SQAD: Spatial-Spectral Quasi-Attention Recurrent Network for Hyperspectral Image Denoising

Erting Pan, Yong Ma, Xiaoguang Mei, Fan Fan, Jun Huang, *Member, IEEE*, and Jiayi Ma, *Senior Member, IEEE*

***Abstract*— This article presents a novel end-to-end model based on encoder–decoder architecture for hyperspectral image (HSI) denoising, named spatial-spectral quasi-attention recurrent network, denoted as SQAD. The central goal of this work is to incorporate the intrinsic properties of HSI noise to construct a practical feature extraction module while maintaining high-quality spatial and spectral information. Accordingly, we first design a spatial-spectral quasi-recurrent attention unit (QARU) to address that issue. QARU is the basic building block in our model, consisting of spatial component and spectral component, and each of them involves a two-step calculation. Remarkably, the quasi-recurrent pooling function in the spectral component could explore the relevance of spatial features in the spectral domain. The spectral attention calculation could strengthen the correlation between adjacent spectra and provide the intrinsic properties of HSI noise distribution in the spectral dimension. Apart from this, we also design a unique skip connection consisting of channelwise concatenation and transition block in our model to convey the detailed information and promote the fusion of the low-level features with the high-level ones. Such a design helps maintain better structural characteristics, and spatial and spectral fidelities when reconstructing the clean HSI. Qualitative and quantitative experiments are performed on publicly available datasets. The results demonstrate that SQAD outperforms the state-of-the-art methods of visual effect and objective evaluation metrics.**

***Index Terms*— Hyperspectral image (HSI) denoising, quasi-recurrent neural network (QRNN), self-attention, skip connection, spatial-spectral feature extraction.**

I. Introduction

WITH the rapid development of remote sensing imaging technology and intelligent interpretation algorithms, remote sensing images have become an indispensable resource in various applications, such as land resources surveys [1], agricultural and forestry monitoring [2], [3], urban planning [4], and military early warning [5], [6]. Compared with other types of remote sensing data, hyperspectral image (HSI) has a tremendous edge with its rich spatial and spectral information. Nevertheless, due to the inevitable distraction and contamination during HSI acquisition, storage, and transmission, real HSIs are usually suffering from noises of sorts, as presented in Fig. 1. Noises can severely degenerate the quality of HSIs, thereby impairing the performance of subsequent feature extraction and comprehension tasks, such as HSI unmixing [7], [8], feature learning [9], classification [10]–[13], and object detection [14]. In consequence, removing degradation caused by noises is an initial and crucial issue before other interpretation tasks.

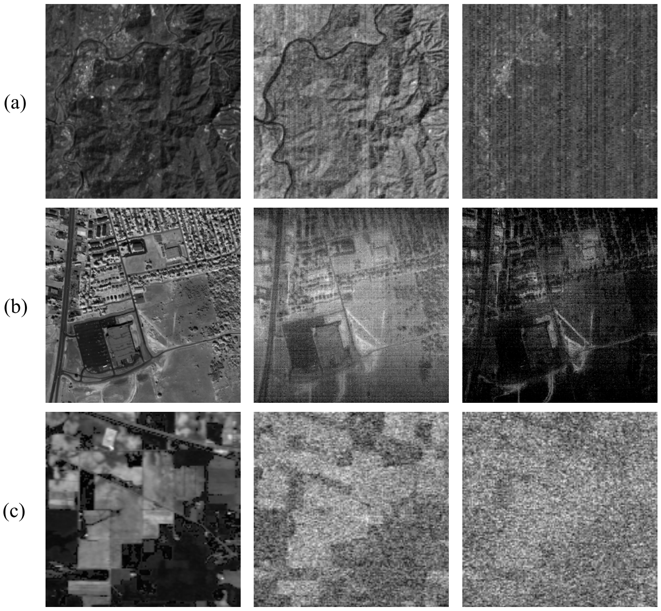


Fig. 1. Examples of unknown complex noise in different HSIs and different bands. (a) Earth Observing-1 (EO-1) Hyperion data in 51th, 101th, and 166th bands. (b) Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Urban data in 81th, 104th, and 108th bands. (c) Indian Pines data in 21th, 105th, and 109th bands.

To date, numerous methods based on traditional machine learning or emerging deep learning have been proposed for HSIs denoising. Traditional knowledge prior-based methods usually constrain the solution space of this ill-conditioned problem by designing certain assumptions or priors. Many methods adopt spatial-spectral priors, involving sparse [15], [16], low-rank [17]–[20], nonlocal similarity [21]–[23], or total variation [24]–[26], to reasonable regularize the model and have achieved promising denoising performance. Some other approaches [27], [28] work on the tensor-based method to explore structure correlations and multilinear prior in

Manuscript received December 12, 2021; revised February 13, 2022; accepted March 1, 2022. Date of publication March 10, 2022; date of current version April 5, 2022. This work was supported in part by the National Nature Science Foundation of China under Grant 61903279. *(Corresponding author: Xiaoguang Mei.)*

The authors are with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: panerting@whu.edu.cn; mayong@whu.edu.cn; meixiaoguang@gmail.com; fanfan@whu.edu.cn; junhwong@whu.edu.cn; jyma2010@gmail.com).

Digital Object Identifier 10.1109/TGRS.2022.3156646

HSIs. Unfortunately, such methods severely rely on manually designed features and require to be solved iteratively with multiple optimizations, which is very time-consuming. Worse still, these assumptions and prior information usually reflect features of HSIs only partially, making the model unable to deal with different types of noise, which seriously limits its practical applications. By contrast, learning-based methods can tackle the HSI denoising problem more flexibly and efficiently with the power of deep learning. Recently, inspired by advancing works of denoising RGB images, more studies emerged. Some of them have demonstrated that separate extraction of spatial and spectral features can be effectively used for denoising. For example, parallel feature extraction can be achieved using cascaded convolutional blocks [29] or multibranch networks [30]. However, this may lead to neglect of the spatial and spectral correlations inherent in HSIs. Moreover, 3-D convolution-based networks [31], [32] have also been designed to efficiently acquire joint spatial and spectral features, but these methods are less effective in recovering contextual details and impose a greater computational burden. In a nutshell, it is defective to process HSI data by directly following the common idea of denoising RGB images. Such ideas ignore intrinsic adjacent dependency of HSI noises in the spectral dimension, leading to unsatisfied denoising results. In addition, HSI noise removal is a relatively low-level feature learning task. It implies that the depth of the network may not be the key to better denoising performance, as confirmed by the literature [33]. These abovementioned HSI denoising methods will be reviewed detailedly in Section II. Accordingly, incorporating the intrinsic properties of HSI noise to construct a practical feature extraction module while maintaining high-quality spatial and spectral information is of great importance for HSI noise removal, and this is the central goal in this work, too.

To this end, on the basis of the encoder–decoder framework in deep learning, we propose a novel end-to-end model of HSI denoising, named the spatial-spectral quasi-attention recurrent network, abbreviated as SQAD. In our SQAD, on the one hand, we design a quasi-attention recurrent unit (QARU) to model the inherent spatial and spectral correlations. Specifically, we adopt 2-D convolution in the spatial dimension and f-pooling in the spectral dimension to fully exploit the spatial-spectral features. In particular, as the noise is distributed with strong dependency in adjacent bands, we raise a spectral attention calculation to further strengthen these correlations in the spectral dimension. On the other hand, as low-level structural features are essential in reconstructing clean images in HSI denoising, we design a specific skip connection involving a channelwise concatenation and a transition block between corresponding levels to achieve information transfer at different levels. This strategy facilitates retaining high spatial and spectral fidelities and makes the final SQAD form in an asymmetric structure. The experimental results of state-of-the-art comparisons prove its superior. The contributions of our work include the following three aspects.

1) We construct an end-to-end deep learning model SQAD, which frees the focus of research from the assumptions of feature extraction or the complex design of *a priori*.
2) We propose QARU consisting of spatial and spectral components, which can fully explore the intrinsic spatial-spectral features of HSIs and precisely capture the adjacent dependency of noises.
3) We design channelwise concatenations with specific transition blocks for feature transmission, forming our SQAD in an asymmetric encoder–decoder structure. Such a design facilitates maintaining structural features better and retaining high spatial and spectral fidelities via the fusion of low-level features with high-level ones.

The remainder of this article is arranged as follows. Section II reviews existing methods related to denoising. In Section III, we give a brief introduction of the HSI denoising degradation model and describe the design of the proposed method in detail. Qualitative and quantitative comparisons, as well as ablation studies, are performed in Section IV. Finally, the conclusion is presented in Section V.

II. Related Work

This section provides a brief review of existing HSI denoising methods. Then, as an extension, we briefly introduce the quasi-recurrent neural network (QRNN) and its typical applications in the image processing community.

A. Existing HSI Denoising Methods

The mainstream of existing methods for HSI denoising can be roughly classified into two types.

