AACNet: Asymmetric Attention Convolution Network for Hyperspectral Image Dehazing

Meng Xu, *Member, IEEE*, Yanxin Peng, Ying Zhang,
Xiuping Jia, *Fellow, IEEE*, Sen Jia, *Senior Member, IEEE*

***Abstract*—Haze in hyperspectral images (HSIs) can lead to crosstalk between multiple bands, resulting in errors that can be amplified and transmitted during data processing. As a consequence, this may cause a reduction in the accuracy and precision of remote sensing data. The purpose of haze removal is to restore high-quality HSIs from degraded ones. The high spectral resolution and typically dozens to hundreds of spectral bands in HSIs pose significant challenges for haze removal. Thus, many methods designed for natural and multispectral images are not effective in removing haze in HSIs. To address this challenge, we develop a model called asymmetric attention convolution network (AACNet) designed for haze removal in HSIs. Specifically, the basic architecture of AACNet is mainly composed of several residual asymmetric attention groups (RAAGs), where the core components are residual asymmetric attention blocks (RAABs). This design enables the full utilization of deep spatial-spectral features while skipping low-frequency regions and focusing more on the haze-affected areas. To more accurately restore the spectral information in areas polluted by haze, a pooling channel self-attention (PCSA) module has been proposed. This module can effectively reconstruct the spectral response curve that is affected by the haze. Our experiments on both simulated and real datasets demonstrate that the proposed AACNet outperforms several leading haze removal methods in both precision and visual quality. The source code and data of this article will be made publicly available at https://github.com/SZU710/AACNet for reproducible research.**

***Index Terms*—hyperspectral images, haze removal, deep learning, attention mechanism.**

I. Introduction

HYPERSPECTRAL sensors can image objects in continuous narrow wavelengths, producing spectral curves that reflect the object's physical properties. Due to the unique spectral characteristics of ground objects, HSIs can accurately detect, identify and analyze ground objects. HSIs are widely used in Earth observation applications, such as resource exploration [1], [2], precision agriculture [3], [4], and disaster monitoring [5]–[7]. Optical imaging is susceptible to scattering and atmospheric particulate matter, resulting in blurry images, reduced visibility, and low contrast, all of which significantly degrade image fidelity. Haze, in particular, can destroy the spectral curve of surface objects and seriously affect the accurate identification of HSIs, ultimately limiting its practical applications. As a result, image dehazing is crucial for many image-processing applications and holds significant practical value.

Researchers have proposed various methods for restoring clear images from hazy images, but these methods are typically specific to natural images. In general, there are two categories of image dehazing methods: prior-based and learning-based. The most striking difference between these two types is that the image priors in prior-based methods are manually crafted, while learning-based methods automatically learn the priors. Several typical image priors, such as dark channel prior [8], color-line prior [9], and color attenuation prior [10] have been proposed to calculate the transmission map for achieving clear images. In addition, Berman *et al.* [11] discovered that image dehazing can also be accomplished through a nonlocal haze-line prior. Although these prior-based methods can produce satisfactory results in some specific scenarios, the reliance on prior assumptions may lead to inaccurate dehazing results. It is worth noting that the atmospheric scattering model-based RGB image dehazing methods [12], [13] are not suitable for haze removal in HSIs. This is mainly because the atmospheric scattering model assumes that the attenuation coefficient of all spectral channels is constant, which does not hold for HSIs due to the more extensive spectral information they possess.

In recent years, deep learning-based methods have demonstrated significant advantages in computer vision and image processing. With the expansion of image datasets, deep learning has become increasingly popular in the study of removing haze from RGB images. Cai *et al.* [13] and Ren *et al.* [12] utilized convolutional neural networks (CNNs) to estimate the medium transmission map from the blurry input image, which was then used to generate the corresponding haze-free image. However, the inaccurate estimation of the transmission map always resulted in a decrease in the quality of the dehazed image. Therefore, subsequent researchers proposed a series of dehazing methods that do not require the estimation of the medium transmission map. Chen *et al.* [14] employed smooth dilation convolution to avoid grid artifacts, and adopted a gating sub-network to fuse features from different levels, thereby recovering clear haze-free images. With the development of attention mechanisms in deep learning, researchers began to use attention mechanisms to enhance the representation ability of dehazing networks. Qin *et al.* [15] combined channel and spatial attention mechanisms to selectively process different channel and pixel weights, enabling the network to focus more

Meng Xu, Yanxin Peng, Ying Zhang, and Sen Jia are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: m.xu@szu.edu.cn; pengyanxin2021@email.szu.edu.cn; 2210273054@email.szu.edu.cn; senjia@szu.edu.cn).

Xiuping Jia is with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2612, Australia (e-mail: x.jia@adfa.edu.au).

on the contaminated areas. Song *et al.* [16] introduced gating mechanisms into convolutional blocks to serve as pixel attention modules, thereby enhancing the dehazing performance of the model.

In the field of remote sensing, a large number of models for multispectral image dehazing have been proposed. These models can be mainly divided into two categories: dehazing models based on traditional methods and dehazing models based on deep learning. The traditional model-based or prior-based methods rely on manually extracted features from the hazy images and use various mathematical models to estimate and remove the haze. Zhou *et al.* [17] utilized radiative transfer model to remove thin clouds by calculating the linear regression coefficient between visible or infrared bands and cirrus band. Xie *et al.* [18] first determined the hazy area through the image brightness, and then repaired the image based on color consistency constraints and global prior information. Xu *et al.* [19] developed a dehazing method that employs a dehazing operator to gradually and iteratively remove haze until satisfactory results are obtained, and updates the transmission rate and scene brightness information of the image through multiple iterations before reconstructing the final clear image using the estimated transmission rate and the hazy scene. Xu *et al.* [20] found that clouds have a higher signal-to-noise ratio, and they removed thin clouds in optical remote sensing images through noise-adjusted principal component transformation. These methods have been widely used and achieved some success in dehazing. However, they have limitations in dealing with complex and diverse hazy scenes, and low accuracy and robustness in the restoration of remote sensing images due to their reliance on handcrafted features and assumptions.

With the development of deep learning technology, researchers have begun to use deep learning techniques to solve the problem of multispectral image dehazing. CNNs with residual structures were applied to increase the depth of the model and improve its ability to extract haze features [21], effectively enhancing the dehazing performance of the model. Guo *et al.* [22] integrated local and global residual learning into the network, and introduced a channel attention module to capture strong channel correlations present in multispectral images. Zi *et al.* [23] employed a neural network designed based on the U-Net [24] architecture to estimate the thickness coefficients of each band with respect to the reference haze thickness map, thereby obtaining haze maps of different bands. The approach proposed in [25] leverages haze distortion control functions and a multiscale grid architecture to intelligently model and restore haze-free images while preserving the spatial structures of ground objects. These techniques aim to alleviate the degradation caused by haze in multispectral remote sensing images. Nevertheless, the straightforward application of these techniques to haze removal in HSIs is a formidable task, mainly due to the augmented spectral complexity inherent in HSIs.

The haze present in remote sensing images is spatially variant and its effect is related to the wavelength, with a tendency to decrease as the wavelength increases [26]. This phenomenon causes color distortion in the image and affects the performance of traditional dehazing methods that are based on hand-crafted features. The impact of haze in HSIs is more complex than that in multispectral images. Different bands are affected by haze to varying degrees, making it more challenging to remove haze from all bands compared to multispectral images. Furthermore, the bands in hyperspectral images are highly correlated, which means that there is a significant overlap of information between them, making it easier to overfit the model. Moreover, HSIs typically contain hundreds of bands, requiring more feature information to reconstruct a clear HSI during the haze removal process than in multispectral images. This necessitates the model to have stronger spatial and spectral feature extraction capabilities to achieve better haze removal results. Therefore, applying the multispectral image haze removal method directly to HSIs may lead to suboptimal results. To address these challenges, we propose a new HSI haze removal method called Asymmetric Attention Convolutional Network (AACNet). Our proposed AACNet can extract comprehensive and complex spatial and spectral features, capture the correlations between different channels to prevent overfitting, and finely adjust the weights of different channels, effectively reconstructing the spectral response curves of the haze-contaminated regions. The three main contributions of this work are described as follows:

1) The backbone network architecture of AACNet is mainly composed of several residual asymmetric attention groups (RRAG), with residual asymmetric attention blocks (RAAB) being the core component of each RRAG. The proposed RAAB adopts a multi-path connection to fully utilize the deep spatial-spectral features and uses residual connections and attention mechanisms to skip low-frequency regions and focus on high-frequency regions.
2) In RAAG, asymmetric attention convolution (AAConv) is the most critical component, which consists of fusion asymmetry convolution (FAConv) and global attention (GA). During the training phase, the proposed FAConv utilizes multi-path connections to fully exploit the deep spatial-spectral features. In the deployment phase, by employing the structural re-parameterization strategy, FAConv can be precisely and losslessly converted to a memory-friendly single-path convolution, which improves the model's running speed while ensuring the same performance. GA enables AAConv to pay more attention to the hazy regions of HSIs.
3) To more accurately restore the spectral information of regions polluted by haze, a pooling channel self-attention (PCSA) module is designed to capture the correlations between different channels and finely adjust the weights of different channels. The developed PCSA can effectively reconstruct the spectral response curve affected by pollution. The proposed method has been evaluated on various simulated and real datasets, and the experimental results show that AACNet outperforms state-of-the-art dehazing methods in terms of both qualitative and quantitative evaluation metrics.

