

MAC-Net: Model Aided Nonlocal Neural Network for Hyperspectral Image Denoising

Fengchao Xiong, *Member, IEEE*, Jun Zhou, *Senior Member, IEEE*, Qinling Zhao, *Member, IEEE*, Jianfeng Lu, *Member, IEEE* and Yuntao Qian, *Senior Member, IEEE*

Abstract—Hyperspectral image (HSI) denoising is an ill-posed inverse problem. The underlying physical model is always important to tackle this problem, which is unfortunately ignored by most of the current deep learning (DL)-based methods, producing poor denoising performance. To address this issue, this paper introduces an end-to-end model aided nonlocal neural network (MAC-Net) which simultaneously takes the spectral low-rank model and spatial deep prior into account for HSI noise reduction. Specifically, motivated by the success of the spectral low-rank model in depicting the strong spectral correlations and the nonlocal similarity prior in capturing spatial long-range dependencies, we first build a spectral low-rank model and then integrate a nonlocal U-Net into the model. In this way, we obtain a hybrid model-based and DL-based HSI denoising method where the spatial local and nonlocal multi-scale and spectral low-rank structures are effectively exploited. After that, we cast the optimization and denoising procedure of the hybrid method as a forward process of a neural network and introduce a set of learnable modules to yield our MAC-Net. Compared with traditional model-based methods, our MAC-Net overcomes the difficulties of accurate modeling thanks to the strong learning and representation ability of DL. Unlike most “black-box” DL-based methods, the spectral low-rank model is beneficial to increase the generalization ability of the network and decrease the requirement of training samples. Experimental results on the natural and remote sensing HSIs show that MAC-Net achieves state-of-the-art performance over both model-based and DL-based methods. The source code and data of this article will be made publicly available at <https://github.com/bearshng/mac-net> for reproducible research.

Index Terms—Hyperspectral image denoising, model-based neural network, low-rank representation, nonlocal representation.

I. INTRODUCTION

Hyperspectral image (HSI) has enabled many practical applications for example medical analysis [1], urban planning [2], object tracking [3], thanks to its material identification ability enabled by numerous contiguous narrow bands. Because of the environment interference, sensor defect,

This work was supported in part by Jiangsu Provincial Natural Science Foundation of China under Grant BK20200466 and BK20190448, and the National Natural Science Foundation of China under Grant 62002169, 61902185 and 62071421, the National Key Research and Development Program of China under Grant 2018YFB0505000 and 2018AAA0100500 and “111” Program (No. B13022).

F. Xiong, Q. Zhao and J. Lu are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, P.R. China (Corresponding author: F. Xiong; fxiong@njust.edu.cn).

J. Zhou is with the School of Information and Communication Technology, Griffith University, Nathan, Australia.

Y. Qian is with the College of Computer Science, Zhejiang University, Hangzhou 310027, China.

data transmission, and other factors, the captured HSI is unavoidably contaminated by noises [4]. As a crucial and indispensable preprocessing step, hyperspectral image denoising is becoming increasingly important for high-quality imaging and has attracted extensive attention, recently [5]–[7].

HSI denoising is an ill-posed inverse problem recovering the underlying clean HSIs from noisy observation and its forward process is the formulation of HSIs. Model-based methods rely on the prior knowledge of image at hand and explicitly introduce various mathematical models depicting the physical properties of underlying clean HSIs and the observation process to carry out denoising. Models can be formulated in various forms and have been gradually improving over the past years from sparsity [8]–[11] and low-rankness [12]–[14] to local/nonlocal similarity [15]–[17] and so on. Model-based methods are interpretable, have high generation ability, and don’t heavily rely on the training samples. However, model-based methods often formulate denoising as an optimization problem which typically requires time-consuming numerical iteration and exhausting parameter twisting to produce high-quality denoising. Moreover, it is important to understand that there is no such a “correct model” that can accurately relate the observed HSIs and the desired clean HSIs, especially in complex scenarios. The absence of an accurate model often leads to denoising failure and also limits the applicability of model-based methods. Therefore, it is always a difficult and important task to establish a flexible and accurate mathematical model.

Thanks to the high representation and strong learning ability, deep learning (DL)-based methods have also been applied in HSI denoising. Instead of principled hand-crafted models, DL-based methods parametrize the model with learnable network architectures such as spectral-spatial attention [18], spectral-spatial convolution [19], [20] and residual learning [21] and introduce various loss functions [20] to directly learn the nonlinear mapping from noisy HSIs to clean HSIs. Compared with model-based methods, data-driven DL methods are independent of the physical models. Their architectures are generic and can handle noises even in scenarios where accurate noise models are unknown when provided with enough samples. The drawbacks of data-driven DL-based HSI denoising are twofold. On one hand, without the guidance of physical models, DL-based methods are black-boxes and there is still no human-interpretable way to understand their denoising mechanism. On the other