRNet [40], we conduct repeated multiscale fusion by exchanging information across parallel multiresolution subnetworks. Furthermore, we design a coattention fusion module to adaptively emphasize informative features from different scales, and therefore, enhance the discriminative learning capability of the network for image denoising.

As illustrated in Fig. 2, the fine-fusion network starts from multiscale feature representations $\{\mathbf{X}_r^1, r = 1, 2, 3\}$, where r denotes the spatial resolution index. In the second layer, the feature representations are $\{\mathbf{X}_r^2, r = 1, 2, 3\}$. Each feature representation is computed as

$$\mathbf{X}_r^2 = \text{CA}(f(\mathbf{X}_1^1), f(\mathbf{X}_2^1), f(\mathbf{X}_3^1)) \qquad (4)$$

where CA is the coattention fusion module. $f(\cdot)$ is a transform function. The transform function $f(\cdot)$ depends on the input and output spatial resolution of the feature map. As depicted in Fig. 4, the strided 3×3 convolution is employed for $2\times$ downsampling. Two consecutive strided 3×3 convolutions are employed for $4\times$ downsampling. At the same time, the nearest neighbor sampling following a 1×1 convolution is used for upsampling.

If the input and the output have the same resolution, we adopt the identity connection. Note that after transformation, the feature maps from different scales are of the same size, and they are fed into the coattention fusion module.

In deep neural networks, features from different states or sources contribute differently to the feature representations [59]. In HRNet, multiscale features are directly fused by element-wise summation. We argue that simply combining multiscale features by concatenation or summation lacks the flexibility to modulate these features, and the discriminative ability of deep models will be influenced. Therefore, this article proposes a coattention fusion module to adaptively emphasize the important information from different scales. The structure of the coattention fusion module is sketched in Fig. 5, which consists of two parts: 1) concatenation and split, 2) fusion and self-calibration.

1) Concatenation and Split: The coattention fusion module receives multiscale features and generates trainable weights for feature fusion. Given input features $\mathbf{Y}_1$, $\mathbf{Y}_2$, and $\mathbf{Y}_3$ are with
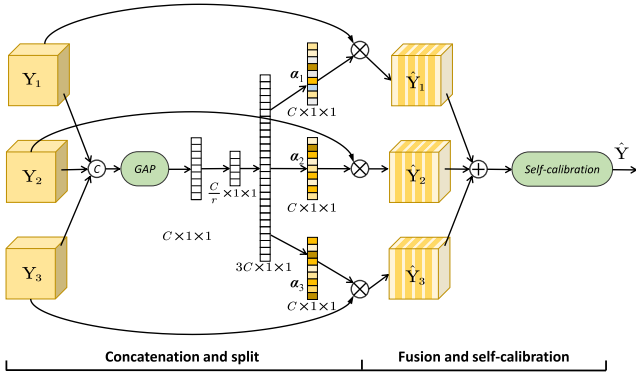
Fig. 5.   Illustration of the proposed coattention fusion module.

the size of $C \times H \times W$, we first conduct the concatenation operation on three features

$$\mathbf{U} = \mathrm{cat}(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3) \tag{5}$$

where $\mathrm{cat}(\cdot)$ is the concatenation operation. Next, The global average pooling is used to compute the channel-wise statistics $\mathbf{s} \in \mathbb{R}^{3C \times 1 \times 1}$ along the spatial dimension of $\mathbf{U} \in \mathbb{R}^{3\,C \times H \times W}$. A downsampling convolution layer is employed to produce a compact feature $\mathbf{u} \in \mathbb{R}^{\frac{3C}{r} \times 1 \times 1}$. Here, $r = 4$ is used in our experiments.

Ultimately, the feature $\mathbf{u}$ is passed through three parallel upsampling layers and provides us with three feature descriptors u_1, u_2, and u_3 each with dimension $C \times 1 \times 1$. Softmax function is applied to u_1, u_2, and u_3, yielding three attention activation vectors α_1, α_2, and α_3, respectively.

2) Fusion and Self-Calibration: The three attention activation vectors generated will be used to recalibrate the input features as

$$\tilde{\mathbf{Y}} = \alpha_1 \cdot \mathbf{Y}_1 + \alpha_2 \cdot \mathbf{Y}_2 + \alpha_3 \cdot \mathbf{Y}_3. \tag{6}$$

Then, we adjust and integrate the features, which are performed by the self-calibration module

$$\hat{\mathbf{Y}} = H_{\mathrm{sc}}(\tilde{\mathbf{Y}}) \tag{7}$$

where $H_{\mathrm{sc}}(\cdot)$ denotes the self-calibrated convolution [60].

As a follow-up operation after fusion, self-calibration convolution uses the convolution filters to operate on the fused feature map to enhance the feature representation ability.

Conclusively, the proposed coattention module transforms the input features into compact descriptors and generates three sets of weights to model channel-wise interdependencies. In this way, the coattention module can adaptively emphasize the important information from multiscale and generate trainable weights for representative feature fusion.

C. Denoising and Reconstruction

At the end of the fine-fusion network, multiscale features are fused by the coattention fusion module. Then, one convolution layer is employed to learn the residual noise image I_N^*. Finally, the noise-free image $\hat{I}$ is computed by subtracting I_N^* from the observation I_N.

We use L_1 loss to optimize our network, and the reconstruction loss is

$$\mathcal{L}_{\mathrm{rec}} = \|\hat{I} - I\|_1 \tag{8}$$

where $\hat{I}$ denotes the estimated noise-free HSI, and I denotes the real noise-free HSI. While in the HSI case, hundreds of bands with abundant spectral information means the noise types and intensity in each band are usually different. Therefore, differences in spatial and spectral direction can provide additional complementary contributions for denoising. We introduce a global gradient regularizer to constrain the details of $\hat{I}$

$$\mathcal{L}_{\mathrm{grad}} = \|\nabla_h \hat{I} - \nabla_h I\|_2^2 + \|\nabla_v \hat{I} - \nabla_h I\|_2^2$$
$$+ \|\nabla_s \hat{I} - \nabla_s I\|_2^2 \tag{9}$$

where ∇_h, ∇_v, and ∇_s denote the gradient operator along the horizontal, vertical, and spectral direction respectively. Then, the total loss function is as follows:

$$\mathcal{L} = \mathcal{L}_{\mathrm{rec}} + \lambda \mathcal{L}_{\mathrm{grad}} \tag{10}$$

where λ is the weight parameter of $\mathcal{L}_{\mathrm{grad}}$. We empirically set λ to 0.01 to balance the loss terms.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Experiment Setup

1) Benchmark Datasets: To verify the effectiveness of the proposed MAFNet for HSI denoising. The proposed MAFNet is employed on several datasets, and training is conducted using data from ICVL [61] and CAVE [62] hyperspectral dataset. The images in the ICVL dataset were collected over 31 spectral bands with the size of 1392×1300, while the images in CAVE dataset were collected over 31 spectral bands with the size of 512×512. The training data are randomly cropped as cube data with the size of $128 \times 128 \times 31$. Basic data augmentation (rotation and scaling) is used for regularization. A total of 20 000 training samples are generated in total. To verify the robustness of the proposed MAFNet in real data, spaceborne hyperspectral data are used in our experiments, including Pavia University, Urban, and Indian Pines. Through experiments on both real noise HSI datasets, we try to verify the generalization ability and denoising effect of the proposed MAFNet.

2) Noise Setting: Hyperspectral data captured by real spaceborne sensors are commonly contaminated by a mixture noise, such as the Gaussian noise, impulse noise, and deadline noise. In the testing phase, five types of complex noise are defined as follows.

1) *Case 1: Non-i.i.d. Gaussian noise.* Data in all spectral bands are contaminated by Gaussian noise with various intensities. The variances of Gaussian noise are randomly selected from 30 to 70.
2) *Case 2: Gaussian + Stripe noise.* Every band is contaminated by *non-i.i.d Gaussian noise*, as mentioned in Case 1. Besides, some spectral bands are randomly selected to add strip noise. In each band, 5% to 15% of columns are polluted with strips.