The rest of the paper is organized below. Section II introduces related work on hyperspectral image dehazing, atten-

tion mechanism, and structural re-parameterization. Section III details the proposed approach. Section IV validates the effectiveness of our approach on both simulated and real datasets. Finally, Section V concludes the paper.

II. Related Works

A. Dehazing Models for HSIs

There is limited literature on dehazing for HSIs. Furthermore, the majority of these studies have relied on model-driven approaches. In [27], the authors first calculated a haze density map by differentiating the average bands belonging to the visible and infrared spectral ranges. The haze intensity map mainly reflects the spatial distribution of haze in the hazy image. Then, the haze abundance and density in different spectral bands were estimated based on the pixel reflectance between two manually selected pixels with different haze levels. Finally, the haze-free HSIs were obtained by subtracting the product of the estimated haze abundance and density maps from the original hazy target image. To automatically estimate the concentration of haze, Li *et al.* [28] proposed a saliency measure method that utilizes the reflectance differences between hazy and haze-free pixels. The RGB bands are extracted from the original hyperspectral image and transformed into the Intensity-Hue-Saturation (IHS) space. Then, the saliency maps for haze-free and hazy pixels are computed. Ultimately, the haze intensity map is integrated with the saliency maps to estimate the reflectance of hazy and haze-free pixels.

Another work based on a spectral grouping network (SGNet) was proposed by Ma *et al.* [29] by utilizing the useful information of each spectral band. SGNet divided HSIs into spectral subsets based on their correlation and performed parallel convolution with multiple branches for feature extraction. To transfer useful information between adjacent branches, a novel attention block was designed. Despite the effectiveness of SGNet, its spectral grouping strategy may compromise the continuity of spectral bands, leading to reduced accuracy in reconstructing HSIs. To overcome these limitations, we propose AACNet, a model that can be applied to haze removal from HSIs with various numbers of bands without grouping, while preserving the continuity of the spectra. AACNet consists of multiple RRAGs, where RAAB is the core component of each RRAG. The proposed RAAB utilizes a multi-path connection to fully leverage deep spatial-spectral features, and employs residual connections and attention mechanisms to skip low-frequency regions and focus on high-frequency regions. Moreover, AACNet can preserve the original features of the HSI through skip connections.

B. Attention Mechanism

The attention mechanism has gained significant attention in remote sensing image processing [30]–[33] due to its ability to selectively emphasize informative features and suppress irrelevant ones, inspired by human perceptual mechanisms. Previous research has explored the effectiveness of attention mechanisms in various remote sensing applications, such as HSI denoising [34], change detection [35], image classification [36], and band selection [37]. However, to the best of our knowledge, attention mechanisms have not been applied to the HSI dehazing problem. In this research paper, we present the PCSA block, which aims to accurately reconstruct the spectra of HSIs contaminated by haze. The PCSA block uses a spatial pooling operation to obtain a channel descriptor, assigns a weight to each channel, and calculates the self-attention matrix of the channel descriptor to finely adjust the weights of different channels and effectively reconstruct the spectra contaminated by haze.

C. Structural Re-parameterization

Structural Re-parameterization is a technique that parameterizes a structure using parameters transformed from another structure [38], [39]. Although this approach may result in heavy parameters, large memory consumption, and slow running speed during model training, it can ensure the model has strong representation capabilities. By using the structural re-parameterization technique during deployment, the model can inherit powerful representation capabilities and achieve a fast running speed. Various such approaches have been proposed to improve the performance of convolutional neural networks. ExpandNets [40] expanded the traditional convolutional layers into three consecutive convolutional layers, preserving the advantages of different sensory field size convolutions to enhance model performance. Re-Parameterization visual geometry group (ReVGG) network [41] added a multi-branch topology to the VGG model, which effectively improved its performance [42]. Asymmetric convolutional network (ACNet) [43] enhanced the robustness of traditional convolution against rotational distortions using one-dimensional asymmetric convolution. During deployment, ACNet is equivalent to a traditional convolution. To enhance the capability of exploring the deep spatial-spectral features of HSIs, we propose FAConv. During the model training phase, FAConv utilizes a multi-branch structure to aggregate features extracted from different branches, effectively improving the model's ability to extract spatial and spectral features. In the deployment phase, FAConv can be mathematically converted into a single 3×3 convolution without compromising accuracy, leading to a faster runtime.

III. Proposed Method

A. Architecture of AACNet

Let $\mathcal{X} \in \mathbb{R}^{H\times W\times C}$ represent the hazy hyperspectral image and $\mathcal{Y} \in \mathbb{R}^{H\times W\times C}$ represent the clear hyperspectral image. Here, C is the number of channels, and H and W are the height and width of the image, respectively. We assume that $\mathcal{X}$ and $\mathcal{Y}$ satisfy the following relationship:

$$\mathcal{X} = \mathcal{Y}t + \boldsymbol{a}(1-t) \tag{1}$$

where $\boldsymbol{a}$ is the global atmospheric light and t is the haze transmission diagram. Thus, a clear HSI can be obtained by the following equation:

$$\mathcal{Y} = \frac{\mathcal{X} - \boldsymbol{a}}{t} + \boldsymbol{a} \tag{2}$$

In hyperspectral imaging, each pixel is represented by a spectrum of intensities at different wavelengths, and the presence of haze in the atmosphere causes the spectral intensities to

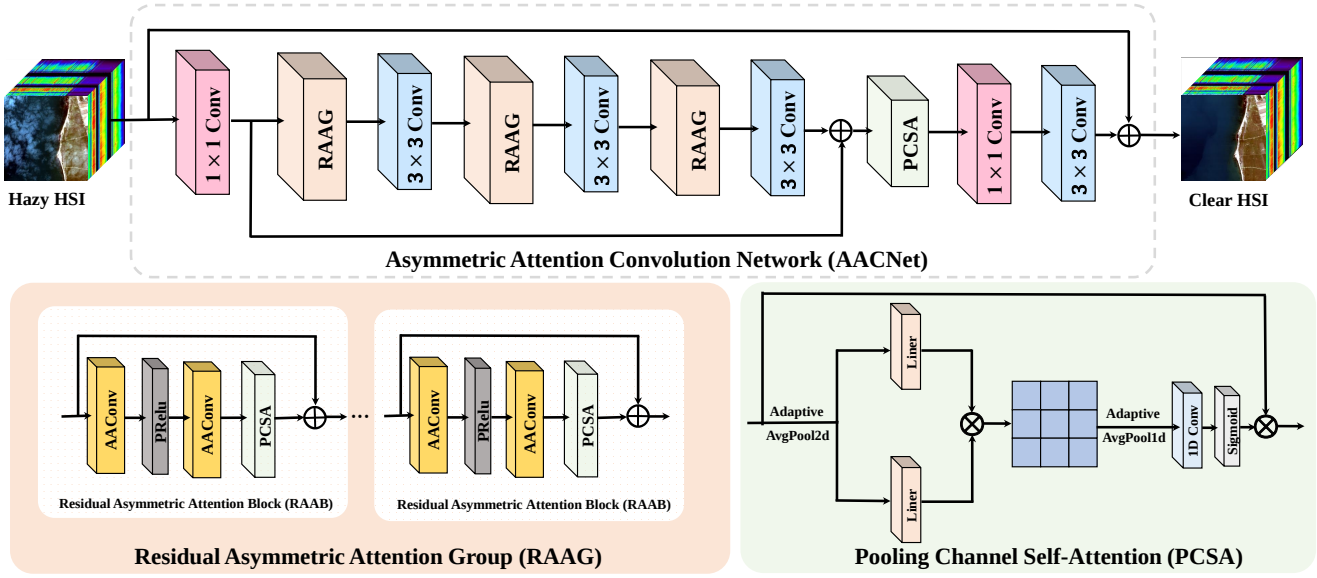


Fig. 1: The network architecture of the proposed asymmetric attention convolution network (AACNet). $\oplus$ denotes the element-wise summation. $\otimes$ denotes the element-wise product.

be attenuated in a wavelength-dependent manner. The amount of attenuation is determined by $\boldsymbol{a}$ and t, both of which are unknown and need to be estimated to perform dehazing. Thus, dehazing from hyperspectral images is an ill-posed problem. To address this issue, we propose a deep learning model called AACNet which is designed to learn deeper and more diverse spatial and spectral information, enabling it to capture a wider range of features compared to models that only focus on one aspect. Furthermore, AACNet can finely adjust the weights of different channels, which enhances its ability to efficiently predict clear hyperspectral images from haze ones.

