Hyperspectral image denoising via spectral noise distribution bootstrap

Erting Pan, Yong Ma*, Xiaoguang Mei, Fan Fan, Jiayi Ma

Electronic Information School, Wuhan University, Wuhan 430072, China

A R T I C L E   I N F O

A B S T R A C T

Hyperspectral image (HSI) denoising is an ill-posed problem, leading to integrating proper *prior* knowledge about hyperspectral noise is critical to developing an efficient denoising method. Most existing methods share a common assumption that all bands have equal noise intensity. However, such assumption runs counter to the practical HSIs, leading to unpleasant denoising results. To tackle this, we intend to investigate the intrinsic properties of real HSI noise in the spectral dimension and construct a novel denoising framework bootstrapping by spectral noise distribution $\hat{N}$, termed $\hat{N}$-Net. On the one hand, we develop dense and sparse recurrent calculations, exploiting intrinsic properties of HSI noise (*i.e.*, diversity, dense dependency, and global sparsity) to estimate spectral noise distribution. On the other hand, having the estimated spectral noise distribution, we develop a bootstrap mechanism with a repetitive emphasis on its guidance for subsequent spatial noise separation and clean HSI recovery, ensuring a more delicate denoising effect. In particular, we verify that the proposed denoising framework can achieve promising denoising performances due to the merit of spectral noise distribution bootstrapping, which also promotes new insights for future related research. The code is avaiable at https://github.com/EtPan/N-Net.

1. Introduction

Benefiting from owing abundant spectral information, hyperspectral remote sensing technology has immeasurable potential in broad applications, like face recognition [1], spaceborne monitoring [2], sea-surface observation [3], and remote sensing interpretation [4,5]. However, constrained by physical and optical mechanism factors, the complex imaging chain of hyperspectral image (HSI) is inevitably affected by various factors such as atmospheric conditions, environmental illumination [6,7], and imaging sensors [8]. They may introduce multiple randomly distributed noises to HSIs, for instance, dense Gaussian noise, sparse noise or structural noise. Noises like these can adversely affect HSI, leading to a significant loss of visual quality. Worse still, it limits qualities of subsequent HSI interpretations, such as super-resolution [9], classification [10], and target detection [11].

Consequently, considering the varied noise degradation existing in HSIs, investigating a noise removal model with solid robustness has essential significance. It has attracted considerable academic attention in the past decades. A great deal of denoising techniques has been proposed, involving traditional algorithms [12–15] and recent deep learning-based algorithms [16–19], which are detailed reviewed in Section 2. There is a notable fact that HSI noise usually obeys a complex distribution in the spectral dimension. However, most existing denoising algorithms assume that all bands share the same noise intensity and simply treat them equally [20–22], which is divorced from the practical scenario to a greater or less extent and causes gross spectral distortion. Taking the EO-1 Hyperion dataset [23] as an example, we demonstrate Figs. 1 and 2 for auxiliary understanding. Some apparent noise characteristics can be summed up as follows: (1) Diversity. The noise intensities across different bands exhibit noticeable differences, which can be attributed to variations in sensor sensitivity across various HSI bands. (2) Dense local correlation. The noisy images and statistical histograms from band 88 to 96 in Fig.r1 reveal that adjacent bands exhibit strong correlations in noises. Since images on the contiguous spectrum are generally collected under slight deviations in both sensor settings and wavelengths, noises along adjacent bands consequently tend to be more or less similar. (3) Global sparsity. As illustrated in Fig. 1, it is easy to note that images located in some bands are relatively clean, while others are extremely noisy. It verifies that not all bands are contaminated by noise. Such a phenomenon is more evident in the typical no-

* Corresponding author.
  E-mail addresses: panerting@whu.edu.cn (E. Pan), mayong@whu.edu.cn (Y. Ma), meixiaoguang@gmail.com (X. Mei), fanfan@whu.edu.cn (F. Fan), jyma2010@gmail.com (J. Ma).
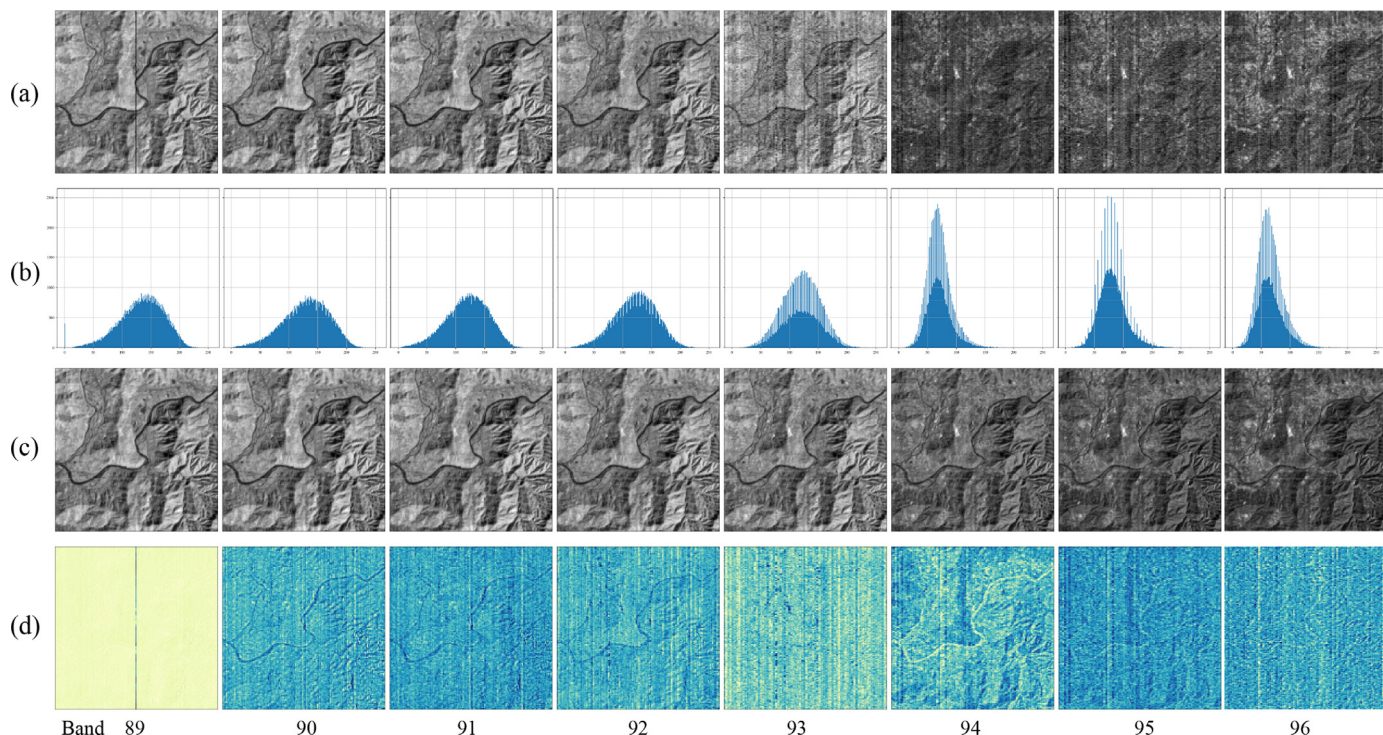
Fig. 1. (a) Images of ROIs from a interval of continuous bands in the original hyperspectral cube; (b) Histograms of ROIs in the selected bands; (c) Restored ROIs obtained by our proposed method; (d) Noise map in the selected bands estimated by the proposed method.
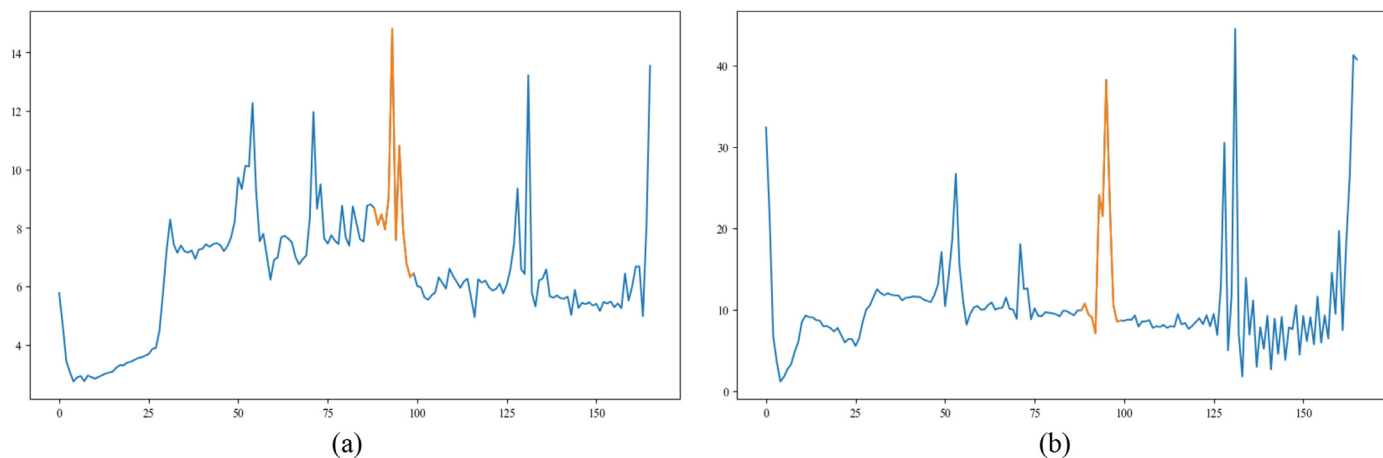

Fig. 2. No-reference image quality metrics on the EO-1 Hyperion dataset: (a) Noise level estimated according to literature [24]; (b) BRISQUE [25]. Note: the metrics of selected bands in Fig. r1 are marked orange. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

references metrics along the global spectral dimension in Fig. 2, where the metric curves present several spikes in a few bands. Accordingly, the real HSI noise distribution in the spectral dimension is generally non-uniform and has more complicated attributes. Treating all bands in an equal manner is not fair. Otherwise, bands with higher noise intensities will play a significant role in denoising algorithms, which may negatively impact those with lower noise intensities and eventually yields unsatisfying results. Unfortunately, previous state-of-the-art deep learning-based denoising methods scarcely concern all noise characteristics in the spectral dimension, too. It inclines to make their performance to deteriorate under more practical cases, which will be clearly observed in our experiments.

To overcome aforementioned challenges, this paper focuses on making full use of HSI noise characteristics and developing a novel HSI noise removal network based on spectral noise distribution bootstrap. Unlike previous denoising methods, we first seek an appropriate model to estimate the spectral noise distribution and capture its attributes, i.e., diversity, dense local correlation, and global sparsity. To this end, we improve the quasi recurrent neural network (QRNN) [26] by designing refined dense and sparse recurrent computations to model these attributes. Then, we integrate these calculators into a Sparse-QRNN and develop a transformation convolution block (Trans-ConvB) to further acquire the spectral noise distribution $\hat{N}$. As the estimated spectral noise distribution enjoys the property of describing the unique noise intensity of each band, we introduce it within a bootstrap convolution block (Bootstrap-ConvB) to provide guidance for subsequent spatial noise separation. Moreover, to acquire a more delicate denoising result, we design a bootstrap mechanism in a repetitive emphasis style. To verify the effectiveness of our method, we compare it with several state-of-the-art methods on data with both synthetic

and real HSI datasets. Overall, the proposed denoising framework conquers the drawbacks mentioned above of previous methods and solves the HSI denoising problem by the bootstrap mechanism of the noise distribution in the spectral dimension. Extensive experiments demonstrate the superiority of our method.

Our main contributions are summarized as follows:

- We construct a novel HSI noise removal framework based on spectral noise distribution bootstrap. Focusing on intrinsic attributes of HSI noise distribution along the spectral dimension enables the proposed method to be naturally generalized to HSIs in practical scenes and ameliorates their quality.
- We raise a specific branch consisting of Sparse-QRNN and Trans-ConvB to fully exploit the intrinsic attributes of HSI noise and progressively transform the attributes vector into the spectral noise distribution $\hat{N}$.
- We develop a bootstrap mechanism that repeatedly incorporates the estimated $\hat{N}$ via multiple Bootstrap-ConvBs to guide spatial denoising. The mechanism allows the spectral noise distribution to function in the spatial noise separation branch to the maximum extent. Such design significantly boosts the precise HSI denoising capability beyond existing techniques.

2. Related works

We have reviewed the mainstream HSI denoising methods and roughly classified them into the following two categories.

2.1. Model-driven based HSI denoising methods

Such methods typically define the denoising task as a maximum a posterior (MAP)-based optimization problem, whose performance relies heavily on the modeled prior. In past decades, scholars have developed a series of approaches concentrating on modeling the characteristics of HSI as the reasonable assumption or priors, involving typical properties such as nonlocal similarity [15,27,28], sparsity [29,30], low rank [23,31], and tensor structure [20,32]. For instance, the tensor dictionary learning (TDL) model [27] utilizes both the nonlocal similarity over spatial domain and the global correlation across the spectral dimension of HSIs and models the HSI as a low-rank tensor in a Laplacian scale mixture. LRMR [31] adopts an efficient HSI restoration method based on low-rank matrix restoration. Later, in Chang et al. [20], Chang et al. additionally consider the non-local self-similarity and present a one-way low-rank tensor recovery method. NMoG [33] defines a non-independent identically distributed (non-i.i.d.) mixture of the Gaussian to model the complicated HSI noise. More recently, some works [29,30,34] adopt band selection strategy to address the noisiest bands, but they ignore other bands with light noise, which results in information loss.