3) *Case 3: Gaussian + Deadline noise.* Every band is corrupted by *non-i.i.d Gaussian noise*, as mentioned in Case 1. Besides, deadline noise is randomly added to one-third of spectral bands. In each band, 5% to 15% of columns are conflicted with deadlines.

4) *Case 4: Gaussian + Impulse noise.* Each band is contaminated by Gaussian noise, as mentioned in Case 1. One-third of bands are randomly selected to add impulse noise with intensity ranging from 10% to 70%.

5) *Case 5: Mixture noise.* Like other cases, every spectral band is corrupted by Gaussian noise. Then, each band is randomly contaminated by a random combination of the other three noises.

3) Competing Methods and Quantitative Metrics: The proposed method was compared with six state-of-the-art methods. Both traditional methods and deep learning-based methods are taken into account. Specifically, For traditional methods, BM4D [48], low-rank methods (LRMR [63] and LRTV [64]) are considered. For deep learning-based methods, the proposed MAFNet is compared with HSID-CNN [36], MemNet [65] and QRNN3D [57]. To give a fair evaluation, three quantitative metrics are used, including peak signal-to-noise ratio (PSNR), structure similarity (SSIM) [66], and spectral angle mapper (SAM) [67]. SAM is a spectral-based index, and a smaller value of SAM indicates a better denoising performance. PSNR and SSIM are spatial-based indexes. Larger values of PSNR and SSIM suggest better denoising performance.

4) Incremental Learning Policy: We use an incremental learning policy for stable training, which can effectively avoid the network converging to suboptimal minimum. Specifically, the training goes through three stages, and the training data of each phase uses the same network. In the first stage, Gaussian noise with fixed noise level ($\sigma = 30, 50, 70$) is sequentially employed to build the training data for network training in turn. The network weights of each training phase are saved, and we load the last trained network weights to initialize the network parameters for the next training phase, instead of retraining the network again from scratch. Next, we use blind Gaussian noise (randomly selected from $\sigma = 30, 50, 70$) to construct the training data, and the method described in the first stage is still used to load the network weight data already trained in the first stage. Finally, the complex noise is employed to produce the training data (from Cases 1 to 5). With the increase of the noise complexity of each stage, the denoising difficulty of the model also increases. Therefore, making full use of the pretrained network model in the previous stage is more conducive to improve denoising performance and enables the network to converge better. In order to explore the effectiveness of the incremental learning strategy, we plotted the loss function curve in the model training process, as shown in Fig. 6, we can see that the incremental learning strategy can converge faster and achieve a better denoising performance.

It should be noted that we make the model handle data with different noises sequentially. Following the easy-to-difficult learning strategy in Curriculum learning [68], we incrementally learning the noise from Cases 1 to 4, and therefore, gradually improving the generalization and learning ability of the model. Finally, the model learns the complex noise, including all kinds
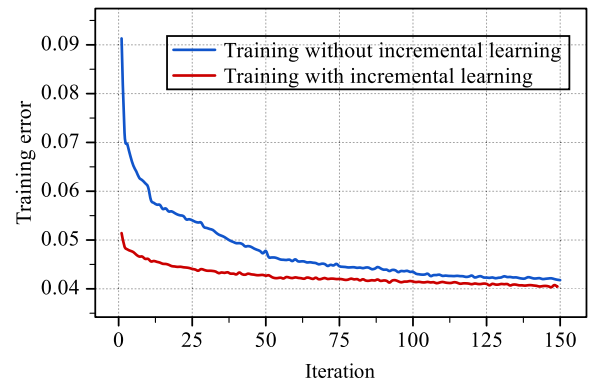

Fig. 6.   Training errors with/without incremental learning. Blue curve denotes the mixture noise data training without incremental learning. Red curve denotes the mixture noise data training with incremental learning.

of noise from Cases 1 to 4. Hence, the final model is robust for complex noise removal. Through incremental learning, the proposed MAFNet achieves a better denoising performance.

We initialize the learning rate at 10^{-4}, and it decayed every epoch to accelerate training. The training process of MAFNet took 100 epochs for Gaussian noise and 150 epochs for complex noise. The network is optimized using the Adam optimizer with the PyTorch framework on a machine with NVIDIA GTX 2080Ti GPU, Intel(R) Xeon(R) E5 CPU of 2.50 GHz and 32 GB RAM.

B. Experiments on Gaussian Noise Cases

This article uses a single model to process various Gaussian noise levels. Specifically, additive Gaussian noise with different variances is imposed to produce a set of noisy HSI patches. The average evaluation indexes are listed in Table I. The best results have been marked in bold. The best performance for each quality index is marked in bold. Figs. 7 and 8 show the denoising results under noise levels $\sigma = 30$ and $\sigma = 70$ to give detailed comparison results.

Through comparison, we can observe that the proposed MAFNet obtains better performance metrics (PSNR, SSIM, and SAM) when dealing with Gaussian noise cases. It is owing to this reason that the proposed MAFNet takes the multiscale contextual information into account. Furthermore, benefiting from the AIN module and coattention fusion, the MAFNet also achieves better denoising results compared with HSID-CNN, MemNet, and QRNN3D. As shown in Figs. 7 and 8, we select one band to give the denoising results. It can be easily seen that the denoising result of the proposed MAFNet is capable of effectively reducing the Gaussian noise while precisely preserving the basic texture details of the original HSI. LRMR can hardly deal with degraded bands with strong Gaussian noise. BM4D introduces undesirable artifacts in some regions. Other deep learning-based methods perform better in noise suppression, but still lose some texture details. Furthermore, the PSNR and SSIM values of other deep learning-based methods are lower than the proposed MAFNet.

C. Experiments of Complex Noise Removal on ICVL Dataset

As mentioned before, the model trained at the final stage of training is used to deal with the five complex noise cases

TABLE I
QUANTITATIVE RESULTS OF DIFFERENT METHODS ON ICVL DATASET

Case	Index	Noisy	LRMR [63]	BM4D [48]	LRTV [64]	MemNet [65]	HSID-CNN [36]	QRNN3D [57]	Ours
$\sigma = 30$	PSNR ↑	18.58	31.73	36.72	34.79	40.17	41.14	43.46	**43.97**
	SSIM ↑	0.107	0.679	0.865	0.769	0.962	0.957	0.973	**0.975**
	SAM ↓	0.703	0.179	0.143	0.157	0.074	0.096	0.038	**0.026**
$\sigma = 50$	PSNR ↑	14.15	29.56	34.92	32.75	37.68	37.21	41.93	**42.03**
	SSIM ↑	0.043	0.634	0.769	0.646	0.941	0.932	0.957	**0.963**
	SAM ↓	0.890	0.214	0.192	0.194	0.093	0.124	0.040	**0.029**
$\sigma = 70$	PSNR ↑	11.22	26.44	31.78	29.13	35.83	35.46	39.73	**40.84**
	SSIM ↑	0.023	0.576	0.601	0.598	0.897	0.896	0.945	**0.954**
	SAM ↓	1.013	0.282	0.231	0.252	0.097	0.148	0.051	**0.031**
Blind	PSNR ↑	14.66	30.16	33.90	32.26	38.84	38.32	42.03	**42.15**
	SSIM ↑	0.058	0.652	0.835	0.774	0.953	0.943	0.960	**0.963**
	SAM ↓	0.868	0.196	0.121	0.151	0.082	0.131	0.044	**0.029**

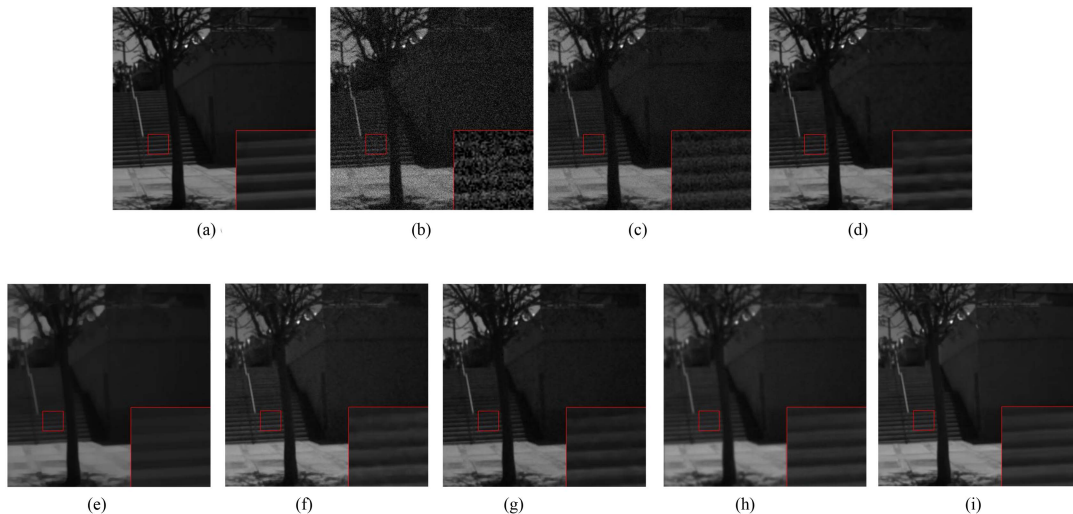"Blind" denotes that each image is corrupted by Gaussian noise with unknown σ.