1) *Knowledge Prior-Based Methods:* Many denoising methods adopt domain prior knowledge of HSI, such as spatial-spectral nonlocal similarity and global spectral correlation, which has successfully mapped noisy observations to clean HSI. Some of them, based on total variation, sparse representation, low-rank, and tensor models, have achieved extraordinary performance. For example, in [21], block matching 4-D (BM4D) is proposed to exploit local correlations in each HSI subcube and nonlocal correlations among different HSI subcubes. Following it, a tensor dictionary learning model (TDL) [22] employs both the nonlocal similarity over the spatial domain and the global correlation across the spectra of HSIs. Subsequently, low-rank matrix recovery (LRMR) [17] adopts an efficient HSI restoration method based on low-rank matrix restoration. In [19], Chang *et al.* claim that nonlocal self-similarity is critical for denoising and present a one-way low-rank tensor recovery method to capture the structural correlations inherent in HSIs. In order to combine the spatial nonlocal similarity with the low-rank characteristics of the global spectrum, nonlocal meets global (NG-Meet) [23] presents a unified spatial-spectral paradigm for HSI denoising. Xue *et al.* [34] designs a nonlocal low-rank regularized tensor decomposition to fully utilize the global correlation across the spectrum and nonlocal self-similarity properties for HSI denoising. The main drawback of the above approaches is that the manually introduced prior knowledge reflects only certain aspects

of the HSI features, limiting their representational capabilities. Besides, they are time-consuming due to the complicated optimization process.

2) *Learning-Based Methods:* The deep learning theory provides a data-driven strategy to solve complex problems in an end-to-end manner. Through training with large amounts of data, deep learning-based methods can eliminate tedious manual feature design and specific parameter tuning under different situations. Recently, many researchers have made some related attempts aimed at HSI denoising [29]–[32], [35]–[38]. For instance, HSI-DeNet [35] first introduces deep convolutional neural networks in HSI denoising, learning a series of multichannel 2-D filters for the spatial and spectral structures of HSIs. HSI denoising convolutional neural network (HSID-CNN) [30] uses spatial and spectral information to recover clean images through two parallel feature extraction branches. Similarly, Maffei *et al.* [36] design a denoising method using a single CNN [HSI single denoising CNN (HSI-SDeCNN)]. Hereafter, considering the directivity and spectral difference of the spatial structure, a spatial-spectral gradient network (SSGN) [29] is proposed for removing mixed noise in HSIs. These methods have demonstrated that extracting spatial and spectral features in parallel can be helpful in denoising HSIs. However, although these methods have achieved promising denoising effects, they also have a common drawback in that they ignore the intrinsic spatial-spectral correlations in HSIs, and there is still great potential for further exploration and promotion in this field.

B. Quasi-Recurrent Neural Network

A variety of features need to be considered when modeling long sequence problems. For example, local semantics can be computed in parallel (e.g., convolution) because of the local context invariance. In contrast, features that rely on long-range global contextual information have to be computed recurrently. Unfortunately, many existing neural network architectures cannot balance the exploration of contextual information with parallel computation.

To address this drawback of standard models, Bradbury *et al.* [39] design QRNN. It can take advantage of parallelism and contextual information, showing the benefits of both convolutional and recurrent neural networks. Many variants of QRNN are tailored to several natural language tasks, and some other subsequent related works have been published successively [40]–[44], showing excellent scalability and potential capability of QRNN. Later, the literature [31] proposes to extend it to the HSI denoising task by proposing the QRNN3D. It introduces QRNN into the 3-D U-net structure using 3-D convolution to extract structural spatial-spectral correlations while using the quasi-recurrent pooling function to extract spectral correlations. Nevertheless, this model has apparent shortcomings. First, 3-D convolution can only extract features locally and consumes extensive computational resources. Second, the quasi-recurrent pooling function actually correlates only with the features in the adjacent bands. It is similar to the computation of the spectral dimension in 3-D convolution. Accordingly, the model has duplicated computations on spectral features and does not really involve the global spectral features. With regards to these issues, we further propose our SQAD model for the HSI denoising task.

III. Proposed Method

In this section, we briefly introduce the noise degradation model for HSI first. Then, we provide the specific design and the formulation of QARU. Finally, we give a description of the network architecture with the particular design in detail.

A. HSI Denoising Problem Formulation

A clean HSI can be described as a 3-D tensor $\mathcal{X} = \{X^1, X^2, \dots, X^B\} \in \mathbb{R}^{W\times H\times B}$, where matrix $X^i \in \mathbb{R}^{W\times H} (i = 1, 2, 3, \dots, B)$ represents the ith band, W and H denote the spatial dimension, and B denotes the spectral dimension.

Denoting the noisy HSI observations as $\mathcal{Y} = \{Y^1, Y^2, \dots, Y^B\} \in \mathbb{R}^{W\times H\times B}$, and the additive noise degradation model for HSI can be described as follows:

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

where $\mathcal{N} \in \mathbb{R}^{W\times H\times B}$ indicates noises involving dense noise, such as Gaussian noise, and sparse noise, such as impulse noise, stripe noise, and deadlines. In consequence, the HSI denoising problem is defined as obtaining the noise-free counterpart $\mathcal{X}$ from the only known $\mathcal{Y}$.

B. Quasi-Attention Recurrent Unit

Recovering a completely clean $\mathcal{X}$ is a very challenging task. According to (1), the degradation effect of HSI noises is additive. It implies that the noisy observation $\mathcal{Y}$ has a very high structural similarity to the reconstructed noisy-free $\mathcal{X}$. Therefore, adequate feature extraction is of great importance for reconstructing clean HSIs. Nevertheless, on the one hand, the noise distribution of each band differs, involving different noisy types or levels, and requires that the spatial feature extraction conducted separately in each band can capture reliable and discriminative information. On the other hand, the adjacent dependency of HSI noises is also considerable, which desires us to pay attention to the spatial-spectral feature correlations and the local dependency in the spectral dimension.

In view of this, corresponding to different focuses in HSI feature extraction, we design QARU as the essential feature extraction block in our algorithm. As illustrated in Fig. 2, the designed QARU consists of two components, including spatial component and spectral component, and each of them involves a two-step calculation, which will be described in detail next.

1) Spatial Component: We find that the HSI denoising and reconstruction performance in each band is markedly dependent on the spatial contextual information at different scales due to spatial nonlocal similarity. A standard solution is to adopt a multiscale convolution for spatial feature extraction, which will significantly increase the model complexity. Consequently, in this article, we choose to employ scaling

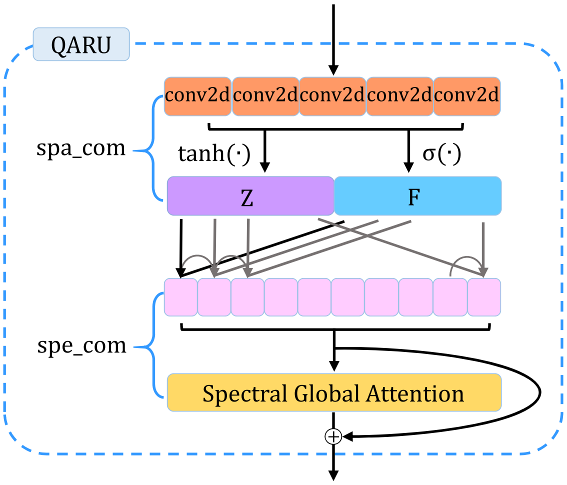


Fig. 2. Inner structure illustration of the proposed QARU. It consists of two components, including spatial component and spectral component, and each of them involves a two-step calculation.

operations to acquire HSI patches at different resolutions during data augmentation to extract spatial features at different scales using the fixed scale of convolution. It facilitates the enlargement of the sample diversity and the capture of multiscale features. The designed spatial convolutional component is achieved by employing 2-D convolutional filter banks to each HSI band in the spatial domain. It allows fully parallel computation across both minibatches and spatial dimensions. The 2-D convolution in the spatial dimensions can mimic a series of image transformation operations widely used in low-level vision to obtain practical local spatial features.

Given the input feature maps $X \in \mathbb{R}^{p\times B\times W\times H}$(in the first layer, X is the original HSI patch with $p = 1$), the spatial convolutional component performs 2-D convolution in the spectral dimension with separate filter banks, producing a candidate tensor $Z \in \mathbb{R}^{q\times B\times W\times H}$ and a forget gate $F \in \mathbb{R}^{q\times B\times W\times H}$ through different activation functions

$$\begin{aligned} Z &= \tanh(W_z * X) \\ F &= \sigma\left(W_f * X\right) \end{aligned} \tag{2}$$

where W_z and W_f, each in $\mathbb{R}^{p\times q\times k}$ (k indicates the $1 \times 3 \times 3$ convolutional kernel), are the convolutional filter banks and $*$ denotes a 2-D convolution. The candidate tensor Z is passed through a tanh nonlinearity, and the forget gate F employs an elementwise sigmoid σ.

2) Spectral Component: The strong spectral correlation in HSIs, in other words, its low-rank property, has been widely used for HSI denoising. However, we observe that modeling the spatial-spectral correlation of HSI has a more positive impact on the denoising and reconstruction process. Therefore, upon acquiring the spatial features of each band in the spatial component, it is necessary to effectively exploit the interspectral relationship of these spatial features.

To this end, we first present a quasi-recurrent f-pooling operation along the spectral bands, in which the dynamic average pooling operation and gating mechanism represented by a forget gate are involved. In particular, the quasi-recurrent f-pooling function is designed to perform along the spectral dimension. To be specific, we split the candidate tensor Z and forget gate F, and generate sequences of z_b and f_b, respectively. Then, we mix these tensors recurrently across the whole spectrum as (3)

$$h_b = f \odot h_{b-1} + (1 - f) \odot z_b \quad \forall b \in [1, B] \tag{3}$$

where $\odot$ denotes an elementwise multiplication and h_{b-1} indicates the previous hidden state and the $(b - 1)$th band in the output (h_0 is initialized as zero). The forget gate f_b is mainly to weigh the current candidate tensor z_b and the previous memory h_{b-1}. The forget gate f_b there acts as a filter, which represents a way of information transmission based on selection. Unlike a fixed convolutional filter in CNNs, the value of f_b relies on the current input feature maps. It could effectively adapt to the current band. Since the dynamic f-pooling recurrently calculates along the whole spectrum, the interspectrum correlations would be fully exploited. Finally, all hidden states $\{h_1, h_2, \ldots, h_B\}$ concatenate along the spectrum, producing the output feature maps H.