Some advanced model-driven methods have demonstrated high denoising performance, but they are still limited by the following two drawbacks: (1) Most model-driven methods involve time-consuming optimization processes and require sacrificing computational efficiency to achieve high performance; (2) They usually train a specific model with manually optimized parameters to eliminate noise of fixed intensity or type, resulting in difficulties tackling complex and diverse HSI noise contamination.

2.2. Data-driven based HSI denoising methods

Deep learning theory provides a data-driven strategy to solve complex denoising problems. Unlike model-driven methods based on domain-specific knowledge, deep learning-based methods directly learn a nonlinear end-to-end mapping between noisy HSIs and clean HSIs, relying on feature representations learned from large amounts of data. Aiming at natural image denoising, several works have been developed such as DnCNN [35], CBDNet [36], etc. In particular, the DnCNN proposed by Zhang et al. [35] employed residual learning and batch normalization (BN) for fast convergence. Inspired by these denoisers, which developed for natural 2D images, a range of HSI denoising methods have been proposed. For instance, HSI-DeNet [16] first introduces deep convolutional neural networks with a series of multichannel 2D filters in HSI denoising. Similarly, HSI-SDeCNN [37] presents a denoising method using a single CNN. Zhang et al. [22] presented a new deep CNN by incorporating the spatial-spectral gradient information. Hereafter, literature [17] proposes the QRNN3D for deep denoising by introducing QRNN into the 3D U-net structure. Cai et al. [38] raises a coarse-to-fine sparse Transformer to embed sparsity of HSIs into deep learning for HSI reconstruction.

In addition, the noise distribution of HSIs in practical scenes is much more complicated, especially in the spectral dimension. However, previous learning-based methods have barely considered the non-uniform noised distribution across the spectrum. Most of HSIs yield noise with non-i.i.d. [33,39], which decreases the robustness and generalization performance of these methods. There is still great potential for further exploration and promotion in this field. Hence, denoising HSIs needs to focus on two key issues to combat the weaknesses of existing methods: (1) consider the distribution characteristics of HSI noise, especially non-i.i.d, (2) deal with complex noise in practical scenarios.

3. The proposed method

3.1. Motivation

Given the noisy observation $\mathcal{Y} \in \mathbb{R}^{B \times W \times E}$, where $W \times E$ indicates the spatial scale and B is the total number of bands, the HSI denoising task is usually defined as to derive a noise-free HSI $\mathcal{X} \in \mathbb{R}^{B \times W \times E}$ from the noisy observation $\mathcal{Y}$. A generalized noise degradation model can be represented as:

$$\mathcal{Y} = \mathcal{X} + \mathcal{N}, \tag{1}$$

where $\mathcal{N} \in \mathbb{R}^{B \times W \times E}$ indicates hyperspectral noises. Due to the complex imaging chain, hyperspectral noises are quite complex and variable, increasing the difficulty of HSI denoising. As previously analyzed, there are two main reasons why current methods cannot deal with complex HSI noises. On the one hand, existing methods seldom consider the effects of varying noise intensity degrees on noise separation, treating these bands equally. This will inevitably deteriorate the final denoising performance. On the other hand, previous methods also scarcely took into account the properties of the noise distribution in the spectral dimension. In particular, they rarely pay attention to their global sparsity, which also limits the denoising performance. In view of these problems, this paper concentrates on investigating attributes of HSI noise distribution in the spectral dimension to promote HSI denoising in practical scenarios.

Before elaborating on our method, we first denote the HSI noise distribution in the spectral dimension as $\hat{N} \in \mathbb{R}^{1 \times B}$ and provide corresponding mathematical formulas of the observed attributes as below:

$$\hat{N} = \Psi(\hat{N}_D, \hat{N}_S), \tag{2}$$

where $\hat{N}_D \in \mathbb{R}^{1 \times B}$ and $\hat{N}_S \in \mathbb{R}^{1 \times B}$ denote the dense dependency and the global sparsity of $\hat{N}$, respectively, and $\Psi(\cdot)$ indicates the function that how to estimate $\hat{N}$ according to $\hat{N}_D$ and $\hat{N}_S$. To be specific, $\hat{N}_D$ represents the attribute that hyperspectral noises in adjacent bands are high-correlated, which can be formulated as:

$$\hat{N}_D = \Phi_D(Y_i; \theta_D(Y_{i-1}, Y_i)|_{i=1,2,\dots,B}), \tag{3}$$

where i indicates the band index and Y_0 is initialized to be zero. operation $\theta_D(\cdot)$ calculates the noise correlation between adjacent
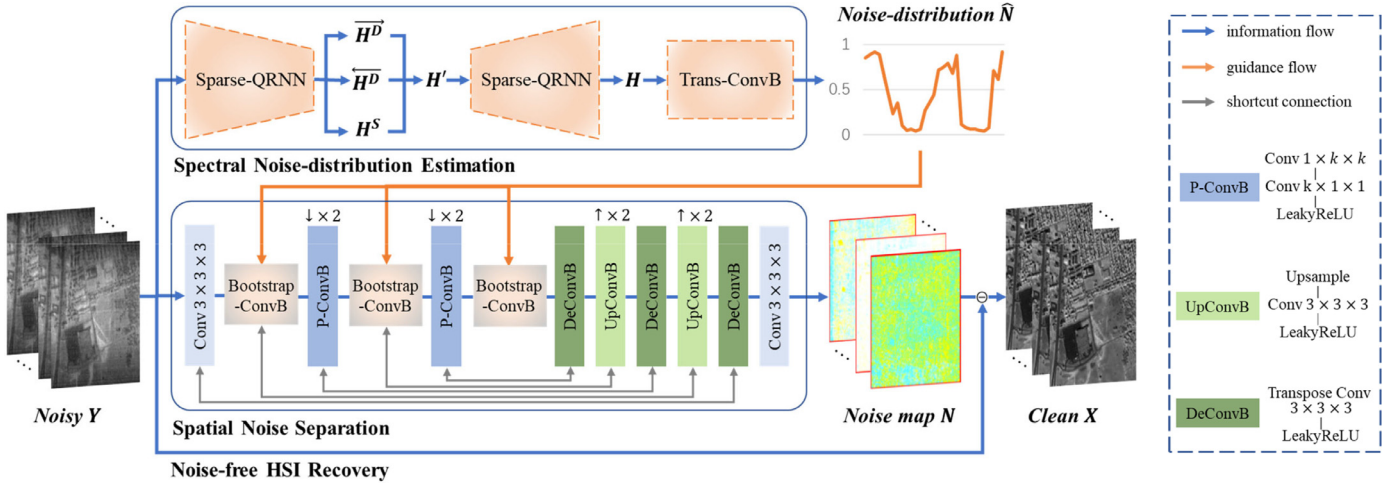
Fig. 3. Illustration of the overall architecture proposed denoising framework. It consists of three phases involving spectral noise distribution estimation, spatial noise separation and noise-free HSI recovery.

bands, after which the function $\Phi_D(\cdot)$ fuses these correlations with current bands along with the spectra and densely map them to $\hat{N}_D \in \mathbb{R}^{1 \times B}$. $\hat{N}_S$ represents the attribute that hyperspectral noises are sparsely distributed in the spectral dimension, which also can be formulated as:

$$\hat{N}_S = \Phi_S\left(P_j; \theta_S(P_{j-1}, P_j)\big|_{j=1,\ldots,\frac{B}{k}}\right), \tag{4}$$

where P_j denotes the jth bands-group, P_0 is initialized to be zero, and each group involves k neighboring bands. In other words, k indicates the sparsity. Correspondingly, $\theta_S(\cdot)$ calculates the correlation between adjacent bands-groups. $\Phi_S(\cdot)$ incorporates the calculated relation and the current bands-group, while iterating along the spectral dimension. In addition, $\Phi_S(\cdot)$ also plays the role of sparsely mapping them to $\hat{N}_S \in \mathbb{R}^{1 \times B}$.

In a nutshell, to finely recover the HSI noise-free observation X, images in different bands should be considered individually. Fortunately, the spectral noise distribution $\hat{N}$ can be roughly estimated by virtue of the last two attributes. The following noise separation can be exquisitely guided by assigning appropriate weights to each band based on $\hat{N}$. Next, we will provide detailed descriptions of the specific design for these phases.

3.2. Network architecture

As shown in Fig. 3, the designed overall denoising framework comprises of three phases, i.e., spectral noise distribution estimation, spatial noise separation, and noise-free HSI recovery. In concrete, in the first phase, the noisy input Y starts with encoding via the proposed Sparse-QRNN to fully extract the aforementioned intrinsic properties H of HSI noises. It follows by decoding the extracted properties with a symmetric Sparse-QRNN and ends with converting them into the spectral noise distribution $\hat{N}$ using the designed transformation convolution block (termed Trans-ConvB). After that, a Unet-based branch under a bootstrap mechanism of the estimated $\hat{N}$ is employed for accurate separating the spatial noise map N. Further, in the third phase, we earn a noise-free HSI X by incorporating the original noisy observation Y and the noise map N. Noteworthy, we adopt a unified learning strategy to train the network in an end-to-end manner, associating in all three phases. In other words, within the back-propagation of model parameters, the recovered result can serve as an auxiliary to the separated noise map and the estimated noise distribution, and the separated noise map can support the estimated noise distribution.

It is worth noting that the spectral noise distribution $\hat{N}$ plays two significant roles here. First, it makes the network stop factoring in the noise effects in each band equally. In contrast, it can guide the network to adapt varying noise intensities in different bands by introducing spectral noise distribution as additive noise biases. Second, it can regularize the network to alleviate overfitting and improve the generalization ability.
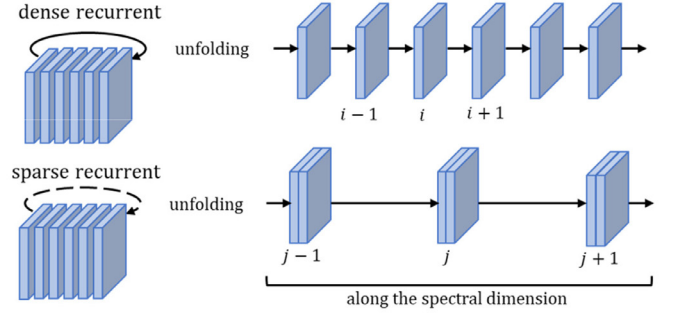
3.3. Spectral noise-distribution estimation

Motivated by Eqs. (3) and (4), we design Sparse-QRNN as the core model by capturing the correlations among the bands within the spectral dimension for the estimation of spectral noise distribution $\hat{N}$. Uniquely, it simultaneously integrates both dense and sparse recurrent computations, offering the merits of capturing both dense local correlation and global sparsity.

As demonstrated in Fig. 4(a), the proposed Sparse-QRNN contains a series of computations involving convolution, non-linear activation, and dense and sparse recurrent calculations. To be specific, given the input $X \in \mathbb{R}^{1 \times B \times W \times E}$, we first utilize a combination of multiple convolution kernels to extract spatial and spectral features $X' \in \mathbb{R}^{m \times B \times W \times E}$, paving the way for the subsequent correlation measurement in the spectral dimension. Notably, we stack one pseudo-3D convolution and one dilated pseudo-3D convolution (dilated by 2) in a group for feature extraction, using one group with the kernel size of $1 \times 3 \times 3$ for the spatial domain and another with the $3 \times 1 \times 1$ kernel for spectral dimension. Such a design can provide local features and relatively sparse features in both spatial and spectral domains. Next, we split the feature map X' in two and activate them non-linearly with $\tanh(\cdot)$ and $\sigma(\cdot)$, respectively, acquiring the representative tensor $T \in \mathbb{R}^{n \times B \times W \times E}$ and the selective gate $G \in \mathbb{R}^{n \times B \times W \times E}$. The selective gate G is characterized by $\sigma(\cdot)$ as a probability from 0 to 1 to execute the feature importance measurement for subsequent feature filtering and delivery.

Hereafter, instead of the original calculation of the quasi-recurrent f-pooling, we take bi-directional dense and sparse recurrent computations (see Fig. 4(b)) to capturing properties of the spectral noise distribution. Specifically, we first design a bi-directional dense recurrent calculation in a band-by-band style, which can effectively exploit the local dependency of features be-
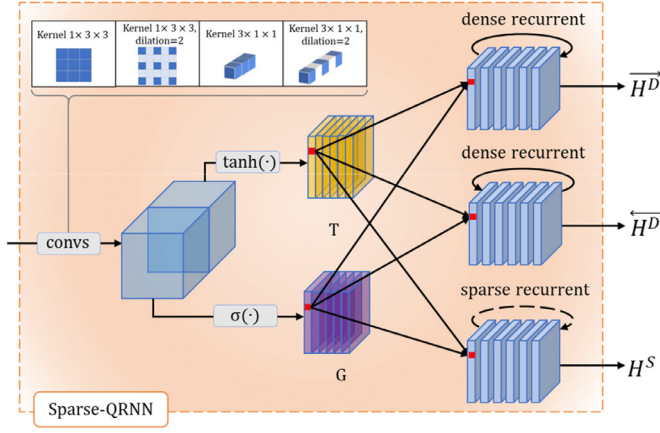
(a) The structure of the designed Sparse-QRNN.