Fig. 7. Gaussian noise removing results of PSNR (dB) at tenth band of the HSI under noise level $\sigma = 30$ (ICVL dataset). (a) Ground truth ($+\infty$). (b) Noisy (18.58). (c) LRMR (32.41). (d) BM4D (36.65). (e) LRTV (34.77). (f) MemNet (39.63). (g) HSID-CNN (40.06). (h) QRNN3D (42.24). (i) Ours (42.57).

simultaneously. Five types of complex noise include non-i.i.d Gaussian noise, Gaussian + stripe noise, Gaussian + deadline noise, Gaussian + impulse noise, and Mixture noise. We conducted experiments on ICVL dataset for complex noise removal. The quantitative results are shown in Fig. 9, and the corresponding quantitative values are listed in Table II, respectively. The best results have been marked in bold.

Fig. 9 shows the visual results of MAFNet denoising under complex noise conditions. The result shows that our MAFNet significantly outperforms the other methods. By comparing the denoising performance indicators in Table II, it can be observed that our MAFNet performs better than LRMR and LRTV, since they are low-rank matrix-based methods and some basic structures get lost in the denoising process. Furthermore, our MAFNet performs better than the other deep learning-based methods (MemNet, HSID-CNN and QRNN3D). It is evident that the multiscale feature exploitation and contextual information integration can help the network to capture more

intrinsic characteristics of HSI. At the same time, it also helps the network to retain more structural information about the input image. As shown in Fig. 9, our MAFNet not only removes complex noise, but also retains the structure and spatial details. Moreover, compared with other methods, the proposed MAFnet generates HSI images with a more natural and vivid appearance. Besides, the HSI images produced by the proposed MAFNet have better global contrast.

D. Experiments of Complex Noise Removal on CAVE Dataset

We conducted complex noise removal experiments on the CAVE dataset [62] so as to verify the effectiveness of the MAFNet's denoising performance. Each image in the dataset is acquired at a wavelength of 10 nm in the range of 400–700 nm. We divided the dataset into two sets, with 20 images for training and 12 images for testing. Table III presents the quantitative evaluation results of different methods. The
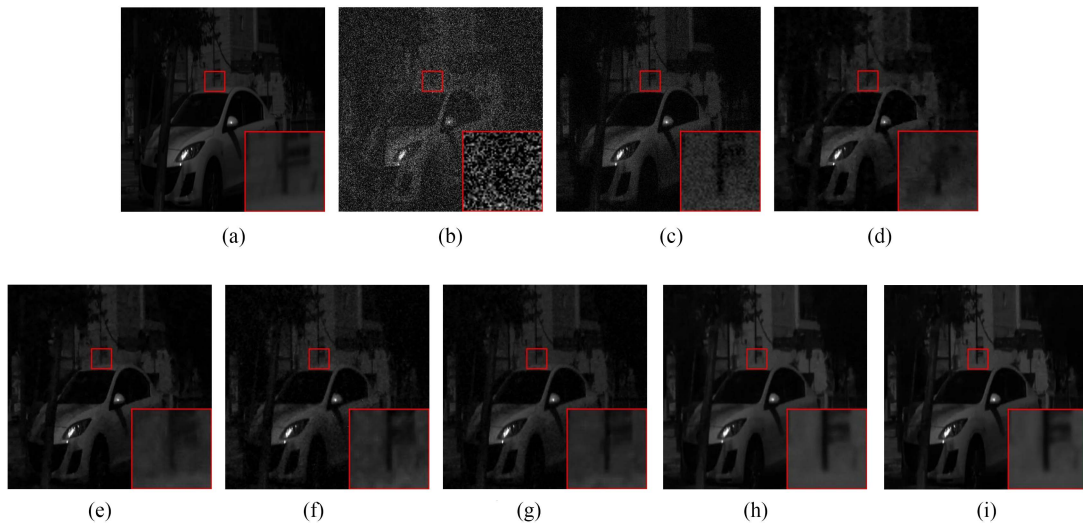
Fig. 8. Gaussian noise removal results of PSNR (dB) at tenth band of image under noise level $\sigma = 70$ (ICVL dataset). (a) Ground truth ($+\infty$). (e) LRTV (29.36). (b) Noisy (14.15). (c) LRMR (26.26). (d) BM4D (30.27). (g) MemNet (34.71). (g) HSID-CNN (35.55). (h) QRNN3D (39.03). (i) Ours (39.78).
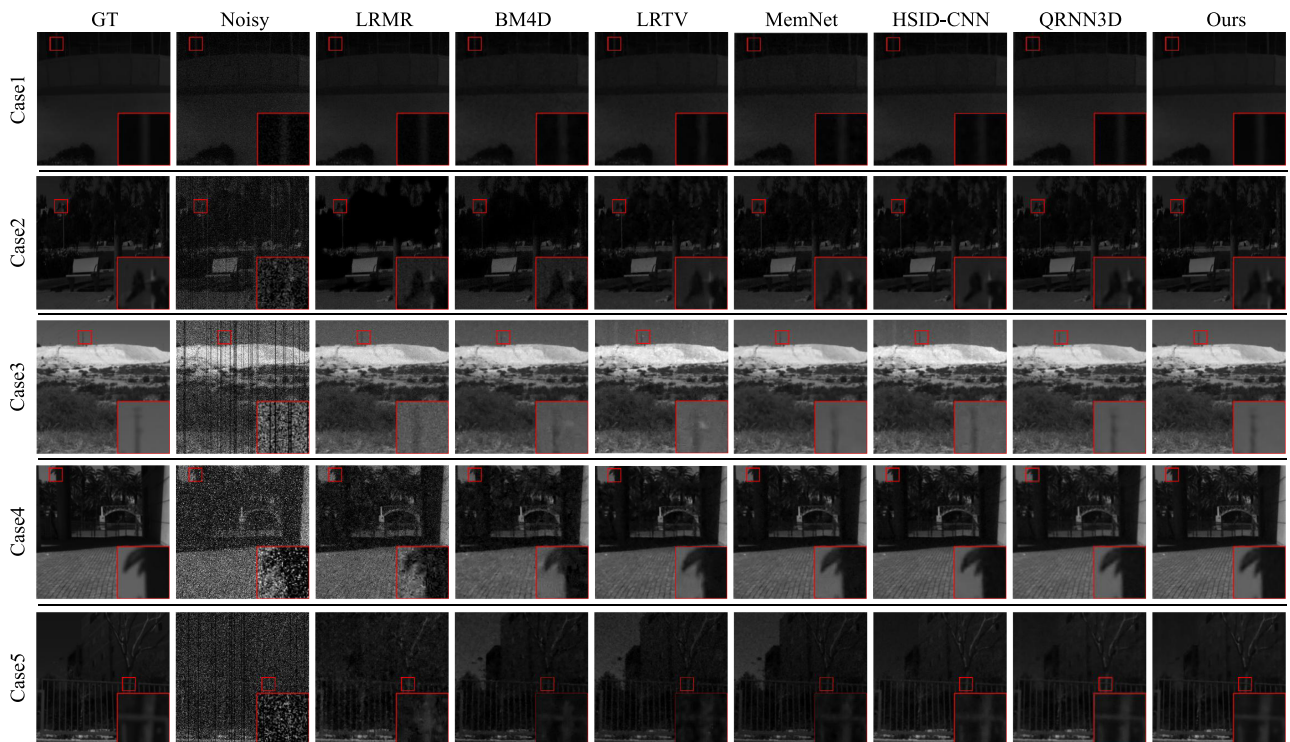

Fig. 9. Complex noise removal results on the ICVL dataset. Examples for non-i.i.d Gaussian noise (Case 1), Gaussian + stripes (Case 2), Gaussian + deadline (Case 3), Gaussian + impulse (Case 4), and mixture noise (Case 5) removal are illustrated, respectively.

best results have been marked in bold. We compare the proposed MAFNet with BM4D [48], OLRT [69], NGMeet [70], MemNet [65], HSID-CNN [36], and QRNN3D [57]. It can be observed that the proposed MAFNet achieves the highest PSNR value, which demonstrates its superior denoising performance.