As shown in Fig. 1, AACNet uses a combination of jump connections, convolutional layers, and attention mechanisms to perform haze removal on hyperspectral images. The jump connections are used to preserve the original features of the hazy hyperspectral image, which can be important for maintaining the overall structure and content of the image. The 1×1 convolutional layer is then used to perform shallow feature extraction, which can help to identify more specific features related to the presence of haze. The three RAAGs and 3×3 convolutional layers connected in series can help to further extract features from the input image, with the RAAGs allowing the model to selectively attend to different parts of the input based on their relevance to the task. The proposed PCSA, 1×1 convolutional layer, and 3×3 convolutional layer are used to further extract and refine the features obtained from the RAAGs. The resulting features are then added to the shallow features obtained previously, allowing the model to take into account both shallow and deeper features in the image. Finally, the resulting features are added to the original features of the hyperspectral image to obtain a clear image.

The RAAG in AACNet is composed of n RAABs, each of which consists of local residual learning, asymmetric attention convolution (AAConv), PRelu, and PCSA. The RAAB utilizes local residual learning to bypass the hazy regions and preserve the regions that are not contaminated by haze. PRelu activation function can help to introduce non-linearity into the model and improve its ability to capture complex relationships between the input features. The combination of AAConv and PCSA can fully explore the spatial and spectral features of the hazy regions and help the model better understand the distribution and characteristics of the haze in the image ultimately leading to more effective haze removal for HSIs.

B. AAConv

As shown in Fig 2, during the training phase AAConv consists of two parts: FAConv and GA. The input features are both fed into FAConv and GA, and the final output features are obtained by cross-multiplying them. In FAConv, we parallelized 3×3 convolution, 1×3 convolution, 3×1 convolution, and 1×1 convolution. This parallel design architecture allows FAConv to learn deeper and more diverse spatial and spectral information compared to ordinary convolutions. When a network has multiple parallel branches, each branch increases the memory consumption of the network, and this may increase the model's running time and deployment costs. Fortunately, using structural re-parameterization techniques, a complex multi-branch network can be transformed into a simple single-branch network without loss during the model deployment phase. Fig. 3 illustrates how the deployment phase FAConv converts the multi-branch training phase FAConv in Fig. 2 into an equivalent single-branch 3×3 convolution process. Once the training phase FAConv in Fig. 2 is trained, the parameters of each convolution are determined. Then, in the structural re-parameterization process shown in Fig. 3, the parameters of 3×3 convolution, 1×3 convolution, 3×1 convolution, and 1×1 convolution are added together and

assigned to the 3×3 convolution, resulting in a new 3×3 convolution that incorporates the rich spatial and spectral contextual information extracted by the training phase FAConv. This 3×3 convolution, which fuses the parameters of four parallel convolutions, inherits the superior performance of the training phase FAConv. Below we will provide a detailed explanation of how to convert the multi-branch FAConv into a 3×3 convolution during model deployment.

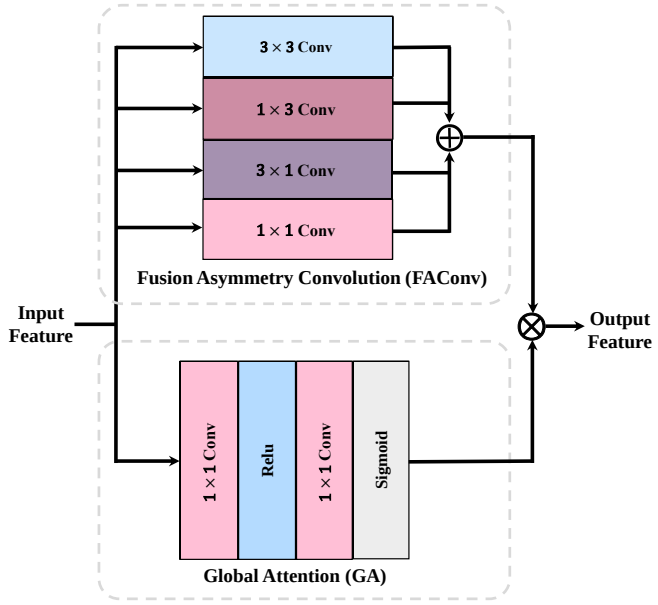


Fig. 2: The architecture of the proposed AAConv, which consists of training phase FAConv and GA. $\oplus$ denotes the element-wise summation. $\otimes$ denotes the element-wise product.

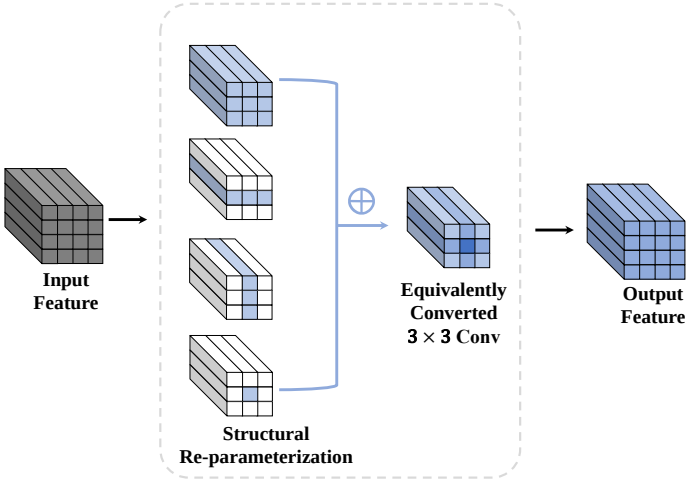


Fig. 3: Deployment phase FAConv with structural re-parameterization. $\oplus$ denotes the element-wise summation.

The d-th convolutional kernel is denoted as $\mathcal{K}^{(d)} \in \mathbb{R}^{K \times K \times C}$ ($d = 1, 2, ..., D$) , where D represents the number of convolutional kernels, K represents the size of convolution kernel. The process of the 3×3 convolutional layer can be represented as:

$$\mathbf{O}_{:,:,d} = \sum_{c=1}^{C} \mathbf{F}_{:,:,c} \ * \ \mathcal{K}_{:,:,c}^{(d)} \tag{3}$$

where $*$ represents the two-dimensional convolution operation, c and d are the indices of C and D, respectively, and $\mathbf{O}_{:,:,d}$ represents the feature map output by the d-th convolution kernel $\mathcal{K}_{:,:,c}^{(d)}$ convolved with the intermediate feature map $\mathbf{F}_{:,:,c}$. For simplicity, we have omitted the bias term in the formula. The d-th convolutional kernel of the 1×3 convolutional layer in the model is denoted as $\mathcal{K}^{v(d)} \in \mathbb{R}^{1 \times 3 \times C}$, the d-th convolutional kernel of the 3×1 convolution is denoted as $\mathcal{K}^{h(d)} \in \mathbb{R}^{3 \times 1 \times C}$, and the d-th convolutional kernel of the 1×1 convolution is denoted as $\mathcal{K}^{p(d)} \in \mathbb{R}^{1 \times 1 \times C}$. The entire convolutional process of these three types of convolutions can be respectively represented as:

$$\mathbf{O}_{:,:,d}^{v} = \sum_{c=1}^{C} \mathbf{F}_{:,:,c} \ * \ \mathcal{K}_{:,:,c}^{v(d)} \tag{4}$$

$$\mathbf{O}_{:,:,d}^{h} = \sum_{c=1}^{C} \mathbf{F}_{:,:,c} \ * \ \mathcal{K}_{:,:,c}^{h(d)} \tag{5}$$

$$\mathbf{O}_{:,:,d}^{p} = \sum_{c=1}^{C} \mathbf{F}_{:,:,c} \ * \ \mathcal{K}_{:,:,c}^{p(d)} \tag{6}$$

where $\mathbf{O}_{:,:,d}^{v}$, $\mathbf{O}_{:,:,d}^{h}$, and $\mathbf{O}_{:,:,d}^{p}$ represent the output feature maps obtained by performing two-dimensional convolution operations between $\mathcal{K}^{v(d)}$, $\mathcal{K}^{h(d)}$, $\mathcal{K}^{p(d)}$ and $\mathbf{F}_{:,:,c}$. As exhibited in Fig. 3, a deployment phase FAConv can be replaced by an equivalent 3×3 convolution:

$$\mathbf{O}_{:,:,d}^{'} = \mathbf{O}_{:,:,d} + \mathbf{O}_{:,:,d}^{v} + \mathbf{O}_{:,:,d}^{h} + \mathbf{O}_{:,:,d}^{p} \tag{7}$$

$$\mathcal{K}_{:,:,c}^{'(d)} = \mathcal{K}_{:,:,c}^{(d)} + \mathcal{K}_{:,:,c}^{v(d)} + \mathcal{K}_{:,:,c}^{h(d)} + \mathcal{K}_{:,:,c}^{p(d)} \tag{8}$$

$$\mathbf{O}_{:,:,d}^{'} = \sum_{c=1}^{C} \mathbf{F}_{:,:,c} * \ \mathcal{K}_{:,:,c}^{'(d)} \tag{9}$$

where $\mathcal{K}_{:,:,c}^{'(d)}$ is the d-th 3×3 convolution kernel obtained after equivalent transformation and $\mathbf{O}_{:,:,d}^{'}$ is the feature map output by performing a two-dimensional convolution operation between $\mathcal{K}_{:,:,c}^{'(d)}$ and $\mathbf{F}_{:,:,c}$.