(b) The unfolding information flow of dense recurrent calculation and sparse recurrent calculation.

Fig. 4. Illustrations of the proposed Sparse-QRNN, which exploring properties of the spectral noise-distribution via dense and sparse recurrent calculations along the spectral dimension.

tween adjacent bands, and $\Phi_D(\cdot)$ in Eq. (3) is defined as:

$$\begin{aligned} \overrightarrow{h_i^D} &= g_i \odot \overrightarrow{h_{i-1}^D} + (1 - g_i) \odot t_i, \\ \overleftarrow{h_i^D} &= g_i \odot \overleftarrow{h_{i+1}^D} + (1 - g_i) \odot t_i, \end{aligned} \tag{5}$$

where i ($i \in \{1, 2, \ldots, B\}$) represents the index of each band with dense intervals. t_i and g_i are derived from tensor T and G, respectively, and operator $\odot$ represents element-wise multiplication. Vectors $\overrightarrow{h_i^D}$ and $\overleftarrow{h_i^D}$ are feature correlations between adjacent bands, expecting to capture the forward and backward dense dependency. It continuously calculates feature correlations between adjacent bands and transfers them recurrently along the spectral dimension. For simplicity, $\overrightarrow{h_0^D}$ and $\overleftarrow{h_0^D}$ are initialized to zero. It is worth noting that g_i here filters the hidden state $\overrightarrow{h_{i-1}^D}$ transmitted from the previous band as well as the representative tensor t_i of the current band, which serves to aggregate features from the former and latter bands and enhance their correlation. Further, we propose a sparse transmission style across bands-groups, which is corresponding to $\Phi_S(\cdot)$ in Eq. (4), aiming to capturing the global sparsity of spectral noise distribution:

$$h_j^S = g_j \odot h_{j-1}^S + (1 - g_j) \odot t_j, \tag{6}$$

where t_j and g_j are derived by bounding the representative tensors and selective gates of the adjacent k bands into bands-groups. $k \in \{1, 2 \ldots, B\}$ denotes both the number of bands in each bands-group and the computational sparsity. Correspondingly, $j \in \{1, \ldots, B/k\}$ represents the index of bands-groups with sparse intervals. It captures global sparsity by recurrently computing feature correlations between bands-groups. If $k = 1$, it means that the sparse recurrent computation degenerates to a dense one. In particular, considering that the real world HSI data have a varying number of bands, we set $z = B$ for real HSI denoising, enabling the model trained on dataset with simulated noise to be transferred to other real-world HSI data. This setting also facilitates in capturing the global sparsity of the noise distribution. Subsequently, we note the results of the recurrent calculations as $\overrightarrow{H^D}, \overleftarrow{H^D}$, and H^S, and yield the final hidden state $H' \in \mathbb{R}^{C \times B \times W \times E}$ (C indicates the channel number of features) by the following calculation:

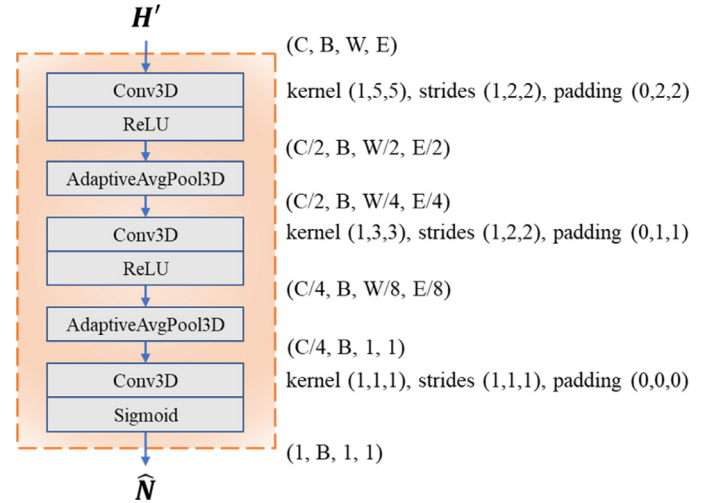$$H' = \overrightarrow{H^D} + \overleftarrow{H^D} + H^S. \tag{7}$$

Fig. 5. The composition of the designed Trans-ConvB. It estimates the spectral noise distribution $\hat{N}$ with progressively feature transformation.

On the basis of H' extracted by Sparse-QRNN, we leverage another Sparse-QRNN in symmetric to further decoding features of spectral noise distribution and acquire $H \in \mathbb{R}^{C \times B \times W \times E}$. Immediately afterward, we design a transformation convolution block (termed Trans-ConvB, illustrated in Fig. 5), involving a series of operations such as adaptive averaging pooling, convolution, ReLU and sigmoid activation. Trans-ConvB progressively maps features to the spectral noise distribution $\hat{N} \in \mathbb{R}^{B \times 1}$ via multi-level calculations and constraints the value ranging $(0,1)$ via sigmoid activation. Followed these, the final $\hat{N}$ is calculated by:

$$\begin{aligned} H &= \text{Sparse-QRNN}(H'), \\ \hat{N} &= \text{Trans-ConvB}(H). \end{aligned} \tag{8}$$

Overall, these designs enable the estimation of spectral noise distribution to effectively characterize its crucial attributes, especially the dense local correlations in adjacent bands and global sparsity in the spectral dimension.

3.4. Spatial noise separation

Upon acquiring the spectral noise distribution $\hat{N}$ in well-estimated, subsequent phases like spatial noise separation and noise-free HSI recovery can be fulfilled accurately. Nevertheless, conventional denoising networks ignore the diversity of noise intensities in the spectral dimension. They simply take these bands into an average consideration, which directly conflicts with the aforementioned attributes (i.e., sparsity) and always leads to poor denoising performance. To this end, we take the estimated spectral noise distribution $\hat{N}$ as knowledge prior to guide spatial noise separation. It is remarkable that, to yield a precise noise map, we design a novel bootstrap mechanism that repeatedly emphasize the spectral noise distribution $\hat{N}$ as guidance throughout the spatial noise separation process. The concrete design details are described below.

First, denoising is generally regarded as a low-level task in computer vision. Hence, we employ a relatively shallower network as the backbone of spatial noise separation (as illustrated in Fig. 3). In more detail, we employ 3D convolution with $3 \times 3 \times 3$ kernel as the head and tail layers of the U-net backbone for jointly exploring or restoring spatial and spectral features. The main body of this backbone contains two down-sampling layers in the encoding process and other two corresponding layers up-sampling in the decoding process. Besides, we also add some shortcut connections between layers in different levels, ensuring that effective features from distant layers are linked. The arrangements mentioned above facilitate the extraction of noise-related features in the spatial domain and the reconstruction of HSI noise maps.

Second, during the encoding process, we develop a bootstrap mechanism based on the guidance under estimated spectral noise distribution $\hat{N}$. It enables our model to treat bands that degrade in various degrees with different weights tailor-made according to $\hat{N}$. Noteworthy, the designed guidance flow can be seen by the lines with arrows colored orange in Fig. 3. The estimated spectral noise distribution $\hat{N}$ is friendly to the varying scales of feature maps f in different layers. Rather than guiding only once, the proposed bootstrap mechanism integrates $\hat{N}$ repeatedly across layers of different resolutions, ensuring that each band is assigned customized weight according to $\hat{N}$ during the feature encoding and transmission process. The merit of such mechanism lies in that feature map from each layer can be characterized and updated on the basis of the estimated spectral noise distribution $\hat{N}$, benefiting in accurate noise separation.

Third, we form the backbone of this branch with four functional blocks, including the pseudo-3D convolution block (P-ConvB), the bootstrap convolution block (Bootstrap-ConvB), the upsampling convolution block (UpConvB), and the deconvolution block (De-ConvB). Taking P-ConvB as an example, it consists of a spatial convolution layer with the kernel size of $1 \times k \times k$, followed by a spectral convolution layer with the kernel size of $k \times 1 \times 1$ and a LeakyReLU activation, where k is generally set to 3. The detailed configurations of these blocks are shown in the dashed box in Fig. 3. Remarkably, unlike the other three, Bootstrap-ConvB incorporates residual learning with multiple convolutions to maximize the efficiency of $\hat{N}$ guidance and feature extraction (see Fig. 6). We devise a specialized calculation that can be formulated as follows:

$$f' = \text{Bootstrap-ConvB}(f * \text{Conv}(\hat{N}; s)) + f, \tag{9}$$

where Bootstrap-ConvB$(\cdot)$ indicates the combination of P-ConvBs in multiple size kernels, 3D convolution layers in multiple size kernels, and LeakyReLU activation. It serves to perform affine transformation upon the band-wise level and regularization upon the channel-wise level according to the spectral noise distribution $\hat{N}$. Conv$(\hat{N}; s)$ acts as an additive noise bias tuner to simulate the
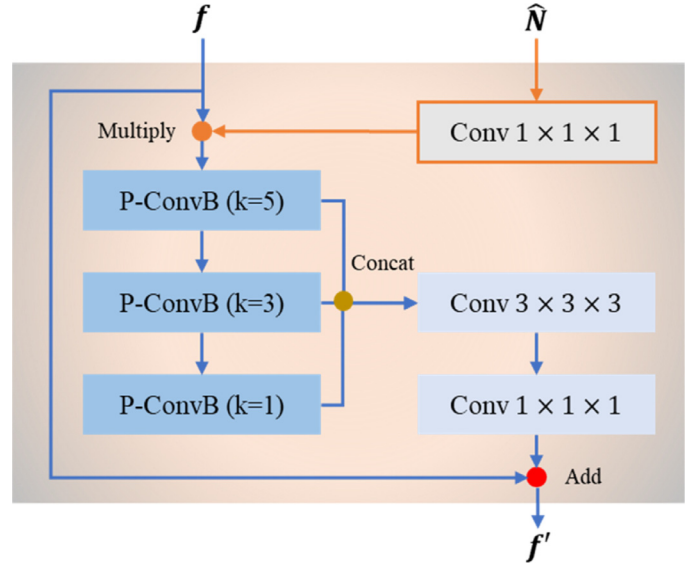

Fig. 6. Illustration of the design of Bootstrap-ConvB. It utilizes the estimated $\hat{N}$ to bootstrap a more accurate spatial noise separation.

varying degrees of bias in different bands. It can model the non-linear biases well, which is profitable to capture the actual noise distribution of the spectral dimension for noise separation. Under the guidance of spectral noise distribution throughout the backbone network, this phase can effectively achieve accurate HSI noise map separation.

4. Experiments and evaluation

4.1. Experimental settings

4.1.1. Data with synthetic noise

The publicly available ICVL[1] hyperspectral dataset [40] is collected by a Specim PS Kappa DX4 hyperspectral camera and a rotary stage for spatial scanning. Each HSI cube in ICVL has a spatial resolution of 1392×1300 and a spectral dimension of 31. The bands are formed in wavelength increments of 10 nm and cover a visible range from 400 nm to 700 nm. Since the ICVL dataset provides a sufficient number of practically noise-free HSIs and widely used as benchmark dataset in Wei et al. [17], Pan et al. [41], we conduct experiments with ablation studies on this dataset. However, HSIs collected from real scenes are always contaminated by complex noises. To mimic noise-contaminated HSIs in the real scenes to the greatest extent, we add a series of complex synthetic noises to the clean ICVL HSIs. Concretely, it involves five cases of complex noise: (1) Non-i.i.d. Gaussian noise; (2) Non-i.i.d. Gaussian + Stripe noise; (3) Non-i.i.d. Gaussian + Dead-line noise; (4) Non-i.i.d. Gaussian + Impulse noise; (5) Mixture noise. Specifically, among these cases, two-thirds of bands are polluted with non-i.i.d Gaussian noise, which is zero-mean and has a random intensity ranging from 30 to 70. Stripe, dead-line noises are randomly added to one-third of bands with 5% to 15% of columns, while impulse noise is randomly added to one-third of bands with an intensity ranging from 10% to 70%. Mixture noise means that the HSI is not only corrupted by Gaussian noise but also by some sparse noise mentioned before. We also conduct synthetic mixture noises ex-

---

periments on the publicly available HYDICE image of Washington D.C. Mall,[2] which has a resolution of $1280 \times 307 \times 191$.

4.1.2. Implementation details

Our model is implemented on the PyTorch[3] platform and accelerated on NVIDIA TITAN RTX GPU. To train our $\hat{N}$-Net model, we select 100 out of 200 HSIs from the ICVL dataset as the training set, along with 5 for validation, and the rest of these HSIs constitute the testing set. Moreover, we set the patch with a spatial size of 64×64 and preserve the complete spectral dimension. We set the mean absolute error loss (MAEloss) as the loss function. Hence, the network is trained by minimizing the MAEloss between the noisy HSI and the clean ground truth pairs. Besides, to improve the robustness of training, we rescale the patches with a rate in $\{1, 0.5, 0.25\}$ and randomly rotate them to diversify the training set further, yielding a total of 40k training samples. Parameters are initialized with Kaiming Initialization and updated by Adam Optimizer. To obtain a denoising model which robust to various types of hyperspectral noise, we follow the easy-to-difficult training policy and set up a two-phase training with a total of 60 epochs with a fixed batch size (i.e., 16). In addition, the initial learning rate is set to 0.001 and decreases exponentially with epochs, where the validation results show stable performance.