It should be noted that we employ state-of-the-art low-rank tensor recovery models (NGMeet [70] and OLRT [69]) on the CAVE dataset. These low-rank tensor methods can effectively utilize both the spatial-spectral information, and preserve the high-dimensional spatial and spectral structure information in HSIs. NGMeet learns the orthogonal basis matrix and reduced image, which produces impressive recovered images. OLRT provides a flexible and promising solution for HSI denoising, deblurring, and inpainting. Additionally, we conducted experiments of Gaussian noise on CAVE dataset. In the low noise case, NGMeet and OLRT demonstrate a better performance than MAFNet. As the noise level increases, the proposed MAFNet

TABLE II
QUANTITATIVE RESULTS OF DIFFERENT METHODS UNDER FIVE COMPLEX NOISE CASES ON ICVL DATASET

Case	Index	Noisy	LRMR [63]	BM4D [48]	LRTV [64]	MemNet [65]	HSID-CNN [36]	QRNN3D [57]	Ours
Case 1	PSNR ↑	17.99	32.20	35.46	36.12	41.32	41.14	42.51	**43.54**
	SSIM ↑	0.160	0.692	0.892	0.905	0.968	0.952	0.969	**0.974**
	SAM ↓	0.853	0.201	0.137	0.115	0.050	0.057	0.043	**0.028**
Case 2	PSNR ↑	17.62	30.87	34.17	34.89	40.32	39.68	41.78	**42.50**
	SSIM ↑	0.151	0.653	0.821	0.890	0.954	0.939	0.960	**0.968**
	SAM ↓	0.855	0.257	0.159	0.129	0.078	0.071	0.049	**0.030**
Case 3	PSNR ↑	17.53	30.17	33.65	34.02	39.26	39.75	40.88	**41.77**
	SSIM ↑	0.150	0.660	0.860	0.847	0.947	0.932	0.954	**0.966**
	SAM ↓	0.871	0.249	0.174	0.160	0.081	0.091	0.059	**0.032**
Case 4	PSNR ↑	15.05	29.37	32.75	33.14	37.22	36.83	37.71	**37.77**
	SSIM ↑	0.117	0.644	0.801	0.795	0.911	0.898	0.938	**0.946**
	SAM ↓	0.900	0.271	0.196	0.182	0.114	0.138	0.095	**0.067**
Case 5	PSNR ↑	14.18	27.16	30.81	31.10	36.14	36.69	37.04	**37.19**
	SSIM ↑	0.095	0.591	0.719	0.735	0.905	0.884	0.922	**0.938**
	SAM ↓	0.918	0.312	0.224	0.211	0.135	0.167	0.105	**0.065**

TABLE III
QUANTITATIVE RESULTS OF DIFFERENT METHODS ON THE CAVE DATASET

Index	Noisy	BM4D [48]	OLRT [69]	NGMeet [70]	MemNet [65]	HSID-CNN [36]	QRNN3D [57]	Ours
PSNR ↑	12.92	29.07	31.09	32.74	32.04	32.82	33.94	**34.19**
SSIM ↑	0.092	0.704	0.785	0.792	0.854	0.890	0.914	**0.931**
SAM ↓	1.203	0.384	0.164	0.152	0.119	0.101	0.095	**0.081**

TABLE IV
QUANTITATIVE RESULTS OF DIFFERENT METHODS ON THE PAVIA UNIVERSITY DATASET

Index	Noisy	LRMR [63]	BM4D [48]	LRTV [64]	MemNet [65]	HSID-CNN [36]	QRNN3D [57]	Ours
PSNR ↑	17.99	27.64	25.16	26.75	29.27	30.38	33.83	**34.03**
SSIM ↑	0.160	0.717	0.613	0.675	0.794	0.809	0.894	**0.912**
SAM ↓	0.853	0.308	0.397	0.327	0.137	0.148	0.109	**0.101**

outperforms NGMeet and OLRT. Therefore, we can conclude that NGMeet and OLRT are effective in spectral regularization, and the proposed MAFNet is beneficial for complex noise removal.

E. Experiments on Remote-Sensing HSI Datasets

To verify the robustness and denoising performance of the proposed MAFNet, we conduct extensive experiments on three remote-sensing HSI datasets. The first dataset is the *Pavia University dataset*. It contains 103 bands, and the spatial size of the image is 610 × 340 pixels. It was captured by the reflective optics system imaging spectrometer sensor over the Pavia University, Italy. The second dataset is the *Urban dataset*. It was captured by the HYDICE sensor. The sensor provides 210 bands ranging from 400 nm to 2500 nm. In order to further verify the denoising performance of MAFNet on real hyperspectral noise, we introduce the Indian Pines dataset, which is captured

by the AVIRIS sensor and contains 220 bands with a resolution size of 145 × 145 pixels.

1) Results on the Pavia University Dataset With Mixture Noise: We added mixture noise on the Pavia University dataset, and the experimental results are listed in Table IV. The best results have been marked in bold. It can be easily seen that the proposed model achieves the highest quantitative metrics. The corresponding visual results are provided in Fig. 10, and the values of the PSNR and SSIM within different bands of the restored HSI on the dataset are depicted in Fig. 11. Our method not only effectively removes the complex noise, but also simultaneously preserves the high-frequency texture details. Furthermore, the mean normalized digital number curves by different methods are shown in Fig. 12, which demonstrates that the proposed MAFNet effectively removes the complex noise without introducing obvious spectral distortion.