On the other hand, most bands in HSIs are of high quality, while only some specific bands are degraded by various noises. Therefore, reducing the noise in the corrupted bands while preserving the high-quality ones is of great importance for HSI denoising. In this case, it is necessary to draw global dependencies of spectral features to relate to different bands. Thus, the spectral feature extraction should be accompanied by an attention-driven and long-range dependency mechanism. Specifically, the designed spectral attention operation constructs a spectral global attention calculation and improves the nonlinearity expression ability of the QARU.

In this spectral attention calculation, it first creates three representations (query, key, and value) for each band by applying learned linear projections. The subsequent attention computation can be done for the entire spectral feature maps H in parallel by grouping queries, keys, and values in H_q, H_k, and H_v representations, as shown in the following formula:

$$H_\alpha = \mathrm{softmax}\left(\frac{H_q H_k^T}{\sqrt{d_k}}\right) H_v \tag{4}$$

where d_k is the dimension of the key representations and H_α indicates self-attention feature maps, which takes softmax(·) to obtain attention scores. With a learned parameter γ, the output feature map H_o can be formulated as

$$H_o = \gamma \times H_\alpha + H. \tag{5}$$

C. Network Architecture

The overall structure of the proposed SQAD is illustrated in Fig. 3. To efficiently leverage the features captured by QARU and reconstruct noise-free HSI in high fidelity, our SQAD first incorporates three pairs of symmetric QARU and DeQARU layers, forming the backbone architecture with an encoder–decoder framework. In particular, the encoder employs QARU to perform layer-by-layer feature extraction and downsampling, and the decoder utilizes DeQARU with deconvolution to fulfill layer-by-layer upsampling. In addition,

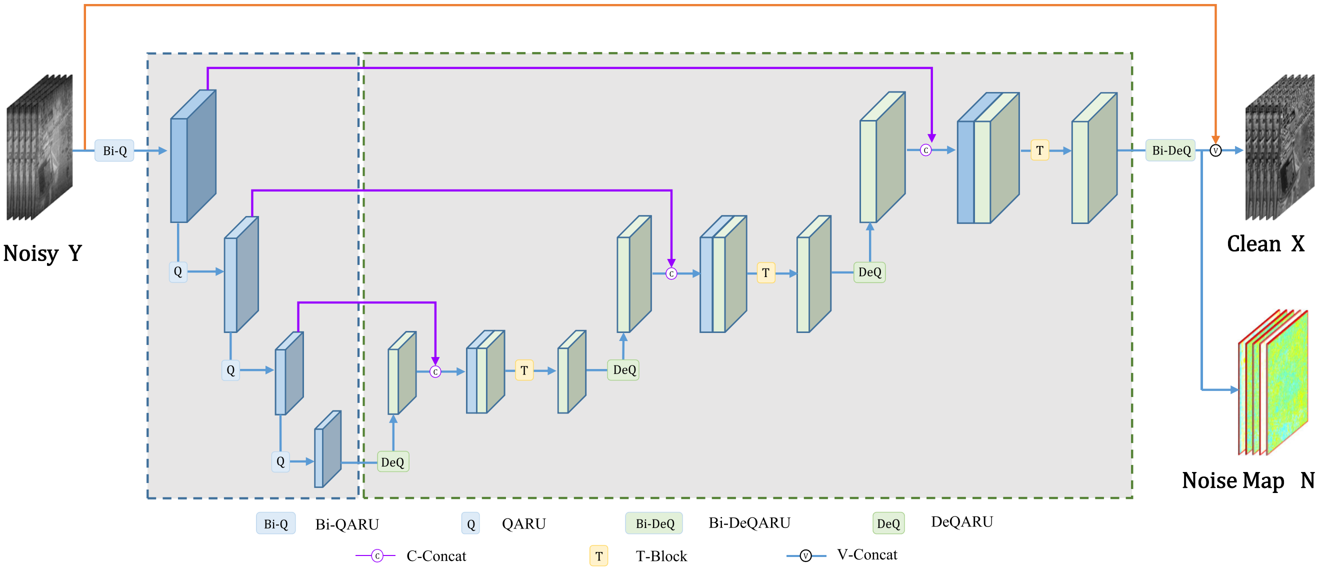


Fig. 3. Overall architecture illustration of the proposed SQAD. The main body consists of stacked downsampling QARU with convolution for the encoder (blue) and upsampling DeQARU with deconvolution for the decoder (green). In addition, we add skip connections to convey features at different levels. The skip connection in each decoder layer and the corresponding encoder layer consists of the channelwise concatenation (C-Concat) and the transition block (T-Block). Another skip connection indicated by the (orange) line in the figure consists of valuewise concatenation (V-Concat).

we raise some other specific designs for enhancing the denoising capability and maintaining spatial and spectral information in high quality.

On the one hand, it is worth mentioning that all the spatial 2-D convolutions in our method actually employ a specific 3-D form, where the stride and kernel of the spectral dimension are fixed to 1. Such a strategy makes our method does not need to constrain the number of spectral bands so that the proposed SQAD can denoise any HSIs with arbitrary bands.

On the other hand, as the network grows deeper, the corresponding feature maps will have larger receptive fields and retain less detailed information. Nevertheless, more reliable denoising results strongly require making up the information loss during the downsampling process. Skip connection [45] has been proven that it can transfer feature information from the previous layer to the later layer, thus maintaining detailed information. Considering the features of HSI noises at different levels, unlike common symmetric encoder–decoder frameworks, we propose a specific skip connection consisting of the channelwise concatenation (C-Concat) and the transition block (T-Block), which forms our SQAD in an asymmetric structure. It is able to fuse the feature maps at the corresponding layers and weaken the gradient disappearance or explosion problem.

To be specific, the channelwise concatenation enhances the information transfer of feature maps between layers via supplementing low-level features of the corresponding scales into the upsampling and deconvolution processes. The structure of the transition block can be described as a sandwich style, as shown in Fig. 4, which can maximize the utilization of features. A batch normalization (BN)-rectified linear unit (ReLU)-Conv combination in the center enables the transition block to further fuse the detailed information from the low-level layer into the high-level one. A BN-ReLU-Conv combination in the center enables the transition block to further fuse the detailed information from the low-level layer into the high-level one. In particular, the 1×1 convolution kernel in its outer layer can dramatically reduce the channel of feature maps and release some computational burden. Such a design finally makes the SQAD form in an asymmetric encoder–decoder structure and have the ability to maintain high spatial and spectral fidelities. On the basis of this, corresponding to (1), another skip connection between the noisy Y and noise map N is designed as valuewise concatenation (V-Concat), which can convey the underlying features for the final reconstruction of noise-free X.

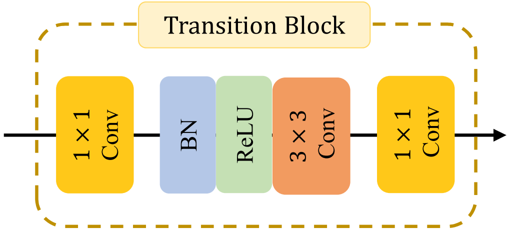


Fig. 4. Illustration of the designed transition block inner structure. It consists of two 1×1 convolutional layers, a combination of a BN layer, an ReLU activation, and a 3×3 convolutional layer.

In summary, beyond the inherent features of HSI noises captured efficiently by QARU, our SQAD in asymmetric structure further retains more high-resolution and detailed information in the process of recovering the clean HSI. Table I lists our network configuration. Stride and output channels in each layer are listed, and other configurations (e.g., padding) can be inferred implicitly.

TABLE I

NETWORK CONFIGURATION OF OUR UNSYMMETRIC ENCODER–DECODER STYLE SQAD FOR HSI DENOSING

Layer		Channel	Stride	Output size
Extractor	Bi-QARU	16	(1,1,1)	$B \times W \times H$
Encoder	QARU	16	(1,1,1)	$B \times W \times H$
	QARU	32	(1,2,2)	$B \times \frac{W}{2} \times \frac{H}{2}$
	QARU	64	(1,2,2)	$B \times \frac{W}{4} \times \frac{H}{4}$
Decoder	DeQARU	64	$(1, \frac{1}{2}, \frac{1}{2})$	$B \times \frac{W}{2} \times \frac{H}{2}$
	C-Concat	128	–	
	T-Block	64	(1,1,1)	
	DeQARU	32	$(1, \frac{1}{2}, \frac{1}{2})$	$B \times W \times H$
	C-Concat	64	–	
	T-Block	32	(1,1,1)	
	DeQARU	16	(1,1,1)	$B \times W \times H$
	C-Concat	32	–	
	T-Block	16	(1,1,1)	
Reconstructor	Bi-DeQARU	1	(1,1,1)	$B \times W \times H$
	V-Concat	1	(1,1,1)	$B \times W \times H$

IV. EXPERIMENTS AND DISCUSSIONS

A. Experimental Settings

1) Data: To validate the performance of the proposed SQAD, we perform qualitative and quantitative experiments on the publicly available the Interdisciplinary Computational Vision Laboratory (ICVL)[1] hyperspectral dataset [46]. It collected 201 images at 1392 × 1300 spatial resolution over 31 spectral bands with spectrum range from 400 to 700 nm at 10-nm increments. Some RGB rendering samples of this dataset are shown in Fig. 5. All the hyperspectral data cubes in this dataset are regarded as clean HSIs. In our experiment, we adopt 100 images for training and five images for validation while the rest for testing. The training data are cropped in a spatial patch size of 64 × 64 with the complete spectrum preserving. Data augmentation is performed by scaling at a rate in $\{1, 0.5, 0.25\}$ and random rotation.