4.1.3. Evaluation metrics

We employ four mainstream quantitative metrics on the denoised results to measure the denoising performance, inclusive of mean peak signal-to-noise ratio (MPSNR), mean structure similarity (MSSIM), mean feature similarity (MFSIM), as well as mean spectral angle mapper (MSAM), mean relative dimensionless global error synthesis (MERGAS) is used to calculate the accuracy of Pan sharpened image considering normalized average error of each band of the result image. The MPSNR, MSSIM, and MFSIM are used for spatial-based image quality evaluation, and a larger value represents better denoising performance. In contrast, MSAM and MER-GAS are spectral-similarity-based metrics, and a smaller MSAM and MERGAS value implies better spectral fidelity for spectral dimension. Additionally, we also record the average running time of each comparative method under all noisy cases. Noteworthy, for deep learning-based methods, the running time refers to the average time consumed per HSI during the testing phase.

4.1.4. Comparative methods

We compare our method with state-of-the-art HSI denoising solutions, including five traditional model-driven based methods, i.e., LRMR [31], LRTV [23], LRTDTV [32], NMoG [33], and FastHyMix [42], and four advanced data-driven based method i.e., QRNN3D [17], SQAD [41], MAC-Net [43], and GRUNet [44]. For these comparative methods, we have made great efforts to reproduce the best denoising results with the publically available codes and optimal parameters settings.

4.2. Denoising on HSIs with synthetic noise

4.2.1. Estimated spectral noise distribution

As represented in Fig. 7, we show noisy HSIs in the selected synthetic noise cases and their corresponding spectral noise distribution $\hat{N}$, which is estimated by our $\hat{N}$-Net. It is observed that values of $\hat{N}$ close to 0 refer to bands heavily contaminated by noise, and vice versa, values close to 1 denote relatively cleaner bands. Intuitively, the curve of the estimated $\hat{N}$ follows a high consistency with the noisy bands along the spectral dimension (as marked by

the red dashed arrows). In addition, the estimated $\hat{N}$ describes the unique properties in the spectral dimension with certainty, providing a reliable prior for dealing with different degradations of noise corruption and can be extremely helpful for subsequent denoising. From another aspect, our $\hat{N}$-Net acquires diverse $\hat{N}$ curves for different noisy cases, which reveals that our proposed model has a strong generalization capability.

4.2.2. Spatial quality comparisons

The visual results of some representative scenes are shown in Fig. 8. We zoom in on two regions of interest in the reconstructed images for a more straightforward and precise observation. Obviously, traditional model-driven methods LRMR and LRTV fail to these complex noises. LRTDTV and NMoG have successfully removed most of the noise at the cost of missing textural details. From the detail aspect, some fine details are blurred in most results, and they are of relatively low quality in comparison with ours. It is evident, especially in case (4), that only NMoG and our $\hat{N}$-Net finely recover the critical lines in such a scene, and our method shows a cleaner result with more contextual details. Grossly, our results achieve more spatial fidelity under all these noisy cases and are much clearer with fewer artifacts and sharp edges. It also implies that our method has a more robust capability of removing various complex noises.

4.2.3. Spectral fidelity comparisons

As shown in Fig. 9, we select two pixels in each noisy case and draw the spectral curves recovered by all comparative methods to verify the superiority of our method in spectral fidelity. Evidently, the spectra recovered by some traditional methods (i.e., LRMR, LRTV, LRTDTV, NMoG, FastHyMix) still show abnormal spikes and fluctuations, especially LRTV in case (3) and (5), as well as NMoG in case (4). Such results represent their unstable performance, suggesting that they fail to maintain the spectral fidelity and cannot denoise HSIs completely. Compared to the other results, the recovered spectra curves of our method are much closer to the reference spectra. It indicates that our method $\hat{N}$-Net accurately eliminates the negative effect of noise in the spectral domain and further confirms the advantage of our method in maintaining high spectral fidelity.

4.2.4. Quantitative comparisons

The averaged denoising results overall test HSss for quantitative evaluation are listed in Table 1. The best results are highlighted in red for each metric among all comparative results. Firstly, an easy observation is that the proposed $\hat{N}$-Net dramatically outperforms other advanced denoising methods under all complex noise cases. Compared to before denoising, the proposed $\hat{N}$-Net brings an average gain of the MPSNR indicator exceeding 21 dB under all noisy cases. This indicator gains even up to 25.797 dB in the case of mixed noise. As for other 2D image quality evaluation metrics (MSSIM and MFSIM), our proposed method also obtains outcomes with absolute advantages, which also corroborates with the previous visualization results (Fig. 8). For a more comprehensive and objective evaluation of the denoising performance of each model, we have collected the spatial quality evaluation indexes (MPSNR and MSSIM) of each band, yielding Fig. 10. Remarkably, our method surpasses other methods in the majority of bands under all cases, demonstrating our method's effectiveness and robustness. On the other hand, the numerical results of the MSAM metric and the MEGRAS confirm the superiority of the proposed $\hat{N}$-Net on spectral fidelity again. To sum up, the quantitative comparison results indicate that the proposed method has superior flexibility in complex noise cases. In the spectral dimension, the stable performance of the spatial quality evaluation indexes (MPSNR, MSSIM) in each band also indicates that our method can make full use of the HSI
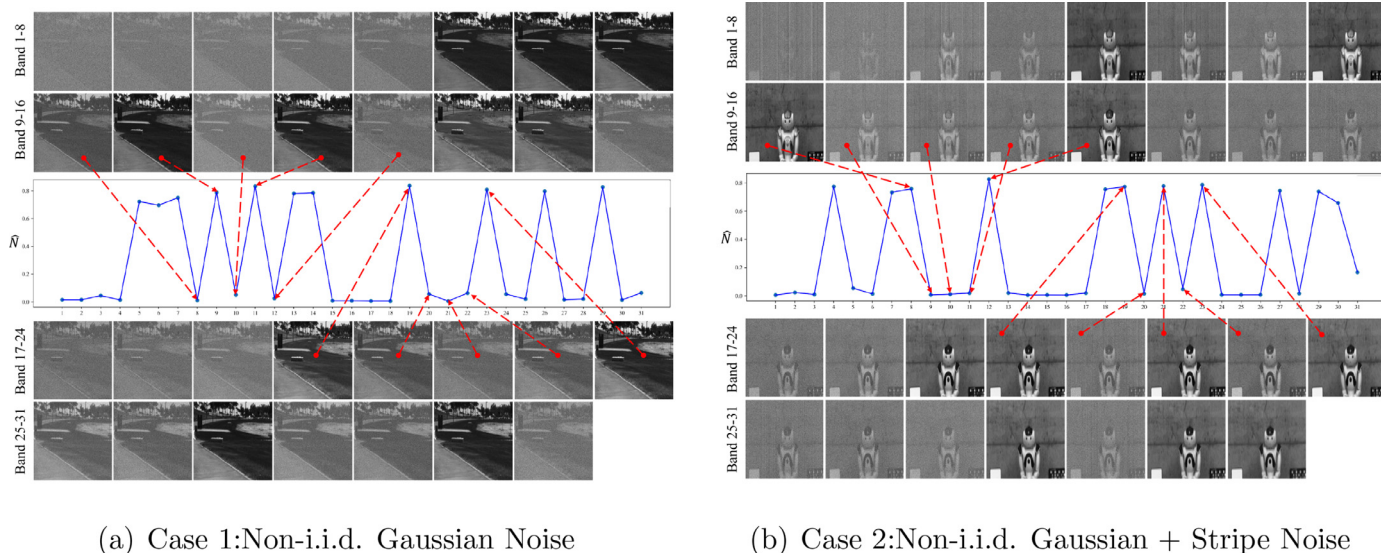
---

[2] It is available at https://engineering.purdue.edu/~biehl/MultiSpec/hyperspectral.html.

[3] http://pytorch.org/

(a) Case 1:Non-i.i.d. Gaussian Noise   (b) Case 2:Non-i.i.d. Gaussian + Stripe Noise

Fig. 7. Noisy HSIs in the synthetic noise cases and their corresponding spectral noise distribution $\hat{N}$, which estimated by $\hat{N}$-Net. The red dashed arrows partially mark their correspondence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
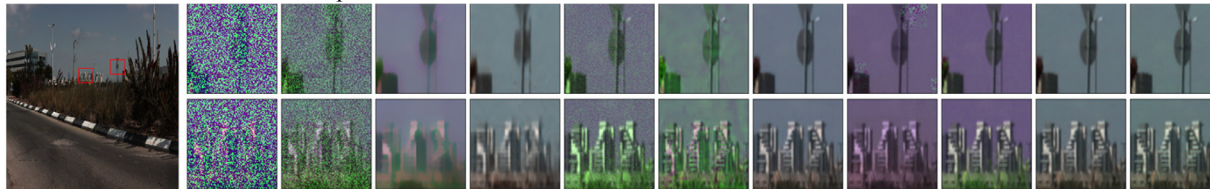

Case 1: Non i.i.d Gaussian Noise
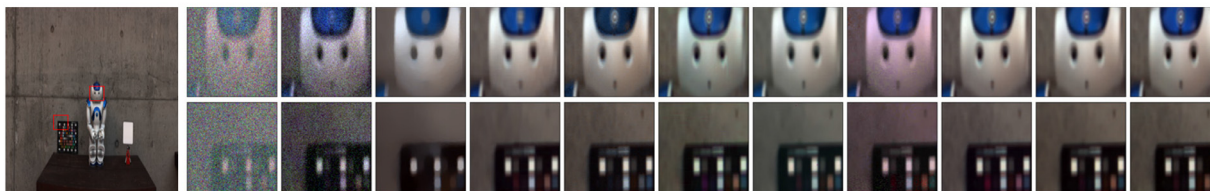
Case 2: Non i.i.d Gaussian + Stripe Noise
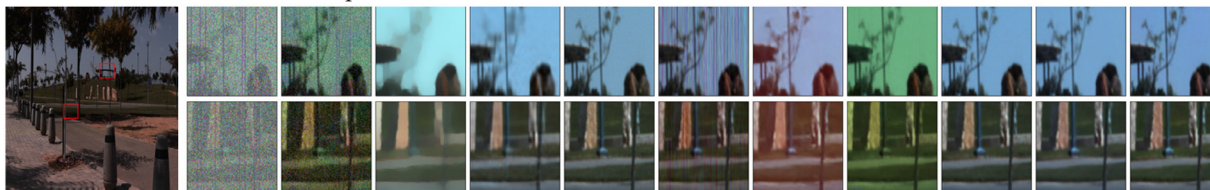
Case 3: Non i.i.d Gaussian + Deadline Noise
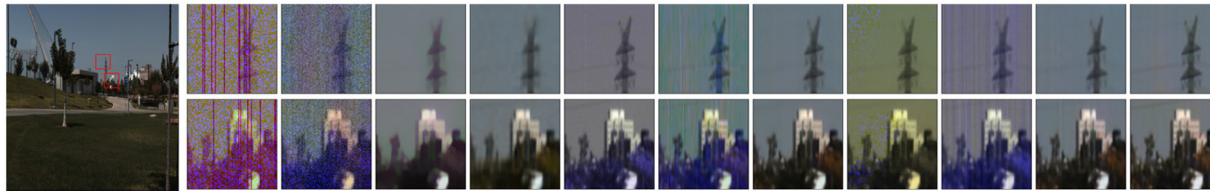
Case 4: Non i.i.d Gaussian + Impulse Noise

Case 5: Mixture Noise

Noisy   LRMR   LRTV   LRTDTV   NMoG   FastHyMix   QRNN3D   SQAD   MAC-Net   GRUNet $\hat{N}$-Net