As demonstrated in Fig. 11, the proposed method achieves the best PSNR values on almost all spectral bands. We find
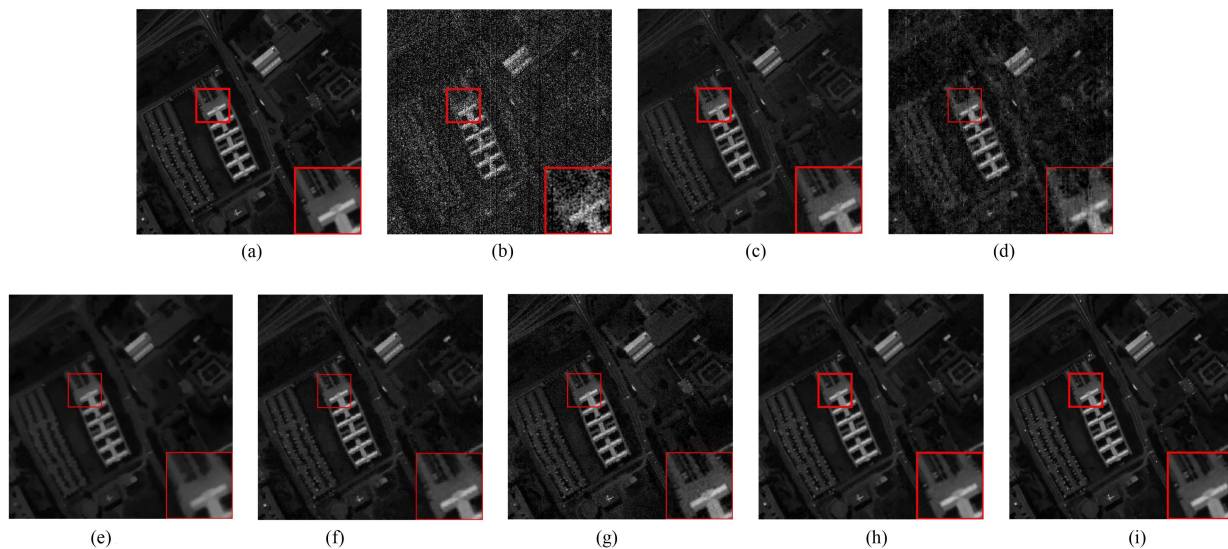
Fig. 10.   Simulated noise removal results of PSNR (dB) at tenth band of mixture noise on the Pavia University dataset. (a) Ground truth (+∞). (b) Noisy (14.15). (e) LRTV (26.75). (f) MemNet (29.27). (c)LRMR (27.64). (g) HSID-CNN (30.38). (d) BM4D (25.16). (h) QRNN3D (33.83). (i) Ours (34.03).
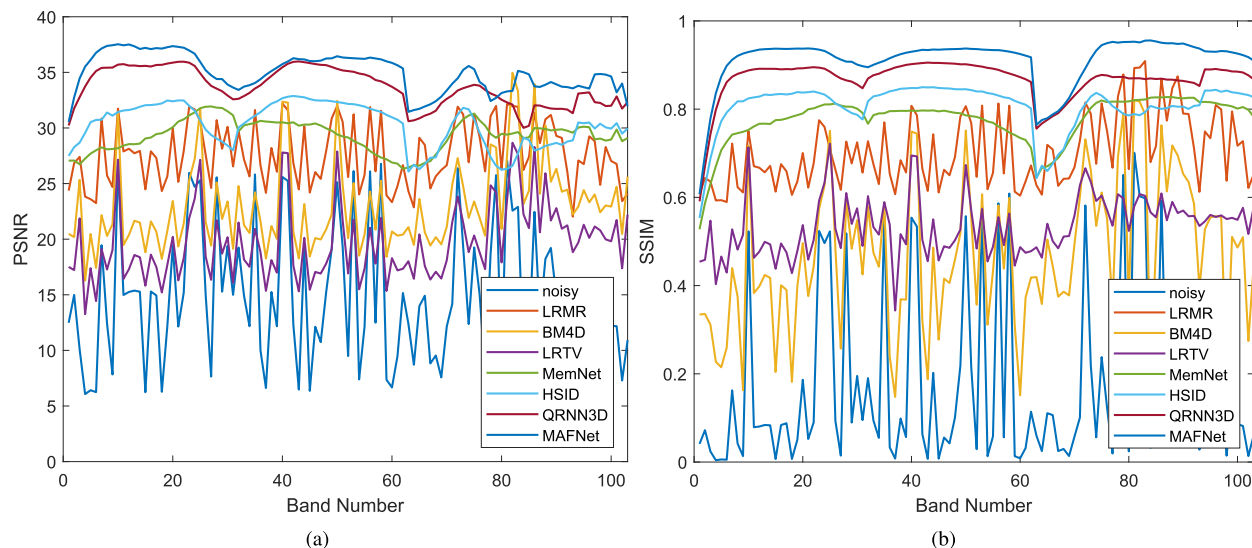

Fig. 11.   PSNR and SSIM values of different denoising methods in each band of the simulated experiments in mixture noise on Pavia University.
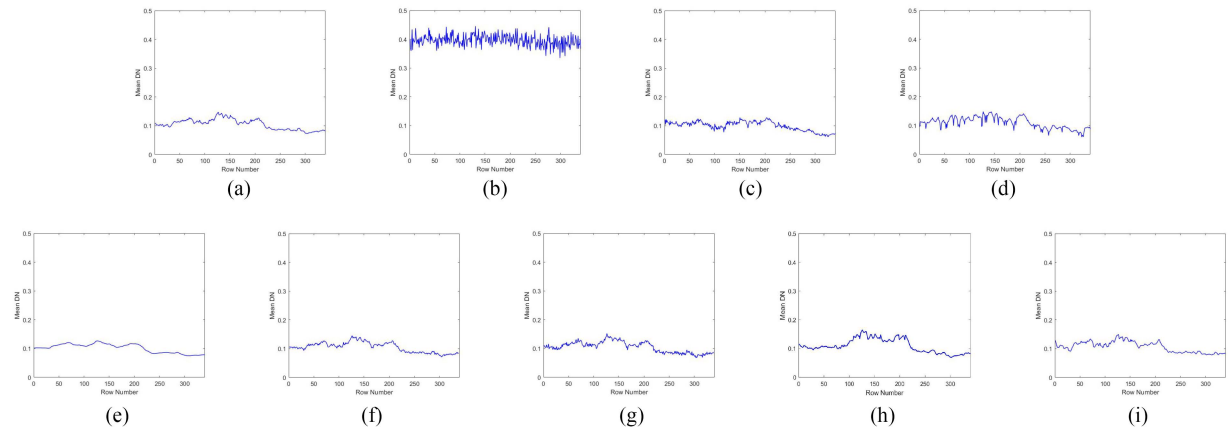

Fig. 12.   Horizontal mean digital number curves at 14th band of real noise by different methods on the Pavia University dataset. (a) GT. (b) Noisy. (c) LRMR. (d) BM4D. (e) LRTV. (f) MemNet. (g) HSID-CNN. (h) QRNN3D. (i) Ours.
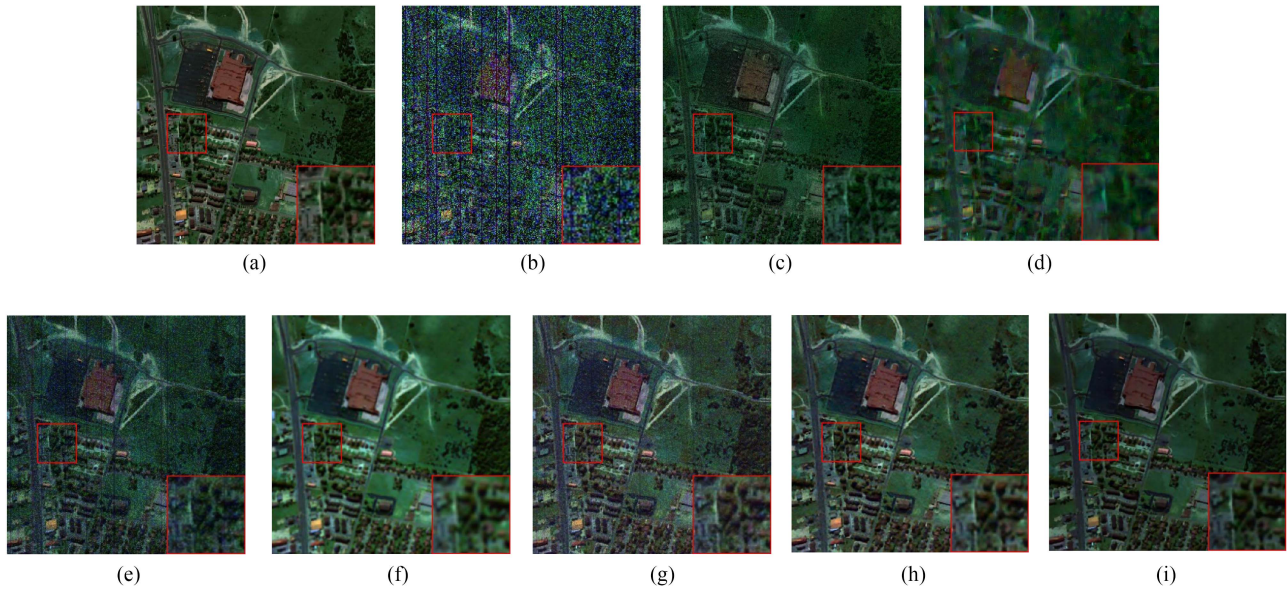
Fig. 13. Noise removal visual results of false color images on band 50, 100, and 150 of the Urban dataset with competing methods. (a) Ground truth. (b) Noisy. (c) LRMR. (d) BM4D. (e) LRTV. (f) MemNet. (g) HSID-CNN. (h) QRNN3D. (h) Ours.
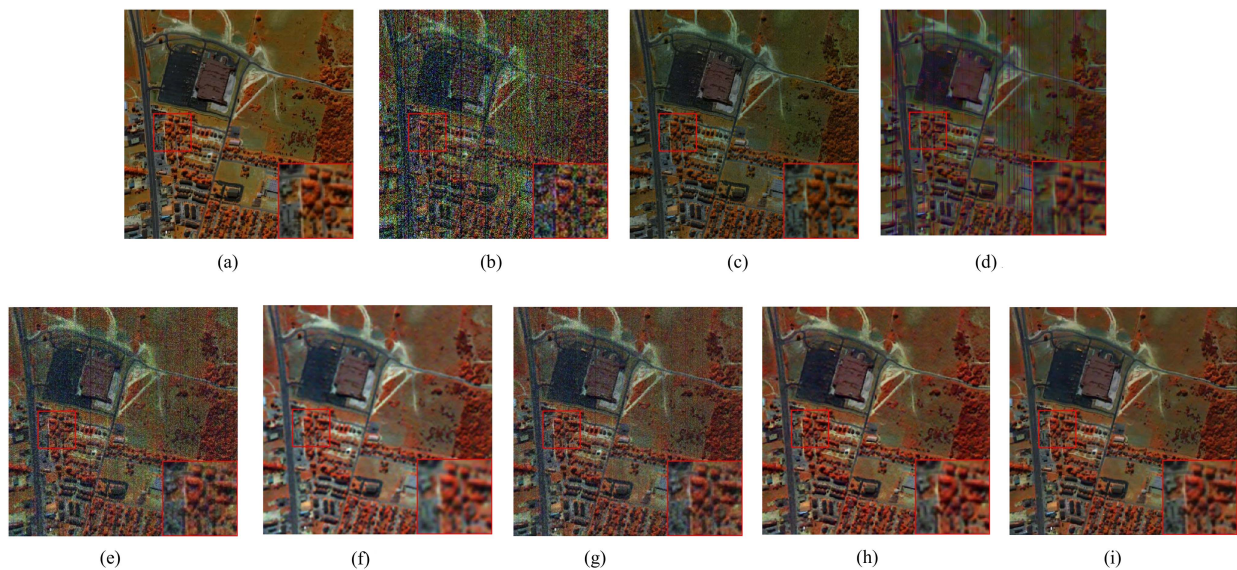