The multi-branch convolutional structure can use different convolutional kernels to extract different spatial and spectral features from the input data, thereby enhancing the network's representational power. Each branch is then added and merged to generate more comprehensive and complex spatial and spectral feature representations. After converting all FAConv modules during deployment, the equivalently converted 3×3 convolution captures the same rich spatial and spectral contextual information as the multi-branch structure used during training. Additionally, the converted convolution maintains the same level of efficiency as the standard 3×3 convolution during deployment.

In order to make AAConv pay more attention to the pixels in hyperspectral images that are affected by haze pollution

TABLE I: Quantitative results of different methods on the simulated validation and test dataset. The best results are highlighted in bold.

Models	Params(M)	Flops(G)	Validation				Test			
			RMSE (↓)	UQI (↑)	SAM (↓)	SSIM (↑)	RMSE (↓)	UQI (↑)	SAM (↓)	SSIM (↑)
MSCNN	0.13	8.24	0.0379	0.6718	11.04	0.9108	0.0361	0.6911	10.27	0.9172
GCANet	0.79	24.58	0.0308	0.7984	4.62	0.9680	0.0287	0.8084	4.38	0.9656
SGNet	0.46	52.3	0.0149	0.9490	1.78	0.9802	0.0152	0.9425	2.02	0.9797
AODNet	4.36	286.15	0.0131	0.9421	2.02	0 9885	0.0115	0.9464	1.94	0.9892
PFDN	11.44	1.57	0.0106	0.9640	1.73	0.9824	0.0105	0.9614	1.77	0.9822
MIAN	200.94	109.25	0.0101	0.9499	1.77	0.9900	0.0095	0.9459	1.78	0.9903
AIDTransformer	10.90	203.23	0.0083	0.9627	1.58	0.9911	0.0088	0.9672	1.57	0.9911
GUNet_S	2.67	14.11	0.0035	0.9942	0.61	0.9983	0.0042	**0.9944**	0.65	0.9978
Ours	1.82	110.43	**0.0029**	**0.9952**	**0.56**	**0.9986**	**0.0033**	**0.9944**	**0.61**	**0.9982**

without increasing the model's computational load and the number of parameters, we designed a simple GA structure consisting of two 1×1 convolutional layers, Relu activation function, and Sigmoid activation function in series. Although GA does not significantly increase the model's computational load, it can significantly improve its performance. AAConv, as the main component, gradually extracts advanced spatial patterns and spectral features, while adaptively adjusting the weights of each pixel to focus more on the areas affected by haze.

C. PCSA

In order to improve the accuracy of restoring spectral information in regions affected by haze, we developed the PCSA module as shown in Fig. 1. The module is proposed to self-learn the correlations between different channels and finely adjust the weights of each channel. Two-dimensional average pooling is designed to transform the global spatial information into the channel descriptor, which can be represented as:

$$p_c = \frac{1}{H \times W} \sum_{i}^{H} \sum_{j}^{W} \mathbf{F}_p^{i,j,c} \tag{10}$$

where $\mathbf{F}_p^{i,j,c}$ represents the value of the c-th channel of feature $\mathbf{F}_p$ at coordinate (i, j), p_c represents the channel descriptor. Two-dimensional average pooling reduces the size of the input feature map from $H \times W \times c$ to $1 \times 1 \times c$.

To finely adjust the weights of each channel, a linear transformation is applied to the pooled feature map, followed by computing the self-attention matrix for channel vectors. The process of computing the self-attention matrix $\mathbf{A}$ can be represented as follows:

$$Q = \mathbf{W}^q p_c,\ K = \mathbf{W}^k p_c,\ \mathbf{A} = K^T \times Q \tag{11}$$

where Q and K represent the vectors obtained after linear transformations $\mathbf{W}^q$ and $\mathbf{W}^k$.

After performing one-dimensional average pooling on $\mathbf{A}$ to obtain a vector, it is input into a one-dimensional convolution and passed through a Sigmoid activation function to obtain the finely-tuned channel weights. Finally, the channel weights are multiplied with the corresponding elements of the input features to obtain the final output of PCSA. The process of one-dimensional pooling can be represented as follows:

$$\mathbf{F}_p^* = (\sigma(\text{1DConv}(v_c))) \times \mathbf{F}_p \tag{12}$$

where v_c represents the vector obtained after one-dimensional pooling, σ is the Sigmoid function, 1DConv is one-dimensional convolution and $\mathbf{F}_p^*$ is the final output of PCSA.

IV. Experiments

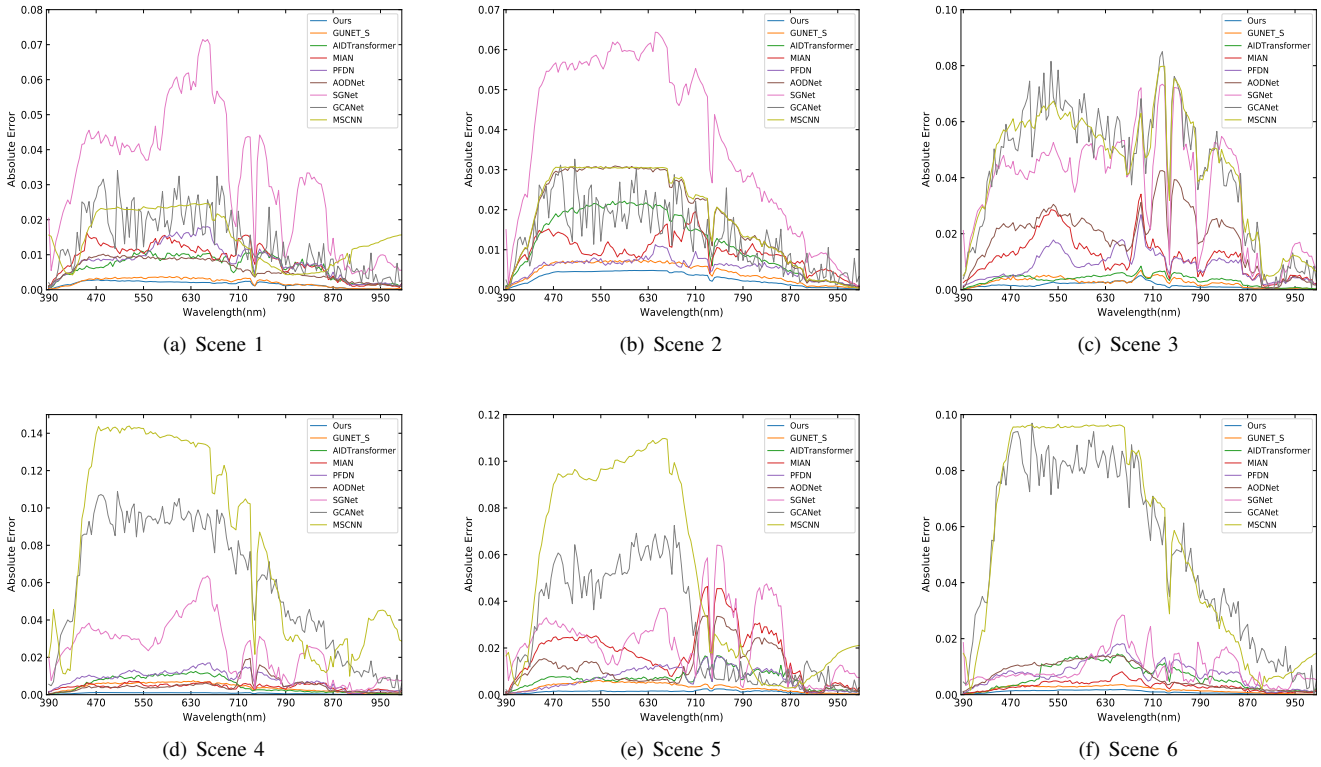
A. Datasets

In this study, we evaluate the performance of our proposed method using both simulated and real datasets. Specifically, we leverage Chinese Gaofen-5 satellite data to construct a baseline simulated haze hyperspectral dataset. Additionally, we generate a real haze hyperspectral dataset by utilizing the Earth Observing-1 (EO-1) satellite data obtained from the United States Geological Survey (USGS). Detailed descriptions of these datasets are provided below.