Fig. 8. Denoising results on ICVL dataset with synthetic complex noise. Examples for all the competing methods on band 5,12 and 20 of the ICVL dataset under 5 complex noise cases are presented respectively.
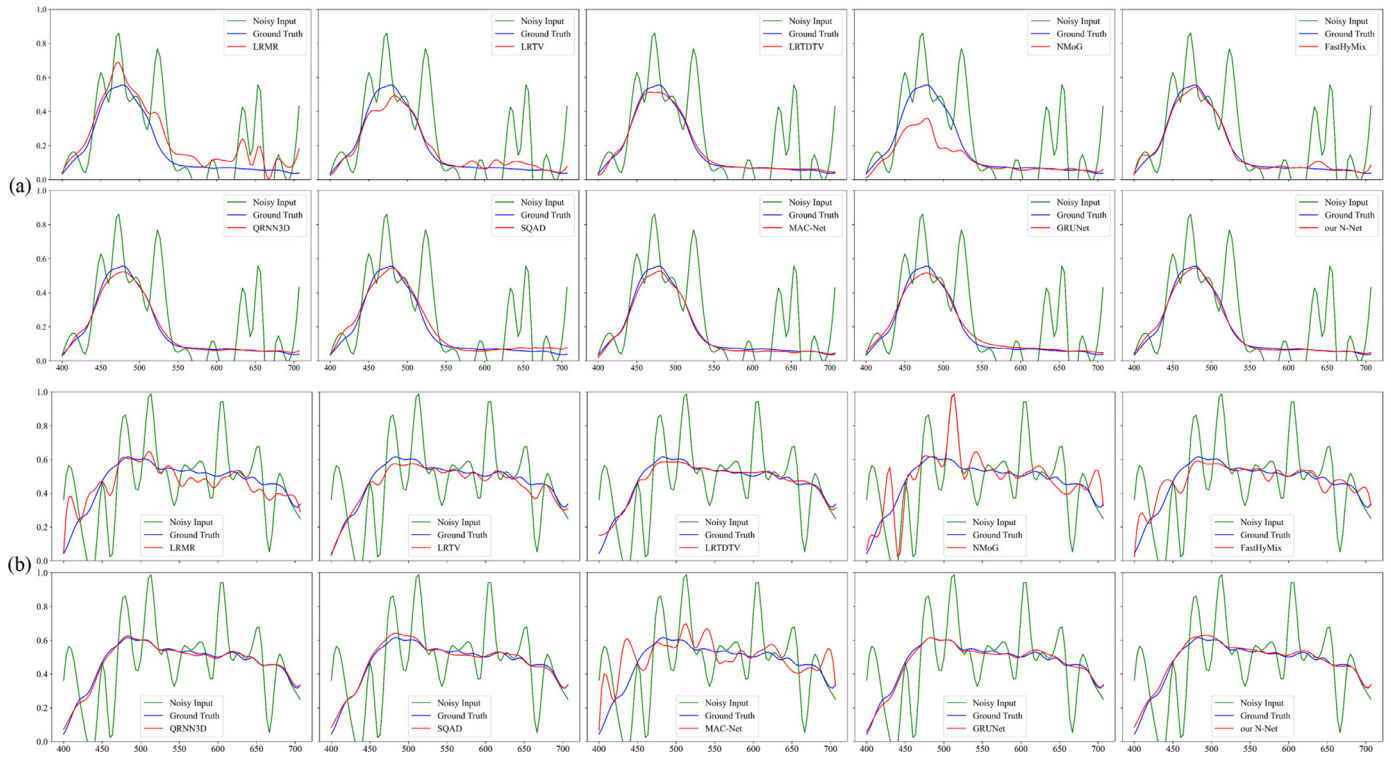
Fig. 9. The spectra of two selected pixels recovered by some advanced comparative methods under all noisy cases from the ICVL dataset are presented, respectively. From the left to right, they are (1) Non-i.i.d. Gaussian noise case; (2) Mixture noise case.

Table 1
Quantitative results of all competing methods under all complex synthetic noise cases on the ICVL dataset. The best results are shown in red, the second best ones are shown in blue, and the third ones are shown in green.

|	Noisy	LRMR [31]	LRTV [23]	LRTDTV [32]	NMoG [33]	FasyHyMix [42]	QRNN3D [17]	SQAD [41]	MAC-Net [43]	GRUNet [44]	our $\hat{N}$-Net
Case (1): Non-i.i.d. Gaussian noise											
MPSNR	23.113	29.142	35.621	39.623	41.063	44.082	44.008	43.354	45.351	44.682	45.56
MSSIM	0.3558	0.5393	0.9171	0.9522	0.9601	0.9788	0.9809	0.982	0.986	0.9844	0.9861
MFSIM	0.6718	0.9106	0.9384	0.9797	0.9907	0.9923	0.9928	0.9948	0.9959	0.9951	0.9963
MSAM	0.7637	0.3287	0.0677	0.0616	0.0927	0.066	0.0388	0.0473	0.0387	0.0362	0.0378
MERGAS	48.828	11.073	5.187	3.706	3.755	2.176	1.709	1.967	1.537	1.608	1.483
Case (2): Non-i.i.d. Gaussian + Stripe noise											
MPSNR	23.082	29.083	35.609	39.294	40.275	41.293	43.93	43.309	45.217	44.568	45.46
MSSIM	0.3553	0.5404	0.9176	0.9513	0.9474	0.9357	0.9809	0.982	0.9848	0.9844	0.9859
MFSIM	0.6672	0.9088	0.9391	0.9792	0.9843	0.9788	0.9927	0.9948	0.9959	0.995	0.9963
MSAM	0.7633	0.3311	0.0681	0.0622	0.1344	0.103	0.0396	0.0486	0.0402	0.0373	0.0388
MERGAS	49.393	11.168	5.201	3.776	5.589	3.582	1.716	1.961	1.591	1.635	1.5
Case (3): Non-i.i.d. Gaussian + Deadline noise											
MPSNR	23.007	28.565	34.115	36.765	38.028	38.525	39.323	42.299	41.815	43.312	43.945
MSSIM	0.3541	0.539	0.9062	0.9333	0.9452	0.8608	0.9547	0.9808	0.9834	0.9841	0.985
MFSIM	0.6651	0.9046	0.937	0.9717	0.9825	0.9476	0.9797	0.9944	0.9964	0.9947	0.996
MSAM	0.7776	0.3383	0.1096	0.0881	0.1603	0.1255	0.0545	0.0553	0.0692	0.0376	0.0471
MERGAS	50.057	11.794	7.979	5.257	6.923	6.938	2.852	2.399	3.525	1.742	1.982
Case (4): Non-i.i.d. Gaussian + Impulse noise											
MPSNR	21.211	24.816	34.155	38.464	33.797	36.578	41.879	40.372	37.23	42.118	42.306
MSSIM	0.3425	0.4066	0.8913	0.9479	0.8228	0.8835	0.9581	0.9545	0.8856	0.962	0.9669
MFSIM	0.6205	0.8332	0.9336	0.9775	0.9335	0.9739	0.993	0.9917	0.9702	0.991	0.9913
MSAM	0.8157	0.4703	0.2118	0.0749	0.4339	0.4043	0.1171	0.1243	0.3761	0.055	0.0761
MERGAS	51.915	20.85	7.41	4.26	17.638	15.632	3.5	3.819	14.98	2.775	2.449
Case (5): Mixture Noise											
MPSNR	15.474	23.906	32.759	36.08	30.296	29.597	40.526	38.959	30.665	39.48	40.807
MSSIM	0.1852	0.3928	0.8832	0.9264	0.7795	0.745	0.9559	0.9476	0.8658	0.9538	0.9604
MFSIM	0.5325	0.826	0.9284	0.9699	0.9201	0.9222	0.9898	0.9862	0.9592	0.9851	0.9877
MSAM	0.8497	0.4775	0.1983	0.0936	0.4364	0.4216	0.1178	0.1215	0.3759	0.066	0.076
MERGAS	57.512	20.944	8.621	5.585	14.902	16.323	3.494	3.769	14.251	2.892	2.677
Time(s)	-	12.42	201.4	411.73	207.92	2.35	0.84	0.78	3.21	0.59	0.56
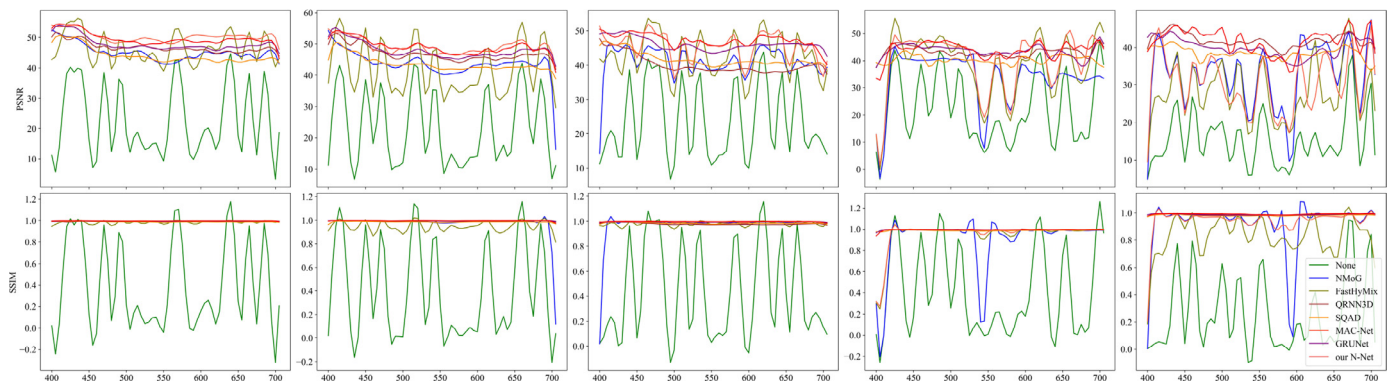
Fig. 10. PSNR and SSIM metrics across the spectrum of some advanced comparative methods under all cases. From the left to right, they are (1) Non-i.i.d. Gaussian noise; (2) Non-i.i.d. Gaussian + Stripe noise; (3) Non-i.i.d. Gaussian + Dead-line noise; (4) Non-i.i.d. Gaussian + Impulse noise; (5) Mixture noise.
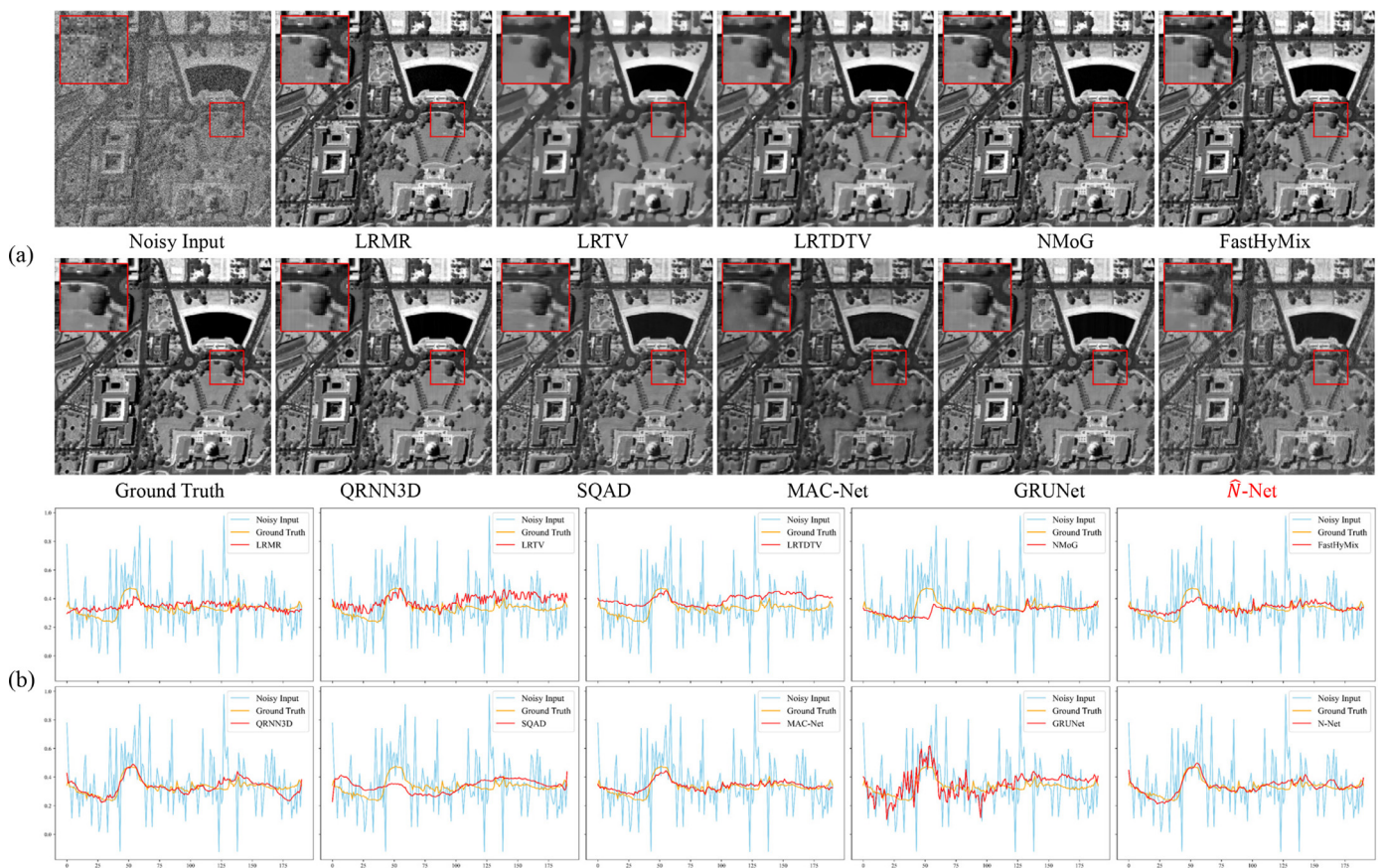

Fig. 11. Comparison in denoising results under complex synthetic mixture noise case on the Washington D.C. Mall dataset. (a) Denoising results of ROIs on band 115; (b) The spectra of the selected pixel recovered by comparative methods.

intrinsic attributes to guide the HSI denoising and recovery, which further ensures the merit of our $\hat{N}$-Net model.

4.3. Denoising on remote sensing HSIs with synthetic noise

The experimental results shown in Fig. 11 and Table 2 demonstrate the effectiveness and robust denoising performance of the proposed $\hat{N}$-Net in handling complex noise degradation in remote sensing scenes.

Fig. 11 shows denoising results of ROIs on band 115 in the Washington D.C. Mall dataset. It can be seen that the proposed $\hat{N}$-Net restores the cleanest denoised image and the most accurate spectral recovery compared to the other methods. Table 2 further quantitatively evaluates the performance of the proposed method

and the other competing methods. FastHyMix achieves the best in MPSNR, MERGAS and Times, while NMoG gets the best on MSSIM and MFSIM. The proposed $\hat{N}$-Net achieves the top three values of all metrics, indicating its robust denoising performance in remote sensing scenes. On the other hand, the performance of the other methods, neither traditional methods LRMR, LRTV, LRTDTV, nor deep learning-based methods QRNN3D, SQAD, MAC-Net, and GRUNet, is relatively poor, with lower MPSNR, MSSIM, and MF-SIM values. One possible reason is that with regard to all deep-learning-based methods, we directly adopted the trained model (all of these methods are trained on ICVL dataset) to test on this new dataset.