Besides, we also evaluate the robustness and flexibility of our model in remote sensing HSIs, including the EO-1 Hyperion[2] dataset with 166 bands [47], [48], the Urban[3] dataset with 210 bands, and the Indian Pines[4] datasets with 200 bands [49].

2) Simulated Noise Settings: As mentioned above, real HSIs are often corrupted by dense noise (i.e., Gaussian noise), sparse noise (i.e., non-i.i.d. Gaussian, stripe, deadline, and impulse noise), or a mixture noise. In our experiments, the training data are generated by adding different types of simulated noise to the clean HSI. The additive noise is simulated according to the following cases.

1) *Case 1 (Gaussian Noise):* Adding Gaussian noise with the fixed intensity for different bands. The mean of the noise distribution is 0, and the variance σ is 50. Here, σ denotes the noise level or noise intensity.
2) *Case 2 (Non-i.i.d. Gaussian Noise):* All bands are polluted by zero-mean Gaussian noise with a random intensity ranged from 30 to 70.
3) *Case 3 (Non-i.i.d. Gaussian + Stripe Noise):* The noise distribution is multiple. All bands are corrupted by non-i.i.d. Gaussian noise as Case 2. In addition, one-third of bands are randomly selected to add stripe noise with 5%–15% of columns.
4) *Case 4 (Non-i.i.d. Gaussian + Deadline Noise):* Each band is corrupted by non-i.i.d. Gaussian noise as Case 2. Besides, one-third of bands are randomly selected to add deadline noise with 5%–15% of columns.
5) *Case 5 (Non-i.i.d. Gaussian + Impulse Noise):* All bands are contaminated by Gaussian noise as in Case 2. One-third of bands are randomly chosen to add impulse noise with intensity ranged from 10% to 70%.
6) *Case 6 (Mixture Noise):* Each band is randomly corrupted by Gaussian noise as Case 1 and at least one kind of noise mentioned in Cases 2–5.

Fig. 5. RGB rendering samples of the hyperspectral ICVL dataset.

The above six cases can be roughly classified into two types, and case 1 is a type of Gaussian noise, while cases 2–6 belong to a complex noise type.

3) Evaluation Measures: Four quantitative quality indices are employed for performance evaluation, including the peak signal-to-noise ratio (PSNR), the structure similarity (SSIM), the feature similarity (FSIM), and the spectral angle mapper (SAM). For one thing, PSNR, SSIM, and FSIM are three mainstream metrics for image quality assessment, all of which can be used as spatial-based indexes. PSNR is a traditional objective fidelity criterion; SSIM is an effectiveness and efficiency metric in line with human intuition. It measures similarity according to luminance, contrast, and structure. FSIM is another metric based on the human visual system and understands images mainly according to its low-level features. For another thing, SAM is a spectral-based metric to measure the spectral similarity between the noisy HSI and the reconstructed clean HSI. For PSNR, SSIM, and FSIM, a larger value implies better performance, while a smaller value suggests better performance for SAM. In addition,

[1] It is available at http://icvl.cs.bgu.ac.il/hyperspectral/

[2] It is available at http://hipag.whu.edu.cn/ziyuanxiazai.html

[3] It is available at https://www.erdc.usace.army.mil/Media/Fact-Sheets/Fact-Sheet-Article-View/Article/610433/hypercube/

[4] It is available at https://engineering.purdue.edu/ biehl/MultiSpec/hyper-spectral.html

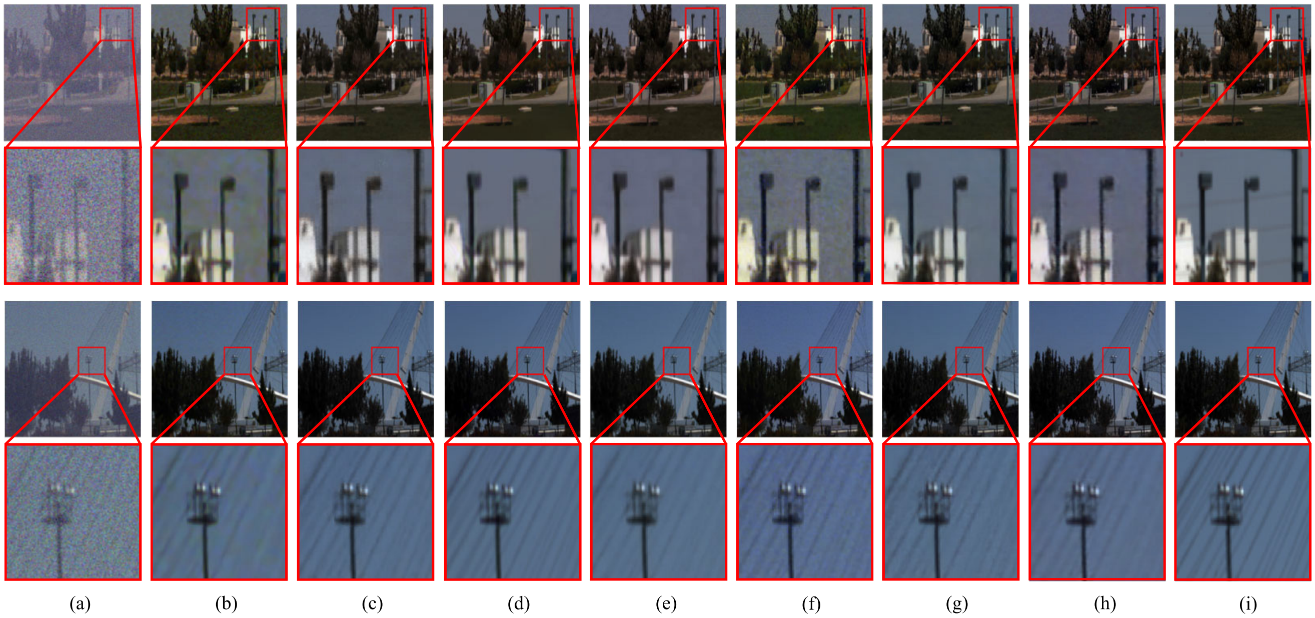


Fig. 6. Simulated Gaussian noise removal results on the ICVL dataset. Examples for all the competing methods on bands 5, 12, and 20 of the ICVL dataset under the Gaussian noise with $\sigma = 50$ and blind Gaussian noise are presented, respectively. Zoomed-in view gives better visualization. (a) Noisy HSI. (b) BM4D. (c) TDL. (d) LLRT. (e) ITSReg. (f) HSID-CNN. (g) QRNN3D. (h) our SQAD. (i) Ground truth.

we report the average running time of all the cases for each method. All running times are measured by processing a noisy HSI with the size of $512 \times 512 \times 31$.

4) Training Details: SQAD is implemented on the PyTorch platform and accelerated on NVIDIA TITAN RTX GPU. We adopt the mean square error loss (MSEloss) as the loss function. The network is trained by minimizing the MSEloss between the noisy HSI and the clean ground-truth pairs. Parameters in SQAD are initialized as Kaiming Initialization and updated by the Adam optimizer. The learning rate is set to 0.001 with exponential decay at epochs, where the validation performance does not increase anymore.

In order to maintain the stability of the training process and improve the denoising performance, we follow the training policy that trains in an easy-to-difficult way. Instead of training multiple networks independently to tackle different types of noise separately, only one network is trained using a training sample with various noise intensities in Gaussian and complex noise types. Thus, the training process of our SQAD contains two phases from the easier Gaussian denoising task with fixed noise intensity to the more difficult noise removal task with complex noise. In the first phase, the zero-mean Gaussian noise with noise level $\sigma = 50$ is added into the training images to construct the training dataset. In the second phase, we use the aforementioned complex noises (from Case 2 to Case 6) to generate the training data. More specifically, a total of 120 epochs are set up throughout the training process, the first 60 epochs are trained for Gaussian denoising, and the latter 60 epochs are trained for complex noise removal. The batch size of the first phase is set to a mini size (i.e., 8) to accelerate training, while a large batch size (i.e., 32) is adopted to stabilize training when handling complex noise cases.

B. Results and Analysis for Synthetic Experiments

To validate the efficiency and robustness of the proposed SQAD, we compare it with several advanced denoising methods on different noise cases. In particular, we first evaluate its performance under Gaussian noise case with different noise levels ($\sigma = 30, 50, 70$, and blind) and then test the complex noise cases involving Cases 2–6.

1) Gaussian Noise Case With Different Intensities: We compare the proposed SQAD with several state-of-the-art HSI denoising methods involving four traditional methods (i.e., filtering-based approach BM4D [21], tensor dictionary learning approach TDL [22], low-rank tensor recovery hyper-Laplacian regularized unidirectional low-rank tensor recovery (LLRT) [19], and the intrinsic tensor sparsity regularization (ITSReg) [15]) and two deep learning methods (i.e., HSID-CNN [30], and QRNN3D [31]). In this case, we adopt three fixed noise intensities (i.e., 30, 50, and 70) and a blind noise intensity to evaluate these methods on denoising Gaussian noise. To provide a fair comparison, all the comparative algorithms use the parameters that have been either optimally assigned or chosen, as described in the references. Next, specific analyses and comparisons are presented. We perform qualitative comparisons on typical HSIs from the ICVL dataset, and the results are shown in Fig. 6, where we additionally amplify an area of interest in the reconstructed images for easier observation. We analyze the results from two aspects, i.e., the whole image and details. From the perspective of the whole image, it can be easily observed that the HSI restored by our method is proficient in sufficiently removing the Gaussian noise while finely retaining the structure underlying the HSI. The results of our method show the best qualitative effect in line with human visual perception compared with other results. From the perspective of details, the proposed method seems to fail in restoring the details and textures of a few thin lines.

TABLE II

Quantitative Performance of All Competing Methods Under Gaussian Noise With Different Intensities on the ICVL Dataset. "Blind" Suggests That Each Sample Is Polluted by Gaussian Noise With Unknown σ (Ranged From 30 to 70). The Best Results Under Different Noise Intensities Are **Boldfaced**, Respectively

σ	Metrics	Noisy HSI	BM4D [21]	TDL [22]	LLRT [19]	ITSReg [15]	HSID-CNN [30]	QRNN3D [31]	SQAD(ours)
30	PSNR	18.59	38.18	40.69	41.65	41.43	39.79	**42.28**	41.73
	SSIM	0.105	0.934	0.955	0.965	0.958	0.959	0.973	**0.985**
	FSIM	0.644	0.955	0.975	0.976	0.976	0.974	0.98	**0.983**
	SAM	0.714	0.109	0.058	0.052	0.078	0.074	0.073	**0.617**
50	PSNR	14.15	35.53	38.52	38.83	39.189	37.56	**40.21**	39.74
	SSIM	0.043	0.893	0.930	0.942	0.935	0.937	0.958	**0.967**
	FSIM	0.489	0.927	0.958	0.957	0.960	0.959	0.972	**0.978**
	SAM	0.907	0.153	0.0836	0.074	0.102	0.091	0.084	**0.073**
70	PSNR	11.23	33.71	36.92	37.22	37.46	36.42	**38.30**	38.14
	SSIM	0.023	0.855	0.910	0.926	0.919	0.923	0.938	**0.958**
	FSIM	0.398	0.903	0.945	0.940	0.945	0.948	0.951	**0.959**
	SAM	1.027	0.182	0.099	0.085	0.113	0.099	0.094	**0.083**
blind	PSNR	17.59	37.66	40.44	41.03	40.89	39.02	**41.65**	41.30
	SSIM	0.121	0.918	0.948	0.956	0.949	0.950	0.965	**0.982**
	FSIM	0.598	0.943	0.968	0.968	0.969	0.968	0.972	**0.978**
	SAM	0.776	0.128	0.069	0.062	0.093	0.080	0.076	**0.065**
	Time/s	-	307.63	50.52	1640.64	1880.25	3.03	1.22	**0.98**

However, QRNN loses some fine-grained details related to these lines and shows oversmoothing regions, while our proposed method preserves more texture details. Besides, other methods produce relatively low-quality results compared with ours. For example, in the Gaussian noise case, traditional methods, such as BM4D and TDL, suffer obvious artifacts in some areas, showing the effect of excessive smoothing, which are evident in Fig. 6.

Table II lists quantitative denoising results. Compared with the leading methods, such as TDL, LLRT, ITSReg, and HSID-CNN under the Gaussian noise cases with different noisy levels, our method acquires comparable or even better denoising results in terms of the three metrics (SSIM, FSIM, and SAM). However, the performance in the PSNR metric is slightly inferior to QRNN3D. In addition, Fig. 7(a) shows the PSNR and SSIM value of each band of a typical HSI in the testing set under Gaussian noise case with blind intensity. We can see that the PSNR metric of the proposed SQAD only reaches comparable values in some bands, while the SSIM metric of SQAD presents the highest value in all bands. These observations indicate that SQAD may generate a few false details and edges in some bands, leading to its inferiority on PSNR measurement. Nevertheless, the superior performance on both SSIM and FSIM metrics demonstrates a more powerful and robust ability of our SQAD to retain the structure features and recover the edge and detail information. The results on the SAM metric also prove that SQAD can keep spectral fidelity. In addition, once the model is trained, the running time of the proposed SQAD method is far less than the traditional methods, and it also can achieve better denoising results with a faster or comparable efficiency than the other two deep learning methods.

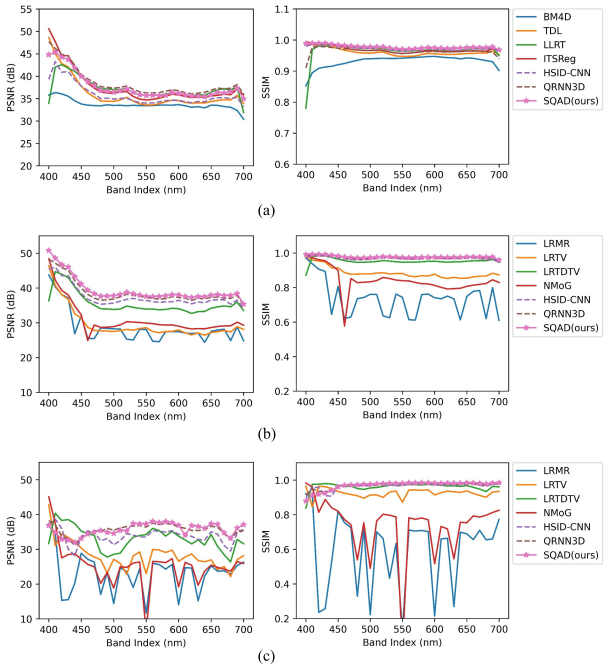
2) Complex Noise Cases: We compare the proposed SQAD with several state-of-the-art denoising methods, including four traditional methods and two deep learning methods too. Uniquely, traditional methods generally rely on strong noise assumptions, resulting in more suitability for specific noise settings. Therefore, we compare our method against different traditional baselines in Gaussian and complex noise cases. In complex noise cases as Cases 2–6, the traditional comparative methods include low-rank matrix recovery-based methods (LRMR [17] and total variation (TV)-regularized low-rank matrix factorization (LRTV) [24]), low-rank tensor method (TV-regularized low-rank tensor decomposition (LRTDTV) [50]), and non i.i.d. mixture of Gaussian denoising (NMoG) [51]. All the comparative algorithms use the optimal parameters as described in the references. Fig. 6 shows some visual results of these methods with amplified details. It is apparent that low-rank matrix recovery methods, i.e., LRMR and LRTV, successfully remove a great mass of noise but at the cost of losing fine details. Compared with the two-deep learning-based methods (i.e., HSID-CNN and QRNN3D), the results of our method achieve more fidelity and are much clearer. Moreover, the reconstructed image of the proposed SQAD has a more consistent tone with the original image, while the tones of the results of LRTV, NMoG, and HSID-CNN are more grayish. It also implies that our method has a more robust capability in exploring spatial and spectral correlations.

In Table III, it can be easily seen that the proposed method can eliminate more noises and achieve better performance in most quantitative assessments compared with all competing methods. Unfortunately, the results under Case 2 and Case 6 reveal its slight disadvantage on the PSNR metric. It might be caused by exceptional edges and details falsely generated

TABLE III

QUANTITATIVE PERFORMANCE OF ALL COMPETING METHODS UNDER COMPLEX NOISE CASES ON THE ICVL DATASET. THE BEST RESULTS UNDER DIFFERENT NOISY CASES ARE **BOLDFACED**, RESPECTIVELY

	Metrics	Noisy HSI	LRMR [17]	LRTV [24]	LRTDTV [50]	NMoG [51]	HSID-CNN [30]	QRNN3D [31]	SQAD(ours)
Case 2	PSNR	14.64	27.76	32.61	36.34	32.44	37.94	**39.95**	39.12
	SSIM	0.058	0.478	0.882	0.919	0.685	0.948	0.957	**0.973**
	FSIM	0.522	0.892	0.852	0.949	0.949	0.966	0.971	**0.979**
	SAM	0.849	0.322	0.073	0.080	0.078	0.065	0.059	**0.050**
Case 3	PSNR	14.58	27.63	32.65	36.18	29.96	37.73	38.79	**39.41**
	SSIM	0.057	0.474	0.882	0.918	0.609	0.946	0.956	**0.962**
	FSIM	0.517	0.889	0.852	0.948	0.927	0.965	0.969	**0.977**
	SAM	0.851	0.324	0.074	0.082	0.223	0.067	0.060	**0.052**
Case 4	PSNR	14.47	27.05	31.21	34.03	29.16	37.18	38.64	**39.07**
	SSIM	0.056	0.471	0.871	0.897	0.603	0.943	0.955	**0.967**
	FSIM	0.513	0.883	0.849	0.939	0.924	0.962	0.966	**0.974**
	SAM	0.866	0.340	0.113	0.107	0.223	0.069	0.061	**0.055**
Case 5	PSNR	12.75	23.99	31.19	35.1	25.58	35.48	**37.16**	36.65
	SSIM	0.046	0.371	0.852	0.908	0.499	0.915	0.928	**0.939**
	FSIM	0.468	0.811	0.842	0.942	0.859	0.955	0.963	**0.967**
	SAM	0.868	0.465	0.216	0.099	0.459	0.107	0.094	**0.087**
Case 6	PSNR	12.44	23.26	29.81	32.67	24.97	34.64	**36.69**	36.22
	SSIM	0.041	0.365	0.839	0.884	0.469	0.908	0.919	**0.937**
	FSIM	0.450	0.801	0.839	0.933	0.851	0.948	0.954	**0.962**
	SAM	0.886	0.478	0.25	0.121	0.466	0.115	0.095	**0.089**
	Time/s	-	11.49	385.89	412.71	258.02	3.22	1.32	**0.97**



(a)

(b)

(c)

Fig. 7. PSNR and SSIM values across the spectrum corresponding to Gaussian and complex noise removal results in Figs. 6 and 8, respectively. (a) Case 1 with blind Gaussian noise. (b) Case 3 with Non-i.i.d. Gaussian nosie and stripe noise. (c) Case 6 with mixture noise.

by SQAD. On the other hand, our method dramatically outperforms some state-of-the-art denoising methods under all complex noise cases, such as LRTDTV, NMoG, and HSID-CNN, especially on SSIM, FSIM, and SAM metrics. In addition, Fig. 7(b) and (c) shows the PSNR and SSIM values of each band in the testing HSIs under cases involving Case 3 with Non-i.i.d. Gaussian noise and stripe noise, and Case 6 with mixture noise. Evidently, some PSNR and SSIM curves have apparent oscillations, such as the NMoG curve and the LRTDTV curve, which shows that these traditional methods cannot have a stable performance on all bands. Besides, the SQAD metric curve of PSNR shows a contrary performance under Case 3 and Case 6, which can be explained by the previously analyzed reason for some weaker results on quantitative PSNR. On the contrary, it also can be observed that the SSIM values of all bands obtained by our SQAD in Fig. 7 are obviously higher than those comparison methods, which exhibits a more robust generalization performance of SQAD for denoising HSIs in different bands. These observations further confirm the superiority and robustness of reconstructing clean HSI with high spatial and spectral fidelities.

C. Results and Analysis for Real HSI Denoising

We also verify our model in real-world noisy HSIs without corresponding ground truth, including Urban and Indian Pines datasets. The size of the Urban dataset is 307 × 307 × 210. The Indian Pines dataset is with the size of 145 × 145 × 220. In these real HSI datasets, some bands are seriously corrupted by the atmosphere and water, and polluted by complex noises (e.g., deadline, stripe, sparse, and Gaussian noise), which causes a big challenge for the reconstruction of the clean HSI.

As we mentioned before, the QARU in our method can denoise any HSIs with arbitrary bands. It makes our method be naturally used for input data with a various number of bands. Based on this flexibility, we directly employ our model pretrained on the ICVL dataset (in complex noise cases)

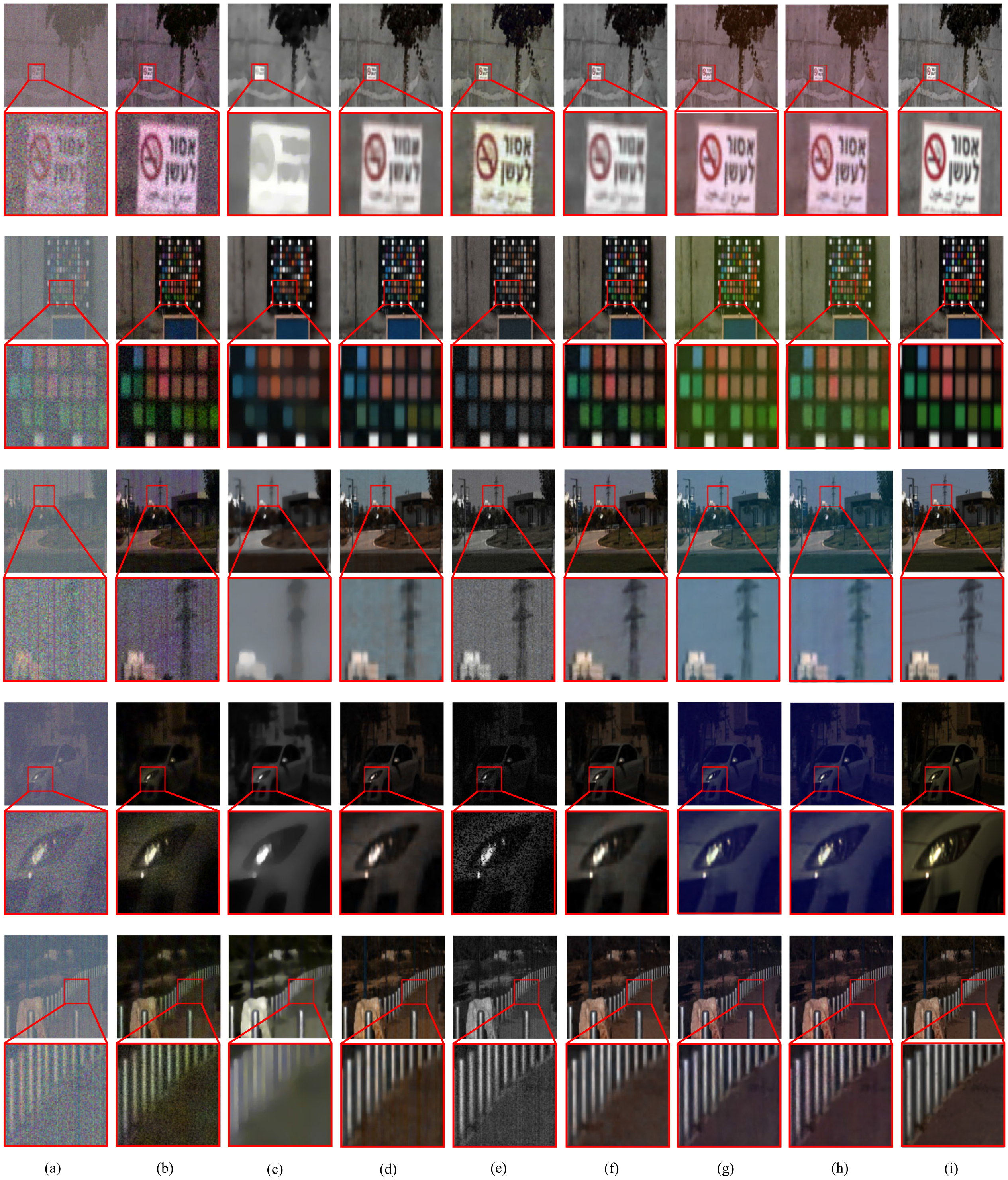


Fig. 8. Simulated complex noise removal results on the ICVL dataset. Examples for all the competing methods on bands 5, 12, and 20 of the ICVL dataset under Cases 2–6 are presented, respectively. Zoomed-in view gives better visualization. (a) Noisy his. (b) LRMR. (c) LRTV. (d) LRTDTV. (e) NMoG. (f) HSID-CNN. (g) QRNN3D. (h) our SQAD. (i) Ground truth.

to denoise these real-world HSIs. Besides, to comprehensively compare the denoising performance, the competing methods in this experiment include TDL, LLRT, ITSReg, LRMR, LRTV, LRTDTV, NMoG, HSID-CNN, and QRNN3D. The qualitative results of this experiment are shown in Figs. 9 and 10.

Specifically, it can be observed in Figs. 9 and 10 that terrible atmosphere and water absorption degrade the quality of HSIs, severely damaging the visual effect of the real scenario. Since the noise of real-world HSIs is usually non-i.i.d., some Gaussian denoising methods, e.g., BM4D and TDL, cannot accurately estimate the clean HSI. From Fig. 9, it can be

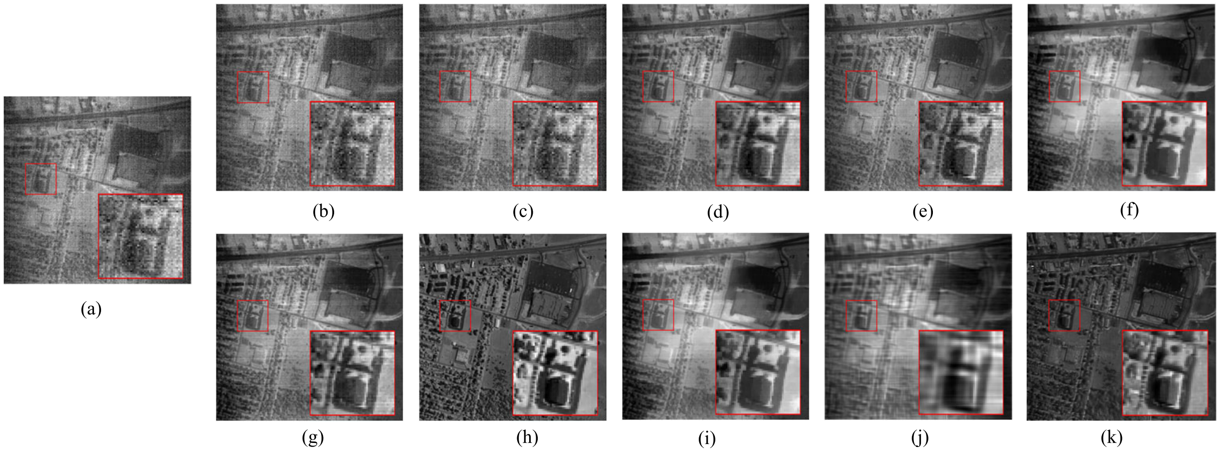


Fig. 9. Denoising results of the Urban dataset at the 103th band. Zoomed-in view gives better visualization. (a) Noisy HSI. (b) TDL. (c) LLRT. (d) ITSReg. (e) LRMR. (f) LRTV. (g) LRTDTV. (h) NMoG. (i) HSID-CNN. (j) QRNN3D. (k) SQAD (ours).

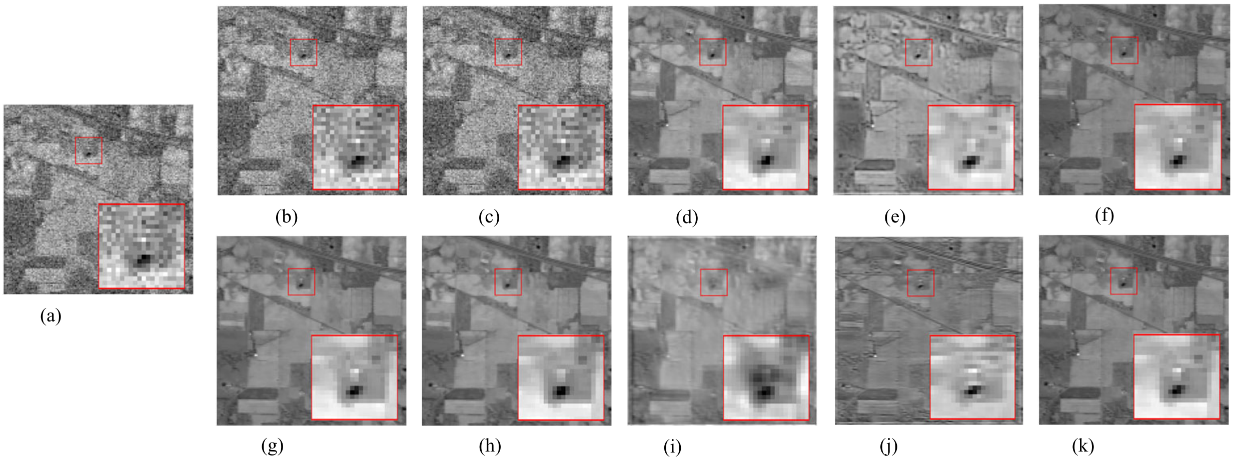


Fig. 10. Denoising results of the Indian Pines dataset at the 102th band. Zoomed-in view gives better visualization. (a) Noisy HSI. (b) TDL. (c) LLRT. (d) ITSReg. (e) LRMR. (f) LRTV. (g) LRTDTV. (h) NMoG. (i) HSID-CNN. (j) QRNN3D. (k) SQAD (ours).

easily seen that other competing methods can not completely remove the complex noise from the real-world HSIs, while our proposed SQAD performs better in both detail preservation and noise removing. In addition, Fig. 10 shows the recovered HSIs of the urban datasets by the competing methods. We can observe that our SQAD can achieve better denoising performance than most of them and obtain comparable performance. These experiments demonstrate that our SQAD has the ability to handle unknown noises and produces sharper and clearer results than other methods. It also consistently indicates the robustness and flexibility of our model, as well as the powerful generalization performance.

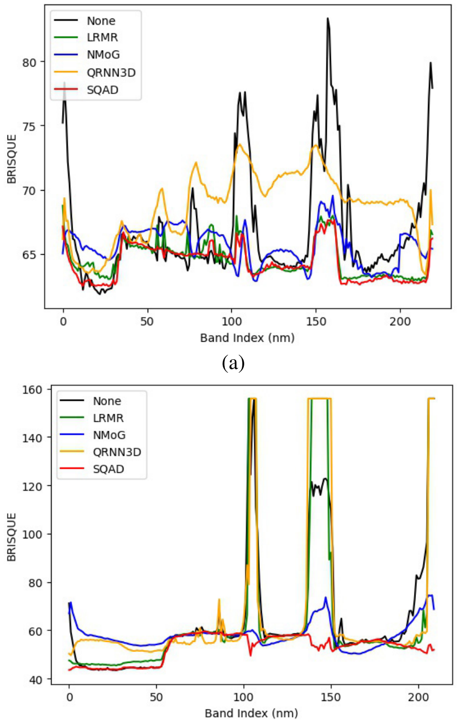
In addition, we add the blind/referenceless image spatial quality evaluator (BRISQUE) for no-referenced quantitative evaluation. For presenting a clearer comparison, we only select a few comparative algorithms for quantitative evaluation of real HSI denoising. Unlike nature images, HSI data contain hundreds of images with a single band, so we calculate the image quality scores in each band and plot them in Fig. 11. We also record the average BRISQUE scores in Table IV. Notably, BRISQUE evaluates the image quality scores according to the degree of distortion calculated by the pixel intensity distribution. The smaller the image quality score, the better the subjective quality. It is obvious that the proposed SQAD earns better quality on the vast majority of bands, demonstrating relatively stable denoising performance, especially for some bands damaged by severe noise. On the other hand, the stability of other comparative denoising algorithms is somewhat suffered by the varying degree of noise contamination in each band. It also reconfirms the positive effect of considering the noise distribution in the spectral dimension and introducing spectral attention calculation on HSI denoising.

TABLE IV
AVERAGE IMAGE QUALITY SCORES EVALUATED BY BRISQUE IN REAL HSI DENOISING

BRISQUE	None	LRMR	NMoG	QRNN3D	SQAD
Urban	64.26	63.82	57.99	66.78	**52.79**
Indian Pines	66.89	64.53	65.62	68.68	**64.25**

D. Ablation Study

To verify the effectiveness of each subcomponent in our SQAD downright, comprehensive ablation studies are



(a)

(b)

Fig. 11. Image quality scores across the spectrum evaluated by BRISQUE corresponding to the real HSI denoising. (a) Indian Pines dataset. (b) Urban dataset.

performed on HSI Gaussian denoising on the ICVL dataset. The main concern of the ablation studies contains the components associated with HSI modeling and the best trade-off between denoising performance and computational cost. We adopt PSNR, SSIM, and the total number of parameters of the network as the evaluation metrics. The proposed encoder–decoder SQAD is the benchmark of these ablation experiments.

To be fair, we keep the same network architecture except for the modification in the investigated component. Table V lists the relevant indicators with various ablation experiments, and then, a detailed analysis is carried out.

1) Subcomponents Investigation: Table V investigates the effect of subcomponents (i.e., spatial convolution, quasi-recurrent pooling, and spectral attention) in QARU, which is the basic building block of our SQAD.

In the ablation experiments, three variants of this basic block are tested, i.e., RES2D, quasi-recurrent unit (QRU) 2D, and QRU3D versus our proposed QARU. RES2D is constructed by removing the quasi-recurrent pooling with the associated gates and the spectral attention calculation. Severe performance loss can be seen in Table V, indicating that a lack of a mechanism to model the spectral characteristics would degrade the denoising performance to a large extent. QRU2D is constructed by a 2-D convolution and quasi-recurrent pooling function. QRU3D is instantiated by replacing the 2-D convolution with the 3-D convolution in QRU2D, which is also the core component of QRNN3D. Both of them involve no spectral attention calculation. In QARU, the spatial convolution in QARU is 3-D convolution in 2-D form, allowing the model to generalize to arbitrary HSIs. As shown in Table V, the number of parameters of QRU3D is comparable to QARU, which implies that the introduction of spectral attention calculation does not bring a computational burden. Instead, it improves the denoising performance via enhancing the spectral correlations. In contrast, our proposed QARU has higher efficiency toward HSI denoising.

TABLE V

Ablations on ICVL HSI Gaussian Denoising (Under Noise Level $\sigma = 50$). We Evaluate the Results by PSNR (dB), SSIM, and the Number of Parameters (Params) of These Networks. Our Benchmark Network Is Indicated by **Boldface**. The Result of QRNN3D Is Also Provided as an Additional Reference

Model	PSNR	SSIM	Params(#)
QRNN3D	40.21	0.961	3.28M
RES2D	35.46	0.873	0.16M
QRU2D	37.91	0.933	0.40M
QRU3D	39.53	0.957	1.18M
QARU	39.74	0.978	1.18M
N-net	37.77	0.937	0.39M
V-net	38.36	0.953	1.17M
C-net	38.87	0.973	1.19M
SQAD	39.74	0.978	1.18M

2) Skip Connections: Table V also shows the results under different kinds of skip connections. N-net does not employ any skip connections, and V-net utilizes additive skip connection directly in feature values of different levels, while C-net employs channelwise concatenation with a transition block.

N-net performs the worst since it does not consider the information loss in the high-level layers. The performance of V-net and C-net significantly exceeds the N-net and requires a relatively close cost of computation resources. Nevertheless, the V-net scheme is relatively more lightweight, while the C-net obtains a better denoising performance, suggesting that the design of channelwise concatenation and transition block in the C-net can be used as an alternative to the typical skip connections. To trade off the denoising performance and computational cost, we adopt C-net in the encoder–decoder framework and V-net between the shallow feature extraction and final reconstruction, forming our SQAD. The result of SQAD indicates that our scheme achieves state-of-the-art performance.

V. Conclusion

In this work, we propose a spectral-spatial quasi-attention recurrent network for HSI denoising, termed SQAD. We first investigate extracting intrinsic spectral and spatial features of HSI noises and elaborate on the design of the core building block QARU of our method. It consists of spatial and spectral components. In particular, we raise spectral global attention calculation in QARU, which could enhance the correlation between adjacent spectra and provide the noise distribution features of HSI in the spectral dimension. Moreover, another focus beyond denoising HSIs is maintaining spatial and spectral fidelities. In this work, we design a specific skip connection composed of the channelwise concatenation with a transition block, which could not only fuse the low-level features with the high-level ones but also release some computational burden. Such a design finally makes the SQAD form in an asymmetric encoder–decoder structure. The ablation studies have proven that it has the ability to maintain better spatial and spectral fidelities when reconstructing the clean HSI. We have validated the efficiency and robustness of our method on different noise cases. Compared with other state-of-the-art fusion methods, our method can achieve advanced performance both qualitatively and quantitatively.

References

[1] J. Transon, R. D'Andrimont, A. Maugnard, and P. Defourny, "Survey of hyperspectral earth observation applications from space in the sentinel-2 context," *Remote Sens.*, vol. 10, p. 157, Jan. 2018.

[2] M. S. M. Asaari *et al.*, "Close-range hyperspectral image analysis for the early detection of stress responses in individual plants in a high-throughput phenotyping platform," *ISPRS J. Photogramm. Remote Sens.*, vol. 138, pp. 121–138, Apr. 2018.

[3] G. V. Laurin *et al.*, "Discrimination of tropical forest types, dominant species, and mapping of functional guilds by hyperspectral and simulated multispectral Sentinel-2 data," *Remote Sens. Environ.*, vol. 176, pp. 163–176, Apr. 2016.

[4] C. Weber *et al.*, "Hyperspectral imagery for environmental urban planning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 1628–1631.

[5] B. Zhang, D. Wu, L. Zhang, Q. Jiao, and Q. Li, "Application of hyperspectral remote sensing for environment monitoring in mining areas," *Environ. Earth Sci.*, vol. 65, no. 3, pp. 649–658, Feb. 2012.

[6] M. Shimoni, R. Haelterman, and C. Perneel, "Hypersectral imaging for military and security applications: Combining myriad processing and sensing techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 101–117, Jun. 2019.

[7] Q. Jin *et al.*, "Hyperspectral unmixing with Gaussian mixture model and spatial group sparsity," *Remote Sens.*, vol. 11, no. 20, p. 2434, Oct. 2019.

[8] Y. Ma *et al.*, "Hyperspectral unmixing with Gaussian mixture model and low-rank representation," *Remote Sens.*, vol. 11, no. 8, p. 911, Apr. 2019.

[9] X. Li, M. Ding, and A. Pizurica, "Fully group convolutional neural networks for robust spectral–spatial feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2022.

[10] E. Pan, X. Mei, Q. Wang, Y. Ma, and J. Ma, "Spectral–spatial classification for hyperspectral image based on a single GRU," *Neurocomputing*, vol. 387, pp. 150–160, Apr. 2020.

[11] E. Pan, Y. Ma, F. Fan, X. Mei, and J. Huang, "Hyperspectral image classification across different datasets: A generalization to unseen categories," *Remote Sens.*, vol. 13, no. 9, p. 1672, Apr. 2021.

[12] X. Zhao, R. Tao, W. Li, W. Philips, and W. Liao, "Fractional Gabor convolutional network for multisource remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–18, 2022.

[13] J. Jiang, J. Ma, Z. Wang, C. Chen, and X. Liu, "Hyperspectral image classification in the presence of noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 851–865, Feb. 2018.

[14] C. Wu, B. Du, and L. Zhang, "Hyperspectral anomalous change detection based on joint sparse representation," *ISPRS J. Photogramm. Remote Sens.*, vol. 146, pp. 137–150, Dec. 2018.

[15] Q. Xie *et al.*, "Multispectral images denoising by intrinsic tensor sparsity regularization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1692–1700.

[16] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips, "Spatial–spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.

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

[18] C. Li, Y. Ma, J. Huang, X. Mei, and J. Ma, "Hyperspectral image denoising using the robust low-rank tensor recovery," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 32, no. 9, pp. 1604–1612, Sep. 2015.

[19] Y. Chang, L. Yan, and S. Zhong, "Hyper-Laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4260–4268.

[20] F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Inf. Sci.*, vols. 397–398, pp. 48–68, Aug. 2017.

[21] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, Apr. 2013.

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

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

[24] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

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

[26] X. Tian, W. Chen, Z. Wang, and J. Ma, "Polarization prior to single-photon counting image denoising," *Opt. Exp.*, vol. 29, no. 14, pp. 21664–21682, 2021.

[27] H. Fan, C. Li, Y. Guo, G. Kuang, and J. Ma, "Spatial–spectral total variation regularized low-rank tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 6196–6213, Aug. 2018.

[28] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.

[29] Q. Zhang, Q. Yuan, J. Li, X. Liu, H. Shen, and L. Zhang, "Hybrid noise removal in hyperspectral imagery with a spatial–spectral gradient network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7317–7329, Dec. 2019.

[30] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, Sep. 2018.

[31] K. Wei, Y. Fu, and H. Huang, "3-D quasi-recurrent neural network for hyperspectral image denoising," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 1, pp. 363–375, Jan. 2021.

[32] W. Liu and J. Lee, "A 3-D atrous convolution neural network for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5701–5715, Aug. 2019.

[33] P. Liu and R. Fang, "Wide inference network for image denoising via learning pixel-distribution prior," 2017, *arXiv:1707.05414*.

[34] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5174–5189, Jul. 2019.

[35] Y. Chang, L. Yan, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, Feb. 2018.

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

[37] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3844–3851.

[38] J. Ma, C. Peng, X. Tian, and J. Jiang, "DBDNet: A deep boosting strategy for image denoising," *IEEE Trans. Multimedia*, early access, Jul. 1, 2021, doi: 10.1109/TMM.2021.3094058.

[39] J. Bradbury, S. Merity, C. Xiong, and R. Socher, "Quasi-recurrent neural networks," 2016, *arXiv:1611.01576*.

[40] F. Bolelli, L. Baraldi, F. Pollastri, and C. Grana, "A hierarchical quasi-recurrent approach to video captioning," in *Proc. IEEE Int. Conf. Image Process., Appl. Syst. (IPAS)*, Dec. 2018, pp. 162–167.

[41] M. Wang *et al.*, "Quasi-fully convolutional neural network with variational inference for speech synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 7060–7064.

[42] Y. Fu, Z. Liang, and S. You, "Bidirectional 3D quasi-recurrent neural network for hyperspectral image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2674–2688, 2021.

[43] Y. Cheng, K. Hu, J. Wu, H. Zhu, and X. Shao, "Auto-encoder quasi-recurrent neural networks for remaining useful life prediction of engineering systems," *IEEE/ASME Trans. Mechatronics*, early access, May 13, 2021, doi: 10.1109/TMECH.2021.3079729.

[44] C. Yang, W. Wang, X. Zhang, Q. Guo, T. Zhu, and Q. Ai, "A parallel electrical optimized load forecasting method based on quasi-recurrent neural network," in *IOP Conf. Ser., Earth Environ. Sci.*, vol. 696, no. 1, 2021, Art. no. 012040.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[46] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2016, pp. 19–34.

[47] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

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

[49] M. F. Baumgardner, L. L. Biehl, and D. A. Landgrebe. (Sep. 2015). *220 Band Aviris Hyperspectral Image Data Set: June 12, 1992 Indian Pine Test Site 3*. [Online]. Available: https://purr.purdue.edu/publications/1947/1

[50] Y. Wang, J. Peng, Q. Zhao, Y. Leung, X.-L. Zhao, and D. Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, Apr. 2018.

[51] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-i.I.d. Noise structure," 2017, *arXiv:1702.00098*.

Erting Pan received the B.S. degree in electrical engineering and its automation from Northeast Normal University, Changchun, China, in 2018, and the M.E. degree in electronic and communication engineering from Wuhan University, Wuhan, China, in 2020, where she is pursuing the Ph.D. degree with the Multispectral Vision Processing Laboratory, Electronic Information School.

Her 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, and the Ph.D. degree from the Huazhong University of Science and Technology (HUST), Wuhan, China, in 2003.

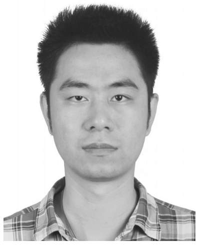
His general field of research is in signal and systems. His research projects include remote sensing of the Lidar and infrared, as well as infrared image processing, pattern recognition, and interface circuits to sensors and actuators. From 2004 to 2006, he was a Lecturer with the University of the West of England, Bristol, U.K. From 2006 to 2014, he was with the Wuhan National Laboratory for Optoelectronics, HUST, where he was a Professor of electronics. He is 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 HUST, in 2016.

From 2010 to 2012, he was a Software Engineer with the 722 Research Institute, China Shipbuilding Industry Corporation, Wuhan. He is an Associate 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 systems from the Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2015, respectively.

He is working as an Associate Professor with the Electronic Information School, Wuhan University. His research interests include infrared thermal imaging, machine learning, and computer vision.



Jun Huang (Member, IEEE) received the B.S. and Ph.D. degrees from the Department of Electronic and Information Engineering, Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

He is working as an Associate Professor with the Electronic Information School, Wuhan University. His main research interest is infrared image processing and infrared spectrum processing pattern recognition.

Jiayi Ma (Senior Member, IEEE) 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.

He is a Professor with the Electronic Information School, Wuhan University. He has authored or coauthored more than 200 refereed journal and conference papers, including the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), the IEEE TRANSACTIONS ON IMAGE PROCESSING (TIP), *International Journal of Computer Vision* (IJCV), IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE International Conference on Computer Vision (ICCV), European Conference on Computer Vision (ECCV), and so on. His research interests include computer vision, machine learning, and robotics. He has been identified in the 2019–2021 Highly Cited Researcher lists from the Web of Science Group.

Dr. Ma is an Area Editor of *Information Fusion* and an Associate Editor of *Neurocomputing*, *Sensors*, and *Entropy*.