Fig. 14. Noise removal visual results of false color images on band 70, 110, and 160 of the Urban dataset with competing methods. (a) Ground truth. (b) Noisy. (c) LRMR. (d) BM4D. (e) LRTV. (f) MemNet. (g) HSID-CNN. (h) QRNN3D. (i) Ours.

that nearly all the methods perform differently across spectral bands. It is mainly caused by the denoising difficulties among different spectral bands, since the intrinsic noise levels from different spectral bands are different. In addition, due to the characteristics of the hyperspectral sensor, the spatial details of the different spectral bands are different. Hence, nearly all the methods perform differently across spectral bands.

2) Results on the Urban Dataset With Mixture Noise: To further validate the efficacy of MAFNet in denoising, we added severely polluted noise to the Urban dataset. Figs. 13 and 14 demonstrate that deep learning-based methods are effective in removing severely polluted noise. Notably, the proposed MAFNet achieves the best denoising performance among all the methods.

3) Results on the Indian Pines Dataset With Real-World Noise: Some bands in the Indian Pines dataset are seriously corrupted by the atmosphere and water, and are polluted by complex noises. We show the denoising result of various methods on this dataset in Fig. 15. It can be observed that MAFNet can still achieve the satisfactory denoising performance in real hyperspectral noise. Furthermore, it performs better in preserving spatial details.

F. Ablation Study and Parameter Sensitivity Analysis

1) Coattention and AIN: The critical parts of the proposed method are the coattention and AIN modules. In order to validate
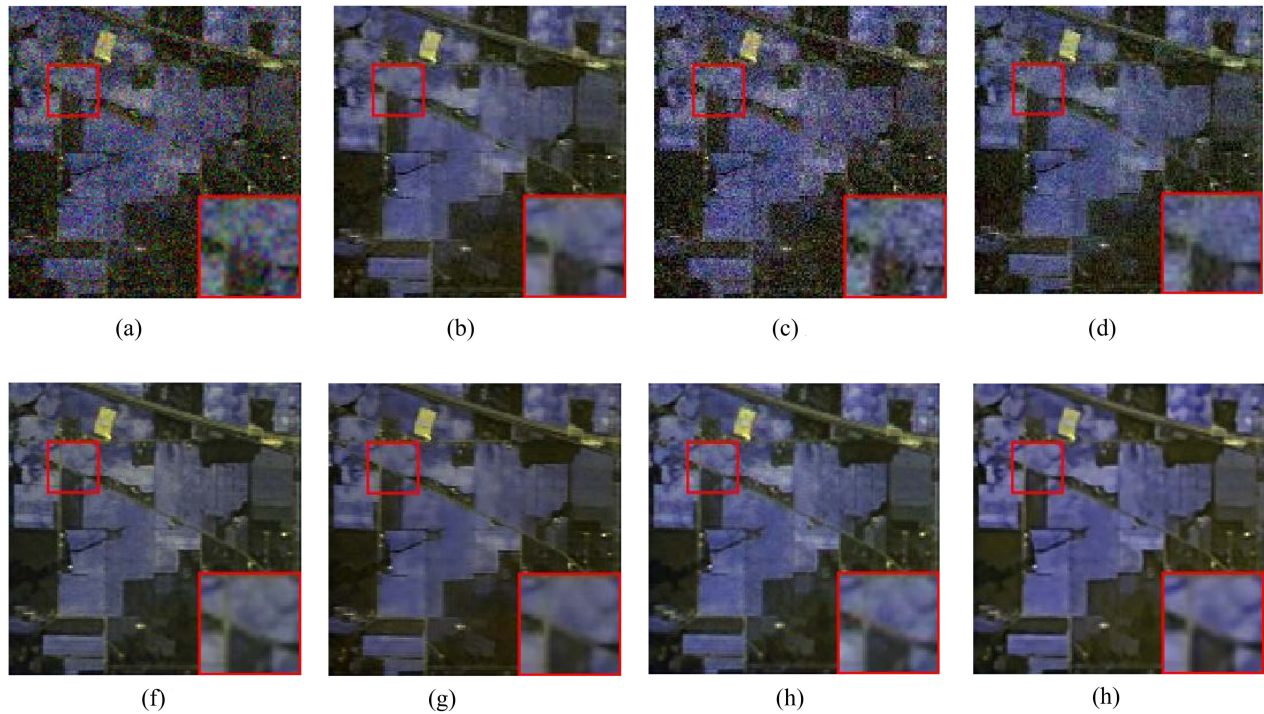
Fig. 15. Noise removal visual results of false color images on band 2, 3, and 203 of the India Pines dataset with competing methods. (a) Noisy. (b) LRMR. (c) BM4D. (d) LRTV. (f) MemNet. (g) HSID-CNN. (h) QRNN3D. (h) Ours.

TABLE V
ABLATION STUDY OF COATTENTION AND AIN

Method	Coattention	AIN	PSNR ↑
(a)	✓		33.93
(b)		✓	33.85
(c)	✓	✓	34.03

Fig. 16. Visualization of γ and β in AIN module.

the effectiveness of both modules, we comprehensively compare three variants on the Pavia University dataset.

1) Basic network with coattention module.
2) Basic network with AIN module.
3) Basic network with both coattention and AIN module (our proposed MAFNet).

The experimental results on the Pavia University dataset are given in Table V. It is evident that the combination of coattention and AIN achieves the best denoising performance.

The AIN module provides semantic guidance via channelwise normalization and pixel-wise affine transformation. The transform parameters γ and β are visualized in Fig. 16. In the AIN module, γ calibrates the input feature and highlights the important regions. As can be observed that in the low-resolution branch, γ highlights the object boundary and, thus, transfers the basic structure to the high-resolution branch. β is used as the complementary information for feature calibration in the AIN module, providing more details to complete the denoising task.

2) Feature Fusion in Coattention Module: Here, we discuss the multiscale feature fusion in coattention module. In the proposed MAFNet, multiscale features are fused by element-wise summation. We also designed two other schemes, and the experimental results on the Pavia University dataset are shown in Table VI. The "Concat" uses concatenation for multiscale feature fusion, and employs 1×1 convolution layer to reduce the channels. The "Multiply" uses element-wise multiplication for feature fusion. Compared with "Concat" and "Multiply," the proposed method could achieve slightly a better denoising performance.
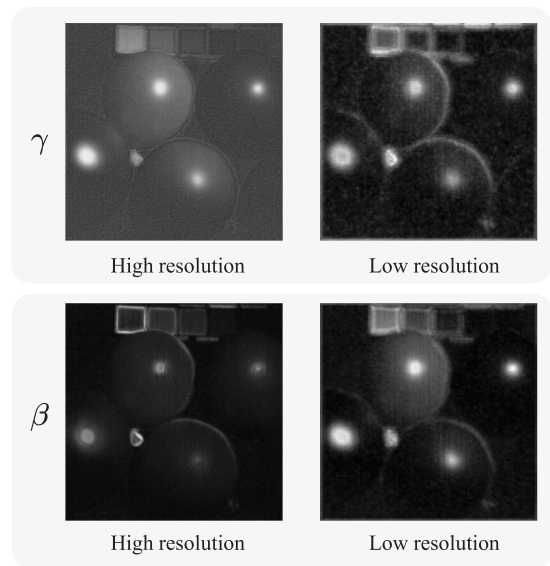
TABLE VI
DIFFERENT FEATURE FUSION SCHEMES FOR COATTENTION MODULE

Setting	Concat	Multiply	Ours
PSNR ↑	33.96	33.93	34.03

TABLE VII
DIFFERENT ATTENTION SCHEMES FOR COATTENTION MODULE

Setting	Split	C-Attn	Ours
PSNR ↑	33.90	39.87	34.03

TABLE VIII
INFLUENCE OF λ IN MAFNET

λ	0.000	0.005	0.010	0.015	0.020
PSNR ↑	33.49	33.74	34.03	33.80	33.78

TABLE IX
DENOISING PERFORMANCE AND THE NUMBER OF PARAMETERS COMPARISON

Method	PSNR ↑	FLOPs (G) ↓	Params (M) ↓
MemNet [66]	29.27	31.61	1.92
HSID-CNN [36]	30.38	6.61	0.63
QRNN3D [58]	33.83	47.62	3.37
MAFNet-S	32.87	5.44	2.03
MAFNet-B	34.03	20.15	7.55
MAFNet-L	34.21	77.71	29.21

is within an acceptable range. Therefore, in our implementations, we set the channel numbers of three scales as (64, 128, and 256). Compared with other deep learning-based methods, our method exhibits an impressive performance in FLOPs and the number of parameters. Specifically, although the FLOPs of our method are higher compared than HSID-CNN, the denoising performance of our method is much better. Compared with MemNet and QRNN3D, the proposed method is more computationally efficient.

3) Split Attention and Self-Calibration: Two attention mechanisms are employed in the coattention module: Split attention and self-calibration. To verify the effectiveness of both mechanisms, we design three variants, as shown in Table VII. The "Split" only uses the concatenation and split part in coattention module, and the self-calibration is removed. The "C-Attn" uses the channel attention [71] instead of the concatenation and split part in coattention module, and the self-calibration is removed. In the proposed MAFNet, both split attention and self-calibration are used. The experimental results on the Pavia University dataset are shown in Table VII. It can be observed that "Split" slightly outperforms "C-Attn," since multiscale features are adaptively emphasized by split attention. The proposed MAFNet performs the best by combining split attention and self-calibration.

4) Global Gradient Regularizer: In our designed loss function, λ is a critical parameter that affects the global gradient regularizer, and can affect the denoising performance. We evaluate the denoising performance by taking different λ on the ICVL dataset while keeping the network unchanged. The results are shown in Table VIII, and it can be observed that the proposed MFANet reaches the bet PSNR value when $\lambda = 0.010$. Therefore, in our implementation, λ is set to 0.010.

5) Number of Channels: We explore the influence of the channel number on the denoising performance, and design three variants of MAFNet according to the channel number and model size, i.e., MAFNet-S, MAFNet-B, and MAFNet-L. The channel numbers of three scales of MAFNet-S are set as (32, 64, and 128). Then, the corresponding channel numbers in MAFNet-B and MAFNet-L are set as (64, 128, and 256) and (128, 256, and 512), respectively. The denoising performance of three variants and other deep learning methods are illustrated in Table IX. It can be observed that MAFNet-L achieves the best PSNR value, but its computational complexity is large. MAFNet-S achieves a good denoising performance while it is rather computationally efficient. It should be noted that MAFNet-B achieves an excellent denoising performance while its computational complexity

V. CONCLUSION

In this article, we present an MAFNet for HSI noise reduction. Two key components contribute to improving the HSI denoising: A progressively multiscale information aggregation framework and coattention fusion module. Specifically, a set of multiscale images are generated and fed into a coarse-fusion network to exploit the contextual texture correlation. Thereafter, a fine fusion network is followed to exchange the information across the parallel multiscale subnetworks. Ultimately, the coattention fusion module adaptively emphasizes informative features from different scales and reinforces the discriminative learning capability for denoising. Experiments on both synthetic and real HSI datasets verified the superiority of the proposed method compared with other state-of-the-art HSI denoising methods.

Although the MAFNet in this article exhibits an outstanding denoising performance, the utilization of multiscale branches for feature extraction also results in an increase in the number of parameters and computational complexity. Consequently, in the future, we will concentrate on devising lightweight HSI denoising techniques.

REFERENCES

[1] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.

[2] Y. Zhang, W. Li, M. Zhang, Y. Qu, R. Tao, and H. Qi, "Topological structure and semantic information transfer network for cross-scene hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, 2021, to be published, doi: 10.1109/TNNLS.2021.3109872.

[3] Y. Zhang, W. Li, M. Zhang, S. Wang, R. Tao, and Q. Du, "Graph information aggregation cross-domain few-shot learning for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, 2022, to be published, doi: 10.1109/TNNLS.2022.3185795.

[4] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[5] D. Hong, J. Yao, D. Meng, Z. Xu, and J. Chanussot, "Multimodal GANs: Toward crossmodal hyperspectral–multispectral image segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5103–5113, Jun. 2021.

[6] X. Tong et al., "A novel approach for hyperspectral change detection based on uncertain area analysis and improved transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 2056–2069, Apr. 2020.

[7] D. Marinelli, F. Bovolo, and L. Bruzzone, "A novel change detection method for multitemporal hyperspectral images based on binary hyperspectral change vectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4913–4928, Jul. 2019.

[8] Q. Wang, Z. Yuan, Q. Du, and X. Li, "GETNET: A general end-to-end 2-D CNN framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 3–13, Jan. 2019.

[9] T. Zhan et al., "TDSSC: A three-directions spectral-spatial convolution neural network for hyperspectral image change detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 377–388, Nov. 2021.

[10] X. Ou, L. Liu, S. Tan, G. Zhang, W. Li, and B. Tu, "A hyperspectral image change detection framework with self-supervised contrastive learning pretrained model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 7724–7740, Sep. 2022.

[11] L. Wang, L. Wang, H. Wang, X. Wang, and L. Bruzzone, "SPCNet: A subpixel convolution-based change detection network for hyperspectral images with different spatial resolutions," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5533314.

[12] S. Sun and C. Hu, "The challenges of interpreting oil–water spatial and spectral contrasts for the estimation of oil thickness: Examples from satellite and airborne measurements of the deepwater horizon oil spill," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 5, pp. 2643–2658, May 2019.

[13] X. Kang, Z. Wang, P. Duan, and X. Wei, "The potential of hyperspectral image classification for oil spill mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5538415.

[14] Y. Li, Q. Yu, M. Xie, Z. Zhang, Z. Ma, and K. Cao, "Identifying oil spill types based on remotely sensed reflectance spectra and multiple machine learning algorithms," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 9071–9078, Sep. 2021.

[15] Z. Liu, Q. Qiu, J. Li, L. Wang, and A. Plaza, "Geographic optimal transport for heterogeneous data: Fusing remote sensing and social media," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6935–6945, Aug. 2021.

[16] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.

[17] J. Jia, X. Zheng, S. Guo, Y. Wang, and J. Chen, "Removing stripe noise based on improved statistics for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Oct. 2022, Art. no. 5501405.

[18] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.

[19] B. Du, Z. Huang, N. Wang, Y. Zhang, and X. Jia, "Joint weighted nuclear norm and total variation regularization for hyperspectral image denoising," *Int. J. Remote Sens.*, vol. 39, no. 2, pp. 334–355, 2018.

[20] 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.

[21] T. Hu, W. Li, N. Liu, R. Tao, F. Zhang, and P. Scheunders, "Hyperspectral image restoration using adaptive anisotropy total variation and nuclear norms," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1516–1533, Feb. 2021.

[22] B. Rasti, P. Ghamisi, and J. A. Benediktsson, "Hyperspectral mixed Gaussian and sparse noise reduction," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 474–478, Mar. 2020.

[23] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, Aug. 2020.

[24] C. Cao, J. Yu, C. Zhou, K. Hu, F. Xiao, and X. Gao, "Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 3, pp. 973–988, Mar. 2019.

[25] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[26] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the PARAFAC model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, Oct. 2012.

[27] M. Wang, J. Yu, J.-H. Xue, and W. Sun, "Denoising of hyperspectral images using group low-rank representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4420–4427, Sep. 2016.

[28] 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.

[29] Z. Chen, X. He, C. Ren, H. Chen, and T. Zhang, "Enhanced separable convolution network for lightweight JPEG compression artifacts reduction," *IEEE Signal Process. Lett.*, vol. 28, pp. 1280–1284, Jun. 2021.

[30] Z. Zou, T. Shi, Z. Shi, and J. Ye, "Adversarial training for solving inverse problems in image processing," *IEEE Trans. Image Process.*, vol. 30, pp. 2513–2525, Feb. 2021.

[31] Y. Zhang, K. Li, K. Li, G. Sun, Y. Kong, and Y. Fu, "Accurate and fast image denoising via attention guided scaling," *IEEE Trans. Image Process.*, vol. 30, pp. 6255–6265, Jul. 2021.

[32] Y. Yu, X. Li, and F. Liu, "E-DBPN: Enhanced deep back-projection networks for remote sensing scene image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5503–5515, Aug. 2020.

[33] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "HPGAN: Hyperspectral pansharpening using 3-D generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 463–477, Jan. 2021.

[34] 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.

[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] 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.

[37] 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.

[38] 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.

[39] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Joint and progressive subspace analysis (JPSA) with spatial–spectral manifold alignment for semisupervised hyperspectral dimensionality reduction," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3602–3615, Jul. 2021.

[40] J. Wang et al., "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021.

[41] Y. Yuan et al., "HRFormer: High-resolution vision transformer for dense prediction," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 7281–7293.

[42] J. Gu et al., "Multi-scale high-resolution vision transformer for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 12094–12103.

[43] W. Wang, T. Zhou, F. Yu, J. Dai, E. Konukoglu, and L. Van Gool, "Exploring cross-image pixel contrast for semantic segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 7303–7313.

[44] S. Chu, D. Kim, and B. Han, "Learning debiased and disentangled representations for semantic segmentation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 8355–8366.

[45] M. Li et al., "Classification of surface natural resources based on HR-Net and DEM," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4988–4991.

[46] 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.

[47] A. Zelinski and V. Goyal, "Denoising hyperspectral imagery and recovering junk bands using wavelets and sparse approximation," in *Proc. IEEE Int. Symp. Geosci. Remote Sens.*, 2006, pp. 387–390.

[48] 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.

[49] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.

[50] 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.

[51] S. Gu, Y. Li, L. Van Gool, and R. Timofte, "Self-guided network for fast image denoising," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 2511–2520.

[52] L. Liu, X. Jia, J. Liu, and Q. Tian, "Joint demosaicing and denoising with self guidance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 2237–2246.

[53] Y. Kim, J. W. Soh, G. Y. Park, and N. I. Cho, "Transfer learning from synthetic to real-noise denoising with adaptive instance normalization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3479–3489.

[54] 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.

[55] 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. 2021, Art. no. 5504714.

[56] B. Lin, X. Tao, and J. Lu, "Hyperspectral image denoising via matrix factorization and deep prior regularization," *IEEE Trans. Image Process.*, vol. 29, pp. 565–578, Jul. 2020.

[57] 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.

[58] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 1510–1519.

[59] J. Yao et al., "Semi-active convolutional neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Sep. 2022, Art. no. 5537915.

[60] J.-J. Liu, Q. Hou, M.-M. Cheng, C. Wang, and J. Feng, "Improving convolutional networks with self-calibrated convolutions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 10093–10102.

[61] 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.

[62] F. Yasuma, T. Mitsunaga, D. Iso, and S. Nayar, "Generalized assorted pixel camera: Post-capture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.

[63] 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.

[64] W. He, 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. 2016.

[65] Y. Tai, J. Yang, X. Liu, and C. Xu, "MEMNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2017, pp. 4549–4557.

[66] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[67] R. Yuhas, J. Boardman, and A. Goetz, "Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques," in *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop*, 1993, pp. 205–208.

[68] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *Proc. Int. Conf. Mach. Learn.*, 2009, pp. 41–48.

[69] Y. Chang, L. Yan, B. Chen, S. Zhong, and Y. Tian, "Hyperspectral image restoration: Where does the low-rank property exist," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6869–6884, Aug. 2021.

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

[71] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 7132–7141.

Haodong Pan received the B.Eng. degree in computer science and technology from the Ocean University of China, Qingdao, China, in 2020. He is currently working toward the M.Sc. degree in computer science with the School of Information Science and Technology, Ocean University of China, Qingdao, China.

His current research interests include computer vision and remote sensing image processing.

Feng Gao (Member, IEEE) received the B.Sc. degree in software engineering from Chongqing University, Chongqing, China, in 2008 and the Ph.D. degree in computer science and technology from Beihang University, Beijing, China, in 2015.

He is currently an Associate Professor with the School of Information Science and Engineering, Ocean University of China, Qingdao, China. His research interests include remote sensing image analysis, pattern recognition, and machine learning.

Junyu Dong (Member, IEEE) received the B.Sc. and M.Sc. degrees in mathematics from the Department of Applied Mathematics, Ocean University of China, Qingdao, China, in 1993 and 1999, respectively, and the Ph.D. degree in image processing from the Department of Computer Science, Heriot-Watt University, Edinburgh, U.K., in 2003.

He is currently a Professor and Dean with the School of Computer Science and Technology, Ocean University of China. His research interests include visual information analysis and understanding, machine learning, and underwater image processing.

Qian Du (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Maryland at Baltimore, Baltimore, MD, USA, in 2000.

She is currently the Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote sensing image analysis and applications, and machine learning.

Dr. Du was the recipient of the 2010 Best Reviewer Award from the IEEE Geoscience and Remote Sensing Society (GRSS). She was a Co-Chair for the Data Fusion Technical Committee of the IEEE GRSS from 2009 to 2013, the Chair for the Remote Sensing and Mapping Technical Committee of International Association for Pattern Recognition from 2010 to 2014, and the General Chair for the Fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing held at Shanghai, China, in 2012. She was an Associate Editor for the *Pattern Recognition*, and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. From 2016 to 2020, she was the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATION AND REMOTE SENSING. She is currently a member of the IEEE Periodicals Review and Advisory Committee and SPIE Publications Committee. She is a Fellow of SPIE-International Society for Optics and Photonics (SPIE).