1) Simulated Dataset: To quantitatively evaluate the dehazing performance, a simulated haze dataset is required due to the unavailable ground truth data after haze removal under realistic conditions. We use the cloud simulation method in [44] to generate the simulated hazy hyperspectral images. First, we randomly selected the cirrus band from a Landsat-8 image as the haze intensity map. Next, the spectral vector of real haze from a high-resolution Gaofen-5 image was extracted. Then, we multiplied the haze intensity map with the spectral vector of haze to obtain a pure haze image. Finally, we added the pure haze image to the haze-free hyperspectral image to obtain the haze-contaminated hyperspectral image. During the fusion process, we used a small coefficient to adjust the thickness of the haze. Finally, a total of 210 clear Gaofen-5 images with a size of $256 \times 256 \times 150$ were employed as the ground truth. The corresponding simulated hazy images were produced by overlaying a haze patch from the hazy hyperspectral image, in which 170 were used for training, 20 each were set aside for validation and testing purposes. Please refer to [44] for more details of the simulation process.

2) Real Dataset: In order to implement AACNet on the real remote sensing data, we built a dataset utilizing EO-1 Hyperion images, which have 242 continuous spectral bands with wavelengths from 0.35 μm to 2.58 μm and a spatial

Fig. 4: Haze removal results of six scenes obtained by different methods. The first and second rows are the images contaminated by haze and their corresponding clear images. The third to eleventh rows show true-color composite images of Band 20 (472 nm), Band 35 (536 nm), and Band 65 (664 nm) of the HSI after dehazing.



(a) Scene 1 (b) Scene 2 (c) Scene 3

(d) Scene 4 (e) Scene 5 (f) Scene 6

Fig. 5: The absolute error curve of the spectral response corresponds to the six scenes in Fig. 4. The absolute error curve of the spectral response is obtained by calculating the absolute error of the average spectral reflectance of hazy areas with a spatial size of 5×5 pixels.

resolution of 30 m. We collected 43 pairs of hazy and clear hyperspectral images, where the clear images were chosen at the same location as the corresponding hazy image, but captured on the nearest dates. Out of the 43 pairs, 37 pairs were used as the training samples and 6 pairs were test samples. Each hyperspectral image in the dataset was cropped to a size of $256 \times 256 \times 242$. Both the simulated dataset and the real dataset include various land cover types, such as farmland, rivers, mountains, cities, plains, and lakes, among others. This ensures sufficient heterogeneity in the dataset, allowing us to evaluate the generalization performance of our proposed model.

B. Implementation Settings

The paired hazy and clear HSIs in simulated and real datasets are all cropped to a spatial size of 64×64 for training and testing. The training batch size is set as 32. The proposed AACNet model is optimized using the ADAM [45] optimizer, where $beta_1 = 0.9, beta_2 = 0.99$ and $epsilon = 10^{-8}$. The initial learning rate is set to 0.0002, and the training stops after 50 epochs. We found experimentally that the model works best when the number of RAABs is 5, so we set n = 5. All experiments are completed on the PyTorch framework via an NVIDIA A40 and Intel(R) Xeon(R) Silver 4314 CPU. In this paper, we optimize the network by adopting the mean squared error (MSE) as the loss function to measure hazy and clear HSIs.

C. Competing Methods

To evaluate the effectiveness of the proposed dehazing method, we compared it against several state-of-the-art dehazing methods in both computer vision and remote sensing domains, including multi-scale convolutional neural networks (MSCNN) [12], gated context aggregation network (GCANet) [14], spectral grouping network (SGNet) [29], all-in-one network (AODNet) [46], physics-based feature dehazing networks (PFDN) [47], multi-input attention network (MIAN) [48], aerial image dehazing Transformer (AIDTransformer) [49] and a variant network of U-Net [24] for dehazing (GUNET_S) [16]. These dehazing methods were selected as our competitors based on their high citation count. All parameters were set according to the author's recommendations in their papers. MSCNN, GCANet, and AODNet are three classic CNN-based dehazing networks for natural images. SGNet divides each HSI into multiple spectral subsets and inputs them to parallel branch convolutions for feature extraction to obtain the dehazed HSI. PFDN employs an encoder-decoder architecture with residual learning and incorporates a physical model unit that simulates the haze formation process to dehaze natural images. MIAN is a multi-input convolutional neural network based on the encoder-decoder architecture designed for multispectral remote sensing image dehazing. AIDTransformer is an aerial image dehazing network based on the vision Transformer [50] architecture. GUNET_S replaces the convolution blocks in U-Net with residual blocks and

introduces attention mechanisms to improve the network's representation capability, and is a recent dehazing network designed for natural images.

D. Experiments on the Simulated Dataset

1) Evaluative Metrics: Four widely used evaluation metrics, including root mean square error (RMSE), universal quality index (UQI) [51], spectral angle mapper (SAM) [52] and structural similarity index measure (SSIM) [53] were adopted:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(H_i - \hat{H}_i)^2} \tag{13}$$

$$\mathrm{UQI} = \frac{4\sigma_{xy}\mu_x\mu_y}{(\sigma_x^2+\sigma_y^2)(\mu_x^2+\mu_y^2)} \tag{14}$$

$$\mathrm{SAM} = \cos^{-1}\left(\frac{\mathbf{t}\cdot\mathbf{p}}{|\mathbf{t}||\mathbf{p}|}\right) \tag{15}$$

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{16}$$

where n is the number of pixels in the image, H_i is the true value for pixel i, $\hat{H}_i$ is the predicted value for pixel i. σ_{xy} is the covariance between the two images, σ_x^2 and σ_y^2 are the variances of the two images, and μ_x and μ_y are the means of the two images. In SAM computation, $\cos^{-1}$ is the inverse cosine function, $\mathbf{t}$ is the spectral signature of the true spectrum, $\mathbf{p}$ is the spectral signature of the predicted spectrum. In SSIM computation, c_1 and c_2 are constants to avoid instability when the denominator is close to zero.

RMSE reflects the average absolute error between the reconstructed image and the ground truth image, with lower values indicating better reconstruction quality. UQI measures the similarity in terms of luminance, contrast, and structure. As the UQI value increases, the image quality is expected to improve accordingly. SAM measures the spectral similarity. A lower SAM value indicates a higher spectral similarity between the reconstructed image and the ground truth image. SSIM calculates the structural similarities, with higher values indicating better reconstruction quality. It is worth noting that all metrics were computed by taking an average over all channels.

2) Quantitative Results: Table I shows the quantitative results of all comparison methods on the simulated validation and test dataset. The results presented in Table I indicate that our method consistently achieves superior performance over all other dehazing methods in the validation dataset across all metrics. In the test dataset, our method demonstrates comparable performance to GUNET_S in terms of the UQI metric, while outperforming other dehazing methods in other metrics. Compared to the representative MSCNN method, AACNet reduces RMSE from 0.0379 to 0.0029, increases UQI from 0.6718 to 0.9952, decreases SAM from 11.4 to 0.56, and increases SSIM from 0.9108 to 0.9986 on the validation dataset. Similarly, on the test dataset, AACNet reduces RMSE from 0.0361 to 0.0033, increases UQI from 0.6911 to 0.9944, decreases SAM from 10.27 to 0.61, and increases SSIM from 0.9172 to 0.9982. Compared to the latest GUNET_S method, AACNet reduces RMSE by 0.0006, increases UQI by 0.001, decreases SAM by 0.05, and increases SSIM by 0.0003 on the validation dataset. On the test dataset, AACNet reduces RMSE by 0.0009, decreases SAM by 0.004, and increases SSIM by 0.0004. The parameters of AACNet decreased by 0.85M and 9.08M compared to GUNet_S and AIDTransformer, respectively. The floating-point operands of AACNet decreased by 92.8G compared to AIDTransformer. Overall, our proposed AACNet achieves a balance between dehazing performance and model complexity.

3) Qualitative Results: For visual evaluation, we randomly selected six different scenes with varying levels of complexity from the validation and test datasets, such as mountains, lakes, urban, and farmlands. To investigate the visual results of different methods, we selected Band 20 (472 nm), Band 35 (536 nm), and Band 65 (664 nm) from the HSI to composite a true-color image as shown in Fig. 4. It can be seen that MSCNN and GCANet can reduce the thickness of haze in hazy regions, but a significant amount of haze was still retained in the image. In scene 2, SGNet, AIDTrasformer and GUNet_S showed obvious color distortion, and AODNet had haze remaining in the removed image. In scenes 3 and 4, SGNet, PFDN and MIAN exhibited color distortion, and AODNet had a small amount of haze remaining. In scene 5, SGNet, AODNet, PFDN, MIAN and AIDTrasformer all show obvious color distortion. In scene 6, SGNet, AODNet, and PFDN showed brighter colors and were not as clear as the reference image in the marked area (highlighted by a red box). Overall, the haze-free images recovered by our proposed AACNet method had the best visual results and were closest to the reference image.