Overall, the experimental results demonstrate that the proposed method outperforms the other competing methods in handling

Table 2

Quantitative results of all competing methods under complex synthetic mixture noise case on Washington D.C. Mall dataset. The best results are shown in red, the second best ones are shown in blue, and the third ones are shown in green.

|	Noisy	LRMR [31]	LRTV [23]	LRTDTV [32]	NMoG [33]	FasyHyMix [42]	QRNN3D [17]	SQAD [41]	MAC-Net [43]	GRUNet [44]	our $\hat{N}$-Net
MPSNR	18.71	31.202	26.642	29.288	34.516	34.793	28.288	22.772	34.111	26.105	34.52
MSSIM	0.6138	0.9562	0.8713	0.942	0.9818	0.9783	0.9499	0.8437	0.9744	0.8949	0.9767
MFSIM	0.781	0.9664	0.8865	0.9525	0.9855	0.9825	0.9511	0.8871	0.9791	0.9257	0.9811
MSAM	0.488	0.1019	0.1145	0.0823	0.0742	0.0720	0.1132	0.1726	0.0761	0.1567	0.0702
MERGAS	27.772	4.831	8.61	5.801	4.025	3.313	6.594	12.345	3.603	8.565	3.43
Time (s)	-	38.71	152.15	459.78	143.63	0.44	5.45	6.01	14.13	7.02	4.42
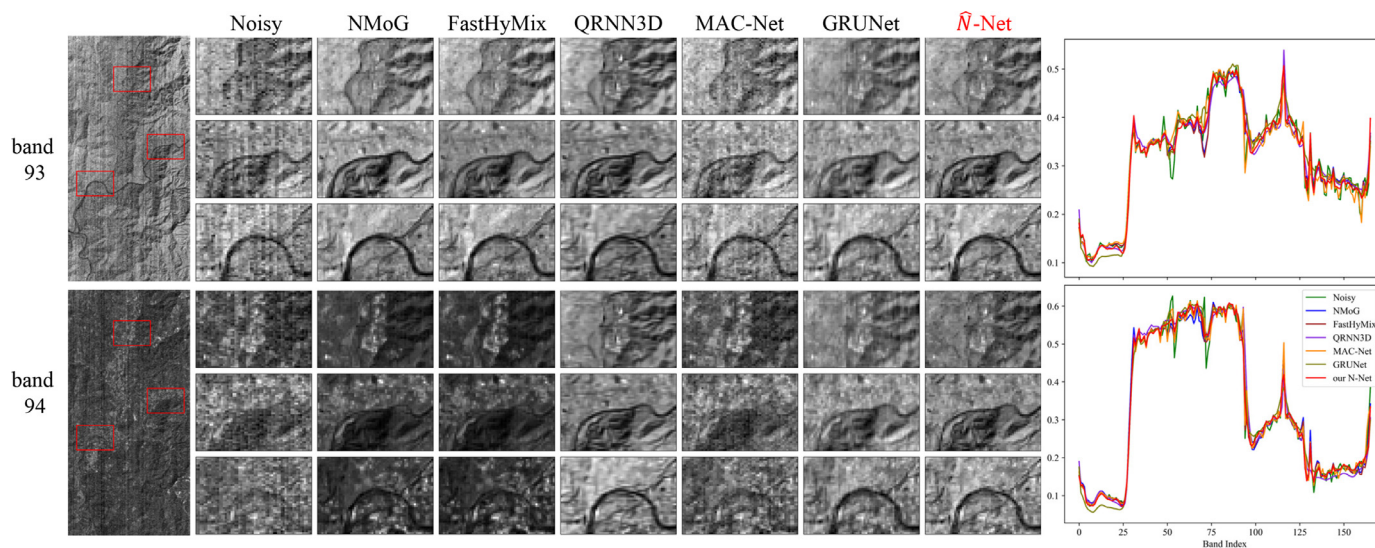
Fig. 12. Comparison in denoising results on EO-1 Hyperion HSI dataset with selected bands and comparison in recovered spectra of two sampled pixels.

complex noise degradations in remote sensing scenes. The generalization capability of the proposed $\hat{N}$-Net to real-world HSIs and its lightweight scale are additional benefits. These results provide valuable insights for advancing the development of effective and robust HSI denoising methods.

4.4. Denoising on real HSI

4.4.1. Real HSI data

We also demonstrate the ability of our proposed method to generalize to real hyperspectral noise without corresponding ground truth. We directly evaluate the trained model in remote sensing hyperspectral datasets collected in natural scenes to verify its denoising performance and robustness. Three experiments are carried out on the EO-1 Hyperion[4] dataset with 166 bands, the Gaofen-5 (GF-5) Shanghai dataset[4] and the GF-5 Baoqing dataset[4] with 155 bands acquired by the Advanced Hyperspectral Imager in GF-5 satellite [45]. Although these data are negatively affected by the atmosphere and water, which heavily contaminated some bands, we still conduct denoising without removing any bands.

We directly utilize the model trained on synthetic noise data from the ICVL dataset to test the real remote sensing datasets to verify its denoising performance and generalization ability. Fig. 12 represents results on EO-1 Hyperion dataset, and Fig. 13 illustrate denoising result on GF-5 Baoqing dataset and GF-5 Shanghai dataset. Due to environmental factors such as atmosphere, water, or hardware factors like photon shot within the sensor, these data sets are heavily contaminated with complex noise. We can easily observe complex noise in some bands involving Gaussian, stripe, deadline, and mixture noise, resulting in a big challenge

to recover clean HSIs. Aiming at a fine comparison of the generalization capabilities of the comparative algorithms, we select two bands suffering from different levels of noise contamination as examples and illustrates the recovered details by zooming in on three small regions. Besides, we also choose two pixels in each real remote sensing dataset and draw the spectral curves recovered by some leading methods to confirm the spectral fidelity of our method.

In these datasets, the original HSIs are damaged by complex mixtures of the Gaussian noise, the stripe noise, and the deadline noise. It is apparent that our method achieves better results in visualization due to its strong capability of self-guiding, especially in terms of better recovery of texture and edge details as well as a more delicate removal of structural noise. Moreover, our method still performs stably in dealing with different contamination levels in various wavebands. Such results further confirm the robustness of our proposed method in a real-world scenario. On the other hand, corrupted by noise, the original spectral curves show some localized dense fluctuations caused by small perturbations and a few spikes caused by significant variations. Generally speaking, it is pretty clear that the spectral curves are smoother after denoising and recovery by our proposed method. Besides, the overall shape of the spectral curve recovered by our method is more consistent with the original data, which verifies the effectiveness and flexibility of our model.

On the other hand, as Fig. 14 shown, the estimated noise distribution curve of these two datasets are are relatively consistent with the actual noise degradation degree, indicating the generalization capability of the trained model to real-world HSIs. Especially the Fig. 14(a), albeit different, the spectral noise distribution it depicted has a certain similarity to typical no-reference image quality metrics in Fig. 2. It also implies the potential positive guid-

---

[4] It is available at http://hipag.whu.edu.cn/resourcesdownload.html.
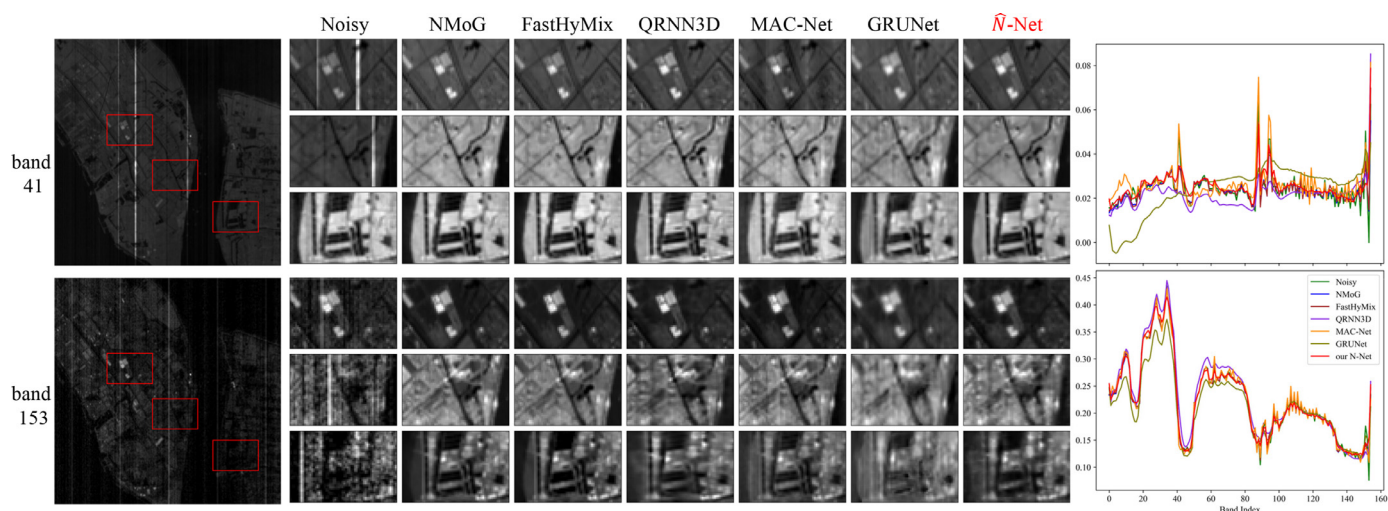
(a) GF-5 Shanghai dataset
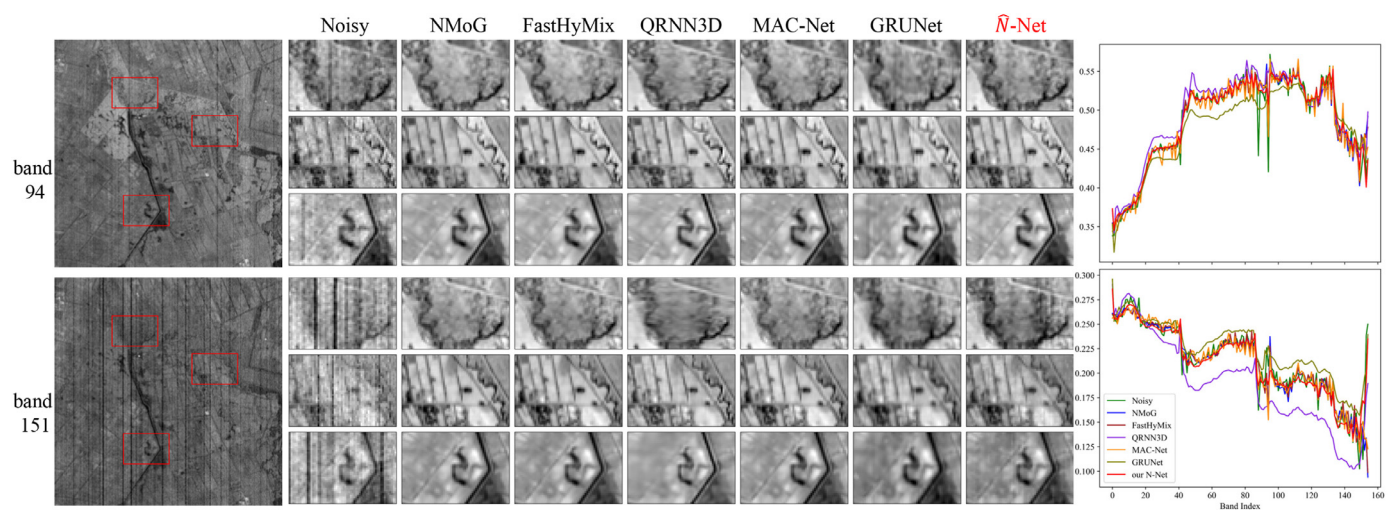
(b) GF-5 Baoqing dataset

Fig. 13. Comparison in denoising results on GF-5 HSI datasets with selected bands and comparison in recovered spectra of two sampled pixels.

ance of the estimated spectral noise distribution for robust denoising performance.

4.5. Ablation study and analysis

Take non-i.i.d. Gaussian noise case and non-i.i.d. Gaussian + stripe noise case as examples, Fig. 15 shows some typical denoising results and corresponding noise maps separated from noisy bands, and Fig. 16 presents the denoised spectra of two sample pixels. Obviously, each band is contaminated with different degrees of noise, and the proposed model effectively removes all the noise and finely reconstructs the clean image. For one thing, the noise map shown in Fig. 15 suggests that our model successfully removes the synthetic non-i.i.d. Gaussian noise and stripe noise in different degrees. For another, the denoised spectra shown in Fig. 16 imply the high spectral fidelity of our model. These results demonstrate the effectiveness of the proposed $\hat{N}$-Net model in coping with different types of complex hyperspectral noise.

To investigate the contribution of each specific design in our method downright, we execute comprehensive ablation studies on

synthetic non-i.i.d. Gaussian noise removal on the ICVL dataset. We take MPSNR, MSSIM, MSAM to evaluate the model performance, and take Total Params and FLOPs to evaluate the model complexity. Table 3 lists the quantitative results of various ablation components, focusing on the specific designs in the spatial noise separation and spectral noise distribution estimation, by which we perform a detailed analysis in terms of denoising performance and computational cost.

In concrete, firstly, through experiments on spatial noise separation, we verify that the introduction of Bootstrap-ConvB could bring over a 1.5 dB gain in the MPSNR metric and a few gains on the MSSIM and MSAM metrics with an increase of only about 72 K in model parameters. It also suggests the advantages of residual learning in Bootstrap-ConvB. However, the unpleasant denoising performance also implies that fine noise removal and HSI recovery cannot be achieved by barely using spatial information. Second, since we developed the spectral dimensional noise distribution branch based on QRNN to guide the subsequent denoising, a series of ablation experiments are conducted to investigate the improvements of each design. The experimental results demon-
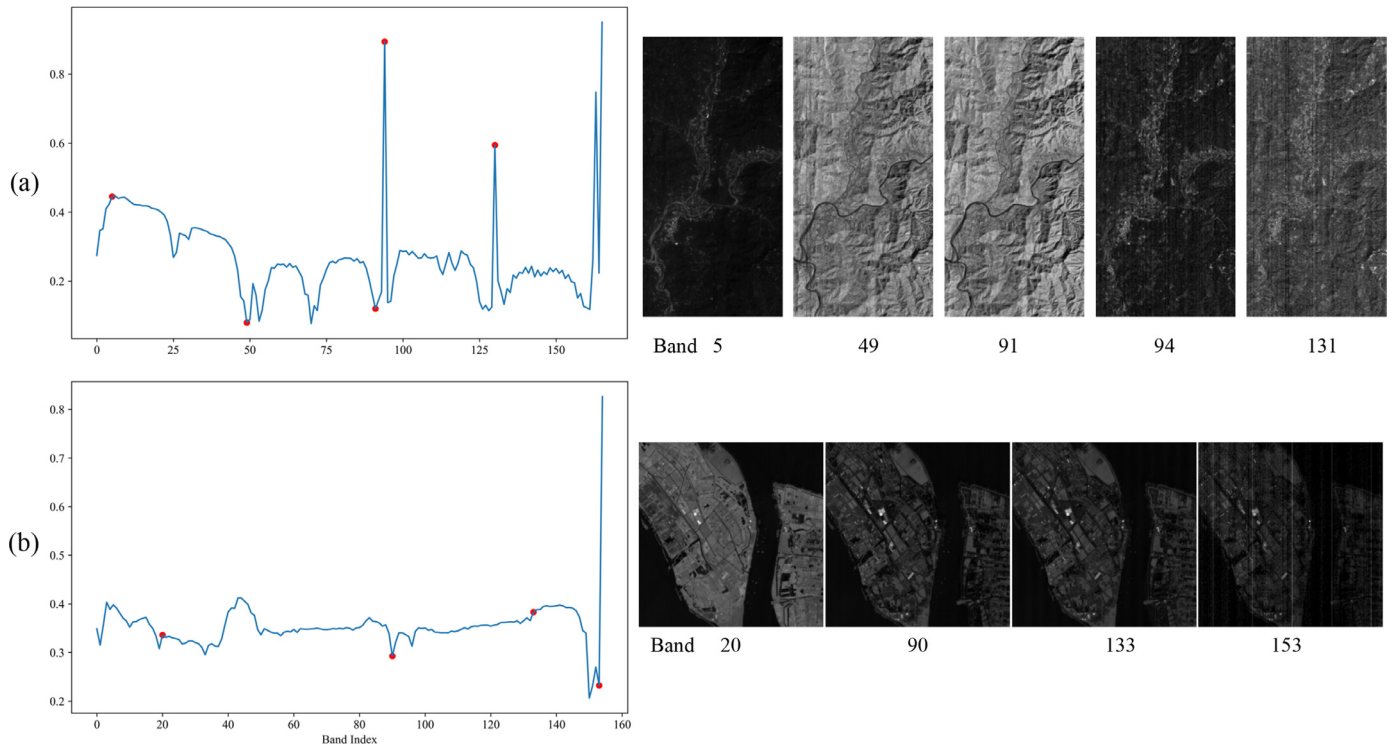
Fig. 14. Illustrations of spectral noise distribution estimated by the proposed $\hat{N}$-Net on different real-word HSIs. Images from some selected bands (marked in red) are also presented. (a) EO-1 Hyperion dataset; (b) GF-5 Shanghai dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
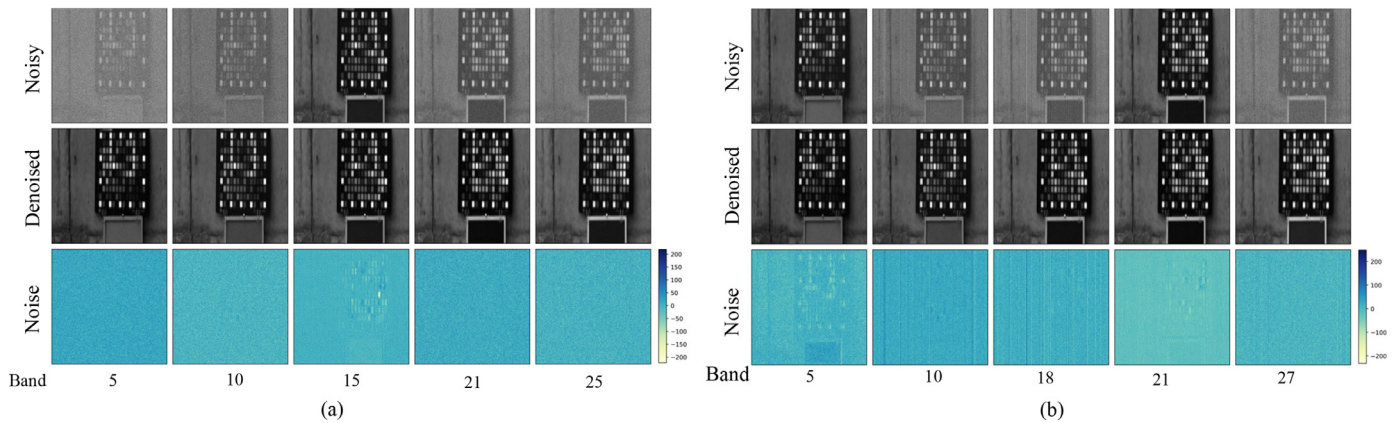

Fig. 15. Denoising performance of the proposed $\hat{N}$-Net model in the spatial dimension under two complex noise cases from the ICVL dataset are presented, respectively, including the noisy images, denoised images and noise maps in different bands. (a) Non-i.i.d. Gaussian noise case; (b) Non-i.i.d. Gaussian + Stripe noise case.

Table 3
Ablations on remove synthetic non-i.i.d. Gaussian noise on ICVL dataset. We evaluate the results by MPSNR, MSSIM, MSAM as well as the number of parameters (Params) of these networks. Our benchmark network is indicated by **boldface**.

| No.	Components						Metrics				
	Spectral noise distribution estimation			Guiding style	Spatial noise separation						
	QRNN	Bi-dense recurrent	Sparse recurrent		P-ConvB	Bootstrap-ConvB	MPSNR	MSSIM	MSAM	Params	FLOPs
1					✓		37.389	0.960	0.134	171.12K	10.48G
2					✓	✓	38.855	0.962	0.111	243.15K	14.47G
3	✓			once (×1)	✓	✓	40.688	0.983	0.071	263.99K	16.90G
4	✓	✓		once (×1)	✓	✓	40.827	0.985	0.068	263.99K	16.90G
5	✓	✓	✓	once (×1)	✓	✓	41.229	0.984	0.066	263.99K	16.90G
6	✓	✓	✓	**repeatedly (×3)**	✓	✓	**43.676**	**0.992**	**0.053**	**263.99K**	**16.90G**
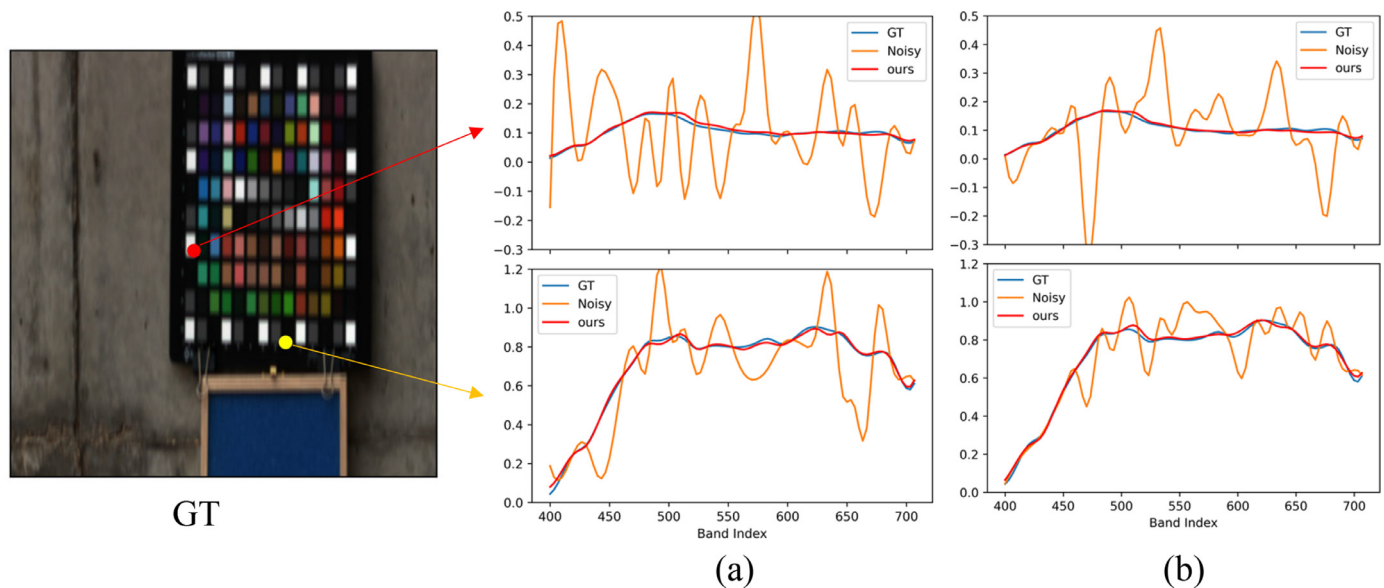
Fig. 16. Denoising performance of the proposed $\hat{N}$-Net model in the spectral dimension under (a) Non-i.i.d. Gaussian noise case; (b) Non-i.i.d. Gaussian + Stripe noise case.

strate that QRNN increases the model parameters from 243.15 K to 263.99 K, while it brings nearly 1.8 dB gain in the PSNR metric and over 0.02 improvement on MSSIM and decreases the MSAM value from 0.111 to 0.071. Besides, we also model the dense local dependence and global sparsity of the spectral dimensional noise and design the bi-directional dense and sparse recurrent calculations. The positive performance on these metrics indicates that it is necessary to consider the intrinsic properties of HSI noise distribution. Notably, these incremental designs that we perform based on QRNN have improved denoising performance without imposing any computational burden according to the metric of params.

Lastly, we assess the effectiveness of the self-guided style. For instance, the addition of spectral noise distribution estimation improves the MPSNR, MSSIM, and MSAM values to 40.688, 0.983, and 0.071, respectively. The repeatedly guiding style achieves superior denoising performance over all metrics. It brings over 2.4 dB gain in the MPSNR metric compared to guiding once, approaching 0.992 on MSSIM and decreasing MSAM to 0.053. Moreover, the guidance of $\hat{N}$ is introduced by leveraging the subbranch of the Bootstrap-ConvB (colored orange in the dashed box of Fig. 3). The repeated guidance is implemented by appending that subbranch to each Bootstrap-ConvB, which explains why it does not raise an extra computational burden on the model parameters. Such design facilitates a more robust ability to cope with varying degrees of noise.

Despite the addition of various optimization modules, the proposed $\hat{N}$-Net has a lightweight scale and the fastest running time compared to other methods. For instance, the proposed method has only 263.99 K total parameters and 16.90 G FLOPs, which is considerably less than other methods that have similar or poorer performance. Therefore, it can be concluded that the proposed $\hat{N}$-Net is effective and robust in handling complex synthetic mixture noise degradations in HSI denoising, while also being computationally efficient and having a lightweight scale

5. Conclusion and discussion

This paper proposes a novel HSI noise removal framework $\hat{N}$-Net based on spectral noise distribution bootstrap. It comprises three phases: spectral noise distribution estimation, spatial noise separation, and clean HSI recovery. Concretely, we design bidirectional dense and sparse recurrent calculations and propose a Sparse-QRNN-based block to capture the dense dependency and

global sparsity of HSI noise in the spectral dimension. Additionally, we progressively map the attributes vector into spectral noise distribution via multi-level convolutions in Trans-ConvB. In particular, we also raise a bootstrap mechanism and incorporate the estimate $\hat{N}$ into spatial noise separation through Bootstrap-ConvBs, addressing various spatial noise degradations in different bands and finely restoring a noise-free HSI. Abundant experiments on data with synthetic and natural noise have demonstrated the superiority of our method. The finding of this study suggests a potential and pleasing prospect of spectral distribution. Such an idea leveraging estimated feature distribution along the spectral dimension can also be easily extended to spatial or temporal dimensions, which is capable of supporting or bootstrapping subsequent tasks in inspiring future insightful research.

The limitation of our method and other existing HSI denoising methods is the unsatisfied fidelity of the recovered spectra in real HSI under extremely severe and complex noise corruption, as shown in Fig. 12 shows. The discrepancy between the real-world and synthetic data causes this phenomenon. In this case, a robust HSI denoising method should precisely remove anomalous spikes or perturbations from the spectral profile while retaining unique and distinguishable features in the spectral dimension. While our solution with spectral noise distribution can facilitate subsequent noise reduction along the spectral dimension with adaptive tuning, this is also primarily limited by the generalization capability of the trained model learned spectral noise distribution estimates over the actual dataset. However, it is worth mentioning that compared to existing methods, as presented in Fig. 13, our method has preserved the details and complete spectral features of the images to the maximum extent under different levels of noise contamination. One possible solution to alleviate the issue mentioned above is to adopt an unsupervised learning-based denoising framework, leveraging the noise distribution prior learned from practical HSIs to perform customized recovery of degraded HSIs. In the future, we will make efforts to construct such an unsupervised learning framework to mitigate this problem.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

References

[1] X. Zhang, H. Zhao, Hyperspectral-cube-based mobile face recognition: a comprehensive review, Inf. Fusion 74 (2021) 132–150.

[2] Y. Zhong, X. Wang, S. Wang, L. Zhang, Advances in spaceborne hyperspectral remote sensing in China, Geo-Spatial Inf. Sci. 24 (1) (2021) 95–120.

[3] V. Capelle, J.-M. Hartmann, Use of hyperspectral sounders to retrieve daytime sea-surface temperature from mid-infrared radiances: application to IASI, Remote Sens. Environ. 280 (2022) 113171.

[4] L. Fang, D. Zhu, J. Yue, B. Zhang, M. He, Geometric-spectral reconstruction learning for multi-source open-set classification with hyperspectral and Lidar data, IEEE/CAA J. Autom. Sin. 9 (10) (2022) 1892–1895.

[5] L. Tang, Y. Deng, Y. Ma, J. Huang, J. Ma, Superfusion: a versatile image registration and fusion network with semantic awareness, IEEE/CAA J. Autom. Sin. 9 (12) (2022) 2121–2137.

[6] D.H. Foster, K. Amano, Hyperspectral imaging in color vision research: tutorial, JOSA A 36 (4) (2019) 606–627.

[7] Z. Shao, W. Wu, D. Li, Spatio-temporal-spectral observation model for urban remote sensing, Geo-Spatial Inf. Sci. 24 (3) (2021) 372–386.

[8] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, H. Huang, Joint camera spectral response selection and hyperspectral image recovery, IEEE Trans. Pattern Anal. Mach. Intell. 44 (1) (2020) 256–272.

[9] W. He, Y. Chen, N. Yokoya, C. Li, Q. Zhao, Hyperspectral super-resolution via coupled tensor ring factorization, Pattern Recognit. 122 (2022) 108280.

[10] Y. Xu, J. Gong, X. Huang, X. Hu, J. Li, Q. Li, M. Peng, Luojia-HSSR: a high spatial-spectral resolution remote sensing dataset for land-cover classification with a new 3D-hrnet, Geo-Spatial Inf. Sci. (2022) 1–13.

[11] H. Qin, W. Xie, Y. Li, K. Jiang, J. Lei, Q. Du, Weakly supervised adversarial learning via latent space for hyperspectral target detection, Pattern Recognit. 135 (2023) 109125.

[12] Y. Chen, Y. Guo, Y. Wang, D. Wang, C. Peng, G. He, Denoising of hyperspectral images using nonconvex low rank matrix approximation, IEEE Trans. Geosci. Remote Sens. 55 (9) (2017) 5366–5380.

[13] H. Fan, Y. Chen, Y. Guo, H. Zhang, G. Kuang, Hyperspectral image restoration using low-rank tensor recovery, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10 (10) (2017) 4589–4604.

[14] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, L. Zhang, Non-local meets global: an iterative paradigm for hyperspectral image restoration, IEEE Trans. Pattern Anal. Mach. Intell. 44 (4) (2022) 2089–2107.

[15] R.C. Evangelista, D.H. Salvadeo, N.D. Mascarenhas, A new Bayesian poisson denoising algorithm based on nonlocal means and stochastic distances, Pattern Recognit. 122 (2022) 108363.

[16] Y. Chang, L. Yan, H. Fang, S. Zhong, W. Liao, HSI-DeNet: hyperspectral image restoration via convolutional neural network, IEEE Trans. Geosci. Remote Sens. 57 (2) (2018) 667–682.

[17] K. Wei, Y. Fu, H. Huang, 3-D quasi-recurrent neural network for hyperspectral image denoising, IEEE Trans. Neural Netw. Learn. Syst. 32 (1) (2021) 363–375.

[18] J. Ma, C. Peng, X. Tian, J. Jiang, DBDNet: a deep boosting strategy for image denoising, IEEE Trans. Multimed. 24 (2022) 3157–3168.

[19] E. Pan, Y. Ma, X. Mei, J. Huang, F. Fan, J. Ma, D2net: deep denoising network in frequency domain for hyperspectral image, IEEE/CAA J. Autom. Sin. 10 (3) (2022) 813–815.

[20] Y. Chang, L. Yan, S. Zhong, Hyper-Laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4260–4268.

[21] Q. Yuan, Q. Zhang, J. Li, H. Shen, L. Zhang, Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network, IEEE Trans. Geosci. Remote Sens. 57 (2) (2018) 1205–1218.

[22] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, L. Zhang, Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network, IEEE Trans. Geosci. Remote Sens. 57 (10) (2019) 7317–7329.

[23] W. He, H. Zhang, L. Zhang, H. Shen, Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration, IEEE Trans. Geosci. Remote Sens. 54 (1) (2016) 178–188.

[24] G. Chen, F. Zhu, P. Ann Heng, An efficient statistical method for image noise level estimation, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 477–485.

[25] A. Mittal, A.K. Moorthy, A.C. Bovik, No-reference image quality assessment in the spatial domain, IEEE Trans. Image Process. 21 (12) (2012) 4695–4708.

[26] J. Bradbury, S. Merity, C. Xiong, R. Socher, Quasi-recurrent neural networks, in: International Conference on Learning Representations (ICLR), 2017.

[27] Y. Peng, D. Meng, Z. Xu, C. Gao, Y. Yang, B. Zhang, Decomposable nonlocal tensor dictionary learning for multispectral image denoising, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2949–2956.

[28] J. Xue, Y. Zhao, W. Liao, J.C.-W. Chan, Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising, IEEE Trans. Geosci. Remote Sens. 57 (7) (2019) 5174–5189.

[29] W. Sun, Q. Du, Hyperspectral band selection: a review, IEEE Geosci. Remote Sens. Mag. 7 (2) (2019) 118–139.

[30] W. Chen, Z. Yang, J. Ren, J. Cao, N. Cai, H. Zhao, P. Yuen, MIMN-DPP: maximum-information and minimum-noise determinantal point processes for unsupervised hyperspectral band selection, Pattern Recognit. 102 (2020) 107213.

[31] H. Zhang, W. He, L. Zhang, H. Shen, Q. Yuan, Hyperspectral image restoration using low-rank matrix recovery, IEEE Trans. Geosci. Remote Sens. 52 (8) (2014) 4729–4743.

[32] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, D. Meng, Hyperspectral image restoration via total variation regularized low-rank tensor decomposition, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (4) (2017) 1227–1243.

[33] Y. Chen, X. Cao, Q. Zhao, D. Meng, Z. Xu, Denoising hyperspectral image with non-iid noise structure, IEEE Trans. Cybern. 48 (3) (2018) 1054–1066.

[34] H. Sun, L. Zhang, L. Wang, H. Huang, Stochastic gate-based autoencoder for unsupervised hyperspectral band selection, Pattern Recognit. 132 (2022) 108969.

[35] K. Zhang, W. Zuo, Y. Chen, D. Meng, L. Zhang, Beyond a gaussian denoiser: residual learning of deep CNN for image denoising, IEEE Trans. Image Process. 26 (7) (2017) 3142–3155.

[36] S. Guo, Z. Yan, K. Zhang, W. Zuo, L. Zhang, Toward convolutional blind denoising of real photographs, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1712–1722.

[37] A. Maffei, J.M. Haut, M.E. Paoletti, J. Plaza, L. Bruzzone, A. Plaza, A single model CNN for hyperspectral image denoising, IEEE Trans. Geosci. Remote Sens. 58 (4) (2019) 2516–2529.

[38] Y. Cai, J. Lin, X. Hu, H. Wang, X. Yuan, Y. Zhang, R. Timofte, L. Van Gool, Coarse-to-fine sparse transformer for hyperspectral image reconstruction, in: European Conference on Computer Vision (ECCV), Springer Nature Switzerland, Cham, 2022, pp. 686–704.

[39] Q. Zhang, Q. Yuan, J. Li, F. Sun, L. Zhang, Deep spatio-spectral Bayesian posterior for hyperspectral image non-iid noise removal, ISPRS J. Photogramm. Remote Sens. 164 (2020) 125–137.

[40] B. Arad, O. Ben-Shahar, Sparse recovery of hyperspectral signal from natural RGB images, in: European Conference on Computer Vision (ECCV), Springer International Publishing, Cham, 2016, pp. 19–34.

[41] E. Pan, Y. Ma, X. Mei, F. Fan, J. Huang, J. Ma, SQAD: spatial-spectral quasi-attention recurrent network for hyperspectral image denoising, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–14.

[42] L. Zhuang, M.K. Ng, Fasthymix: fast and parameter-free hyperspectral image mixed noise removal, IEEE Trans. Neural Netw. Learn. Syst. (2021) 1–15.

[43] F. Xiong, J. Zhou, Q. Zhao, J. Lu, Y. Qian, MAC-Net: model-aided nonlocal neural network for hyperspectral image denoising, IEEE Trans. Geosci. Remote Sens. 60 (2022) 1–14.

[44] Z. Lai, K. Wei, Y. Fu, Deep plug-and-play prior for hyperspectral image restoration, Neurocomputing 481 (2022) 281–293.

[45] H. Zhai, H. Zhang, L. Zhang, P. Li, Cloud/shadow detection based on spectral indices for multi/hyperspectral optical remote sensing imagery, ISPRS J. Photogramm. Remote Sens. 144 (2018) 235–253.

Erting Pan received the B.S. degree in electrical engineering and its automation from the Northeast Normal University, Changchun, China, in 2018, the M.E. degree in electronic and communication engineering from Wuhan University, Wuhan, China, in 2020. She is currently studying for the Ph.D. degree in the Multi-spectral Vision Processing Lab, Electronic Information School, Wuhan University, Wuhan, China. Her current research interests include remote sensing image processing, computer vision and pattern recognition.

Yong Ma graduated from the Department of Automatic Control, Beijing Institute of Technology, Beijing, China, in 1997. He received the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003. His general field of research is in signal and systems. His current research projects include remote sensing of the Lidar and infrared, as well as Infrared image processing, pattern recognition, interface circuits to sensors and actuators. Between 2004 and 2006, he was a Lecturer at the University of the West of England, Bristol, U.K. Between 2006 and 2014, he was with the Wuhan National Laboratory for Optoelectronics, HUST, Wuhan, where he was a Professor of electronics. He is now a Professor with the Electronic Information School, Wuhan University.

Xiaoguang Mei received the B.S. degree in communication engineering from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2007, the M.S. degree in communications and information systems from Huazhong Normal University, Wuhan, in 2011, and the Ph.D. degree in circuits and systems from the HUST, in 2016. From 2010 to 2012, he was a Software Engineer with the 722 Research Institute, China Shipbuilding Industry Corporation, Wuhan. From 2016 to 2019, He was a PostDoctoral Fellow with the Electronic Information School, Wuhan University, Wuhan. He is currently an assistant professor with the Electronic Information School, Wuhan University. His research interests include hyperspectral imagery, machine learning, and pattern recognition.

Fan Fan received the B.S. degree in communication engineering and the Ph.D. degree in electronic circuit and system from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2015, respectively. He is currently an assistant professor with the Electronic Information School, Wuhan University, China. His current research interests include infrared thermal imaging, machine learning, and computer vision.

Jiayi Ma received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively. From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He was a Post-Doctoral with the Electronic Information School, Wuhan University from August 2014 to November 2015, and received an accelerated promotion to Associate Professor and Full Professor in December 2015 and December 2018, respectively. He has authored or co-authored more than 120 refereed journal and conference papers, including IEEE TPAMI/TIP/TSP/TNNLS/TIE/TGRS /TCYB/TMM/TCSVT, IJCV, CVPR, ICCV, IJCAI, AAAI, ICRA, IROS, ACM MM, etc. His research interests include computer vision, machine learning, and pattern recognition. Dr. Ma has been identified in the 2019 Highly Cited Researchers list from the Web of Science Group. He was a recipient of the Natural Science Award of Hubei Province (first class), the CAAI (Chinese Association for Artificial Intelligence) Excellent Doctoral Dissertation Award (a total of eight winners in China), and the CAA (Chinese Association of Automation) Excellent Doctoral Dissertation Award (a total of ten winners in China). He is an Editorial Board Member of Information Fusion and Neurocomputing, and a Guest Editor of Remote Sensing.