To further compare the spectral reconstruction performance of different methods, we calculated the absolute error between the average spectral reflectance of regions with a spatial size of 5×5 pixels in different HSIs and the spectral reflectance of the corresponding regions in the reference HSI (Fig. 5). Our analysis showed that the spectral response curves restored by MSCNN and GCANet differed significantly from the curves of the reference HSI, indicating poor spectral reconstruction performance. The spectral response curves restored by SGNet, AODNet, PFDN, MIAN, AIDTrasformer and GUNET_S were relatively close to the curves of the reference HSI, but the spectral reconstruction performance was poor in bands with high spectral reflectance. In contrast, our proposed AACNet method restored a spectral response curve that was closer to the curve of the reference HSI, indicating superior spectral reconstruction performance and the ability to effectively reconstruct spectra contaminated by haze pollution.

Fig. 6 shows the absolute error heatmaps of six scenes recovered by different methods to the corresponding ground truth image. The images reconstructed by MSCNN and GCANet methods have significant errors in all scenes, while the HSI reconstructed by SGNet has large errors in scenes 1-5. The HSI reconstructed by PFDN and MIAN has noticeable errors in scenes 2 and 3. The HSI reconstructed by AIDTransformer has noticeable errors in scenes 2-6. The

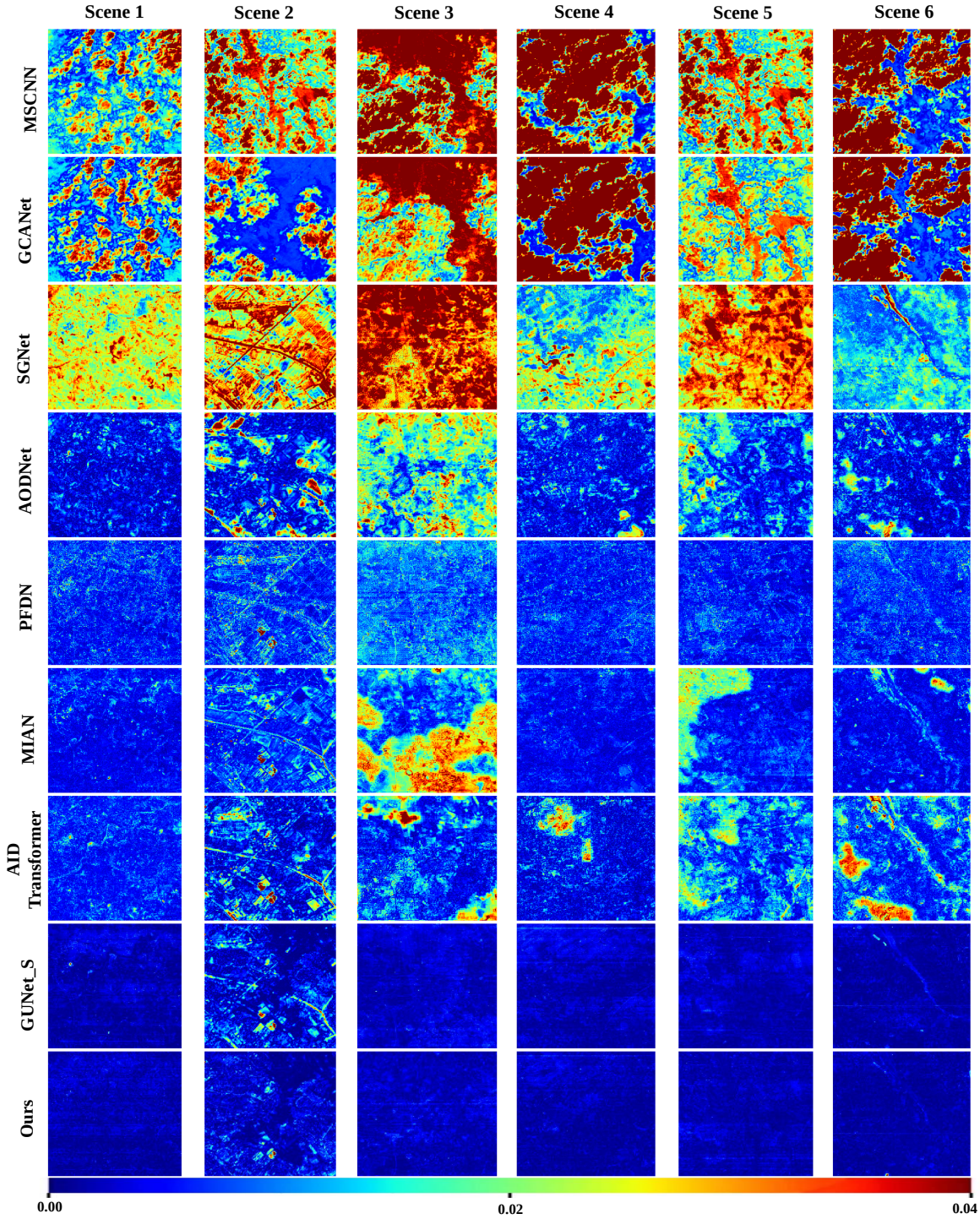


Fig. 6: The average absolute error heatmaps between the restored HSIs using nine different dehazing methods and the corresponding ground truth HSIs.

TABLE II: The ablation results on the Gaofen-5 test set.

Description	GA	FAConv	PCSA	RMSE	SSIM
$Backbone$	✗	✗	✗	0.0059	0.9974
E_a	✓	✗	✗	0.0058	0.9975
E_b	✓	✓	✗	0.0051	0.9976
E_c	✓	✗	✓	0.0045	0.9977
E_d	✓	✓	✓	0.0033	0.9982

HSI reconstructed by GUNet_S has much larger errors in scene 2 compared to the image reconstructed by our proposed AACNet. Obviously, our analysis demonstrated that the haze-free HSIs reconstructed by AACNet had the smallest error compared to the ground truth image Additionally, our proposed AACNet outperforms other methods in terms of dehazing performance. Overall, the reconstruction results of all methods in Scene 2, as indicated by the heatmaps, are the worst compared with other scenes. This is primarily due to the presence of diverse land types such as trees, farmland, and rivers in Scene 2. Additionally, the texture information in Scene 2 images is complex. These two challenges make the reconstruction even more difficult. However, our model still achieves the best result among all methods in Scene 2.

4) Ablation Analysis: In this section, we demonstrate the impact of AAConv and PCSA on our proposed network through ablation studies on the simulated dataset. The experimental results on the test dataset are shown in Table II, where "$Backbone$" refers to the baseline network with a 3×3 convolution replacing AAConv and PCSA. We recorded the best RMSE and SSIM values. AAConv: We investigate the impact of the two sub-modules of AAConv, FAConv and GA, on the performance of the backbone model. Table II presents the evaluation metrics on the test dataset, where the values of RMSE and SSIM for the backbone network are 0.59 and 0.9974, respectively. We introduce E_a, which represents the model with GA merging into the backbone network, to unequally process different features and pixels, making the backbone network more attentive to the hazy areas and outperforming the baseline. Building on the results of E_a, we further conduct another experiment E_b to examine the impact of FAConv in our proposed model. Compared to the specific network E_a without FAConv, E_b achieves a decrease of 0.0007 in RMSE and an increase of 0.0001 in SSIM. Our experiments demonstrate that the AAConv composed of GA and FAConv achieves superior performance.

PCSA: Embedding PCSA into E_a and E_b respectively resulted in E_c and E_d. Compared to E_a, E_c exhibits a decrease of 0.0013 in RMSE and an increase of 0.0002 in SSIM. Similarly, E_d outperforms E_b by achieving a decrease of 0.0018 in RMSE and an increase of 0.0006 in SSIM. These results demonstrate that PCSA contributes to the spectral reconstruction of the hazy area and improves the accuracy of the restored spectra.

E. Experiments on the Real Dataset

The evaluation of haze removal methods on real HSIs is crucial, but obtaining real ground truth images is challenging, and therefore, visual comparison is generally employed. In Fig. 7, we present the results of seven dehazing methods applied to six different real HSI scenes. The visual results reveal that MSCNN performs similarly to its performance on the simulated dataset, and it is ineffective in removing haze in real HSIs. On the other hand, using GCANet for dehazing may result in blurry HSIs with residual haze. SGNet effectively removes haze from most scenes, but a small amount of haze remains in scene 10, and severe color distortion is observed in scenes 7 and 12. AODNet exhibits severe color distortion in scene 8, and there is still haze in scenes 9 and 12. PFDN exhibits severe color distortion in scenes 8, 10, and 12. MIAN exhibits color distortion in scenes 7, 10, and 12, and noticeable noise in scenes 8 and 9. AIDTransformer suffers from severe color distortion in scenes 7 and 10. GUNet_S effectively removes haze in most scenes, but a small amount of haze remains in scenes 8 and 9. In summary, the visual results demonstrate that the comparison methods suffer from issues such as blurring, residual haze, and severe color distortion, making them unsuitable for practical applications.

In contrast, our proposed AACNet effectively removes haze and reconstructs objects contaminated by haze. Compared to other state-of-the-art methods, AACNet can reconstruct the HSI with colors closer to the clear image without causing blurring or residual haze. It can be seen from the simulated and real experimental results, AACNet showed better dehazing results on the crop and mountainous areas, which is mainly because of the simple texture structure in these areas. These results demonstrate that AACNet is an effective method for haze removal in real HSIs, and it can be used to significantly improve the quality of HSIs, especially in scenarios with challenging atmospheric conditions.

It is important to note that while the proposed method delivers promising results in terms of haze removal, it does exhibit slight color distortion in the restored HSIs due to the suboptimal quality of some training data. Moreover, the varying time periods between the hazy and clear images may result in changes in ground objects, causing inconsistent ground objects in the image pairs. This highlights the need for high-quality datasets to improve haze removal results. The presence of complex atmospheric conditions such as multi-layered haze and non-uniform haze distribution can also affect the performance of haze removal algorithms. There are other challenges associated with the practical application of haze removal algorithms in real-world scenarios. Nonetheless, the proposed method demonstrates outstanding performance in removing haze from real datasets, effectively restoring the obscured ground objects, and significantly improving the overall quality of HSIs.

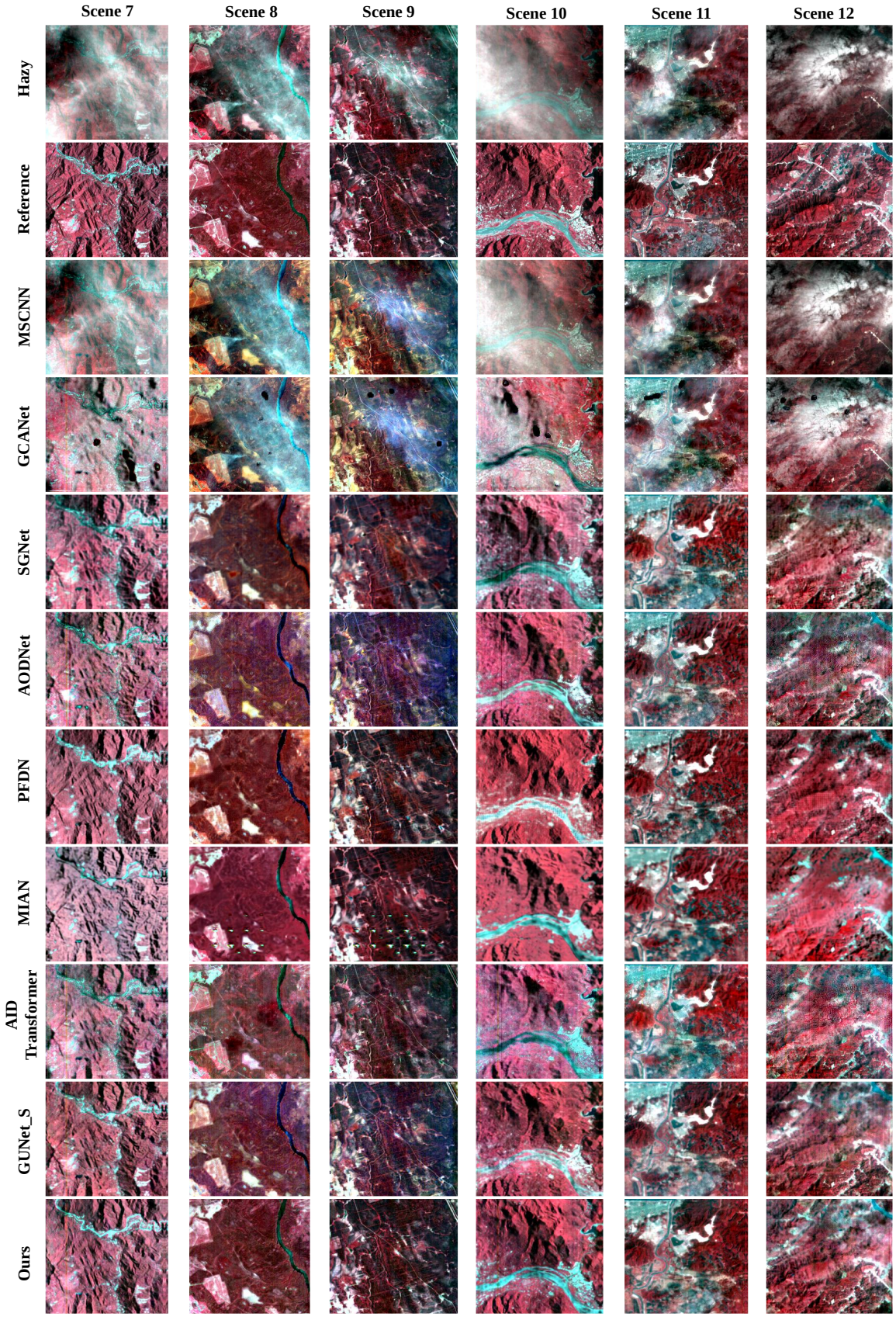


Fig. 7: Haze removal results obtained by different methods on six EO-1 Hyperion image scenes. The first row is the image contaminated by real haze. The second row is the clear reference image collected on the adjacent date. The third to eleventh rows show the false-color composite images after dehazing by nine methods.

V. Conclusion

Compared to haze removal in computer vision or multispectral remote sensing fields, haze removal in hyperspectral images presents a greater challenge. Despite this, the number of studies on haze removal in HSIs is limited. To address this gap, this paper proposes the AACNet model, which is tailored specifically for HSI haze removal. The model's backbone structure is composed of multiple RAAGs, each containing multiple RAABs, which effectively exploit the spatial and spectral correlations between clear and polluted HSIs affected by haze. During deployment, the trained AACNet can successfully convert all the multi-branch FAConvs to a 3×3 convolution while maintaining the powerful feature extraction ability of FAConv. Furthermore, to further improve the haze removal performance of AACNet, we developed a PCSA block to capture the spectral correlations between different spectra of the intermediate features. This block adaptively enhances important spectral features while suppressing redundant features, improving the spectral reconstruction quality of hazy regions. The effectiveness of AACNet for haze removal was validated on both simulated and real hyperspectral datasets, demonstrating superior performance compared to other advanced haze removal methods. Our work highlights the importance of addressing the unique challenges of haze removal in hyperspectral images and provides a promising solution for future research in this area. Overall, the proposed AACNet model showcases the potential for further advancements in haze removal for HSIs. Of course, our work has some limitations. First, the proposed method requires a large amount of Graphics Processing Unit (GPU) memory during training due to the use of multi-branch convolution to improve performance and a large amount of data in a single HSI. Second, the quality of training samples can significantly affect the results, which poses a challenge in creating high-quality hyperspectral training data. Third, various noises in HSIs may also affect the haze removal results.

References

[1] G. Yang, K. Huang, W. Sun, X. Meng, D. Mao, and Y. Ge, "Enhanced mangrove vegetation index based on hyperspectral images for mapping mangrove," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 189, pp. 236–254, 2022.

[2] S. Zhang, D. Tang, N. Li, X. Jia, and S. Jia, "Superpixel-guided variable gabor phase coding fusion for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2022.

[3] H. Huang, Y. Tang, Z. Tan, J. Zhuang, C. Hou, W. Chen, and J. Ren, "Object-based attention mechanism for color calibration of UAV remote sensing images in precision agriculture," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[4] Y. T. Solano-Correa, F. Bovolo, L. Bruzzone, and D. Fernández-Prieto, "A method for the analysis of small crop fields in sentinel-2 dense time series," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 3, pp. 2150–2164, 2020.

[5] S. Veraverbeke, P. Dennison, I. Gitas, G. Hulley, O. Kalashnikova, T. Katagis, L. Kuai, R. Meng, D. Roberts, and N. Stavros, "Hyperspectral remote sensing of fire: State-of-the-art and future perspectives," *Remote Sensing of Environment*, vol. 216, pp. 105–121, 2018.

[6] P. Duan, X. Kang, P. Ghamisi, and S. Li, "Hyperspectral remote sensing benchmark database for oil spill detection with an isolation forest-guided unsupervised detector," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–11, 2023.

[7] S. Zhang, P. Duan, X. Kang, Y. Mo, and S. Li, "Feature-band-based unsupervised hyperspectral underwater target detection near the coastline," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–10, 2023.

[8] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1956–1963.

[9] R. Fattal, "Dehazing using color-lines," *ACM transactions on graphics (TOG)*, vol. 34, no. 1, pp. 1–14, 2014.

[10] Q. Zhu, J. Mai, and L. Shao, "A fast single image haze removal algorithm using color attenuation prior," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3522–3533, 2015.

[11] D. Berman, S. Avidan *et al.*, "Non-local image dehazing," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1674–1682.

[12] W. Ren, S. Liu, H. Zhang, J. Pan, X. Cao, and M.-H. Yang, "Single image dehazing via multi-scale convolutional neural networks," in *ECCV*. Springer, 2016, pp. 154–169.

[13] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, "Dehazenet: An end-to-end system for single image haze removal," *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187–5198, 2016.

[14] D. Chen, M. He, Q. Fan, J. Liao, L. Zhang, D. Hou, L. Yuan, and G. Hua, "Gated context aggregation network for image dehazing and deraining," in *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019, pp. 1375–1383.

[15] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 07, 2020, pp. 11 908–11 915.

[16] Y. Song, Y. Zhou, H. Qian, and X. Du, "Rethinking performance gains in image dehazing networks," *arXiv preprint arXiv:2209.11448*, 2022.

[17] B. Zhou and Y. Wang, "A thin-cloud removal approach combining the cirrus band and rtm-based algorithm for landsat-8 oli data," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 1434–1437.

[18] F. Xie, J. Chen, X. Pan, and Z. Jiang, "Adaptive haze removal for single remote sensing image," *IEEE Access*, vol. 6, pp. 67 982–67 991, 2018.

[19] L. Xu, D. Zhao, Y. Yan, S. Kwong, J. Chen, and L.-Y. Duan, "Iders: Iterative dehazing method for single remote sensing image," *Information Sciences*, vol. 489, pp. 50–62, 2019.

[20] M. Xu, X. Jia, M. Pickering, and S. Jia, "Thin cloud removal from optical remote sensing images using the noise-adjusted principal components transform," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 149, pp. 215–225, 2019.

[21] M. Qin, F. Xie, W. Li, Z. Shi, and H. Zhang, "Dehazing for multispectral remote sensing images based on a convolutional neural network with the residual architecture," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 5, pp. 1645–1655, 2018.

[22] J. Guo, J. Yang, H. Yue, H. Tan, C. Hou, and K. Li, "Rsdehazenet: Dehazing network with channel refinement for multispectral remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 3, pp. 2535–2549, 2021.

[23] Y. Zi, F. Xie, N. Zhang, Z. Jiang, W. Zhu, and H. Zhang, "Thin cloud removal for multispectral remote sensing images using convolutional neural networks combined with an imaging model," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 3811–3823, 2021.

[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*. Springer, 2015, pp. 234–241.

[25] W. Yu, X. Zhang, and M.-O. Pun, "Cloud removal in optical remote sensing imagery using multiscale distortion-aware networks," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[26] J. Li, H. Shen, H. Li, M. Jiang, and Q. Yuan, "Radiometric quality improvement of hyperspectral remote sensing images: a technical tutorial on variational framework," *Journal of Applied Remote Sensing*, vol. 15, no. 3, pp. 031 502–031 502, 2021.

[27] X. Kang, Z. Fei, P. Duan, and S. Li, "Fog model-based hyperspectral image defogging," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–12, 2022.

[28] Z. Li, P. Duan, S. Hu, M. Li, and X. Kang, "Fast hyperspectral image dehazing with dark-object subtraction model," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[29] X. Ma, Q. Wang, and X. Tong, "A spectral grouping-based deep learning model for haze removal of hyperspectral images," *ISPRS journal of photogrammetry and remote sensing*, no. Jun., p. 188, 2022.

[30] S. Ghaffarian, J. Valente, M. Van Der Voort, and B. Tekinerdogan, "Effect of attention mechanism in deep learning-based remote sensing image processing: A systematic literature review," *Remote Sensing*, vol. 13, no. 15, p. 2965, 2021.

[31] R. Xu, Y. Tao, Z. Lu, and Y. Zhong, "Attention-mechanism-containing neural networks for high-resolution remote sensing image classification," *Remote Sensing*, vol. 10, no. 10, p. 1602, 2018.

[32] Z. Rao, M. He, Z. Zhu, Y. Dai, and R. He, "Bidirectional guided attention network for 3-D semantic detection of remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 7, pp. 6138–6153, 2020.

[33] H. Cao, X. Luo, Y. Peng, and T. Xie, "MANet: A network architecture for remote sensing spatiotemporal fusion based on multiscale and attention mechanisms," *Remote Sensing*, vol. 14, no. 18, p. 4600, 2022.

[34] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, "Hyperspectral image denoising using a 3-d attention denoising network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 12, pp. 10 348–10 363, 2021.

[35] G. Wang, Y. Peng, S. Zhang, G. Wang, T. Zhang, J. Qi, S. Zheng, and Y. Liu, "Pyramid self-attention mechanism-based change detection in hyperspectral imagery," *Journal of Applied Remote Sensing*, vol. 15, no. 4, pp. 042 611–042 611, 2021.

[36] W. Li, H. Chen, Q. Liu, H. Liu, Y. Wang, and G. Gui, "Attention mechanism and depthwise separable convolution aided 3dcnn for hyperspectral remote sensing image classification," *Remote Sensing*, vol. 14, no. 9, p. 2215, 2022.

[37] K. He, W. Sun, G. Yang, X. Meng, K. Ren, J. Peng, and Q. Du, "A dual global–local attention network for hyperspectral band selection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[38] X. Ding, X. Zhang, J. Han, and G. Ding, "Diverse branch block: Building a convolution as an inception-like unit," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 10 886–10 895.

[39] S. Zagoruyko and N. Komodakis, "Diracnets: Training very deep neural networks without skip-connections," *arXiv preprint arXiv:1706.00388*, 2017.

[40] S. Guo, J. M. Alvarez, and M. Salzmann, "Expandnets: Linear over-parameterization to train compact convolutional networks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 1298–1310, 2020.

[41] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "Repvgg: Making vgg-style convnets great again," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 13 733–13 742.

[42] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[43] X. Ding, Y. Guo, G. Ding, and J. Han, "Acnet: Strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1911–1920.

[44] M. Xu, X. Jia, M. Pickering, and A. J. Plaza, "Cloud removal based on sparse representation via multitemporal dictionary learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 5, pp. 2998–3006, 2016.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[46] B. Li, X. Peng, Z. Wang, J. Xu, and D. Feng, "AOD-Net: All-in-one dehazing network," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4770–4778.

[47] J. Dong and J. Pan, "Physics-based feature dehazing networks," in *ECCV*. Springer, 2020, pp. 188–204.

[48] Z. He, C. Gong, Y. Hu, F. Zheng, and L. Li, "Multi-input attention network for dehazing of remote sensing images," *Applied Sciences*, vol. 12, no. 20, p. 10523, 2022.

[49] A. Kulkarni and S. Murala, "Aerial image dehazing with attentive deformable transformers," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 6305–6314.

[50] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[51] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE signal processing letters*, vol. 9, no. 3, pp. 81–84, 2002.

[52] L. Loncan, L. B. De Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geoscience and remote sensing magazine*, vol. 3, no. 3, pp. 27–46, 2015.

[53] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

Meng Xu (Member, IEEE) received the B.S. and M.E. degrees in electrical engineering from the Ocean University of China, Qingdao, China, in 2011 and 2013, respectively, and the Ph.D. degree from the University of New South Wales, Canberra, ACT, Australia, in 2017.

She is currently an Associate Research Fellow with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. Her research interests include cloud removal and remote sensing image processing.

Yanxin Peng received the B.S. degree in information and computational science from Jiujiang University in 2021. He is currently pursuing a master's degree in Computer Technology at the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

His research interest is hyperspectral image reconstruction.

Ying Zhang received the B.E. degree from the Guangzhou University, Guangzhou, China, in 2022.He is pursuing the master's degree in computer technology with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

His research interests include hyperspectral object detection, machine learning, and deep learning.

Xiuping Jia (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering from the University of New South Wales, Canberra, ACT, Australia, in 1996.

Since 1988, she has been with the School of Information Technology and Electrical Engineering, University of New South Wales, where she is currently a Senior Lecturer. She is also a Guest Professor with Harbin Engineering University, Harbin, China, and an Adjunct Researcher with the China National Engineering Research Center for Information Technology in Agriculture, Beijing. She has coauthored the *remote sensing textbook Remote Sensing Digital Image Analysis* (Springer–Verlag, 3rd edition 1999 and 4th edition 2006). Her research interests include remote sensing, image processing, and spatial data analysis.

Dr. Jia is a Subject Editor of the *Journal of Soils and Sediments* and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.

Sen Jia (Senior Member, IEEE) received the B.E. and Ph.D. degrees from the College of Computer Science, Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively.

Since 2008, he has been with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, where he is currently a Full Professor. His research interests include hyperspectral image processing, signal and image processing, and machine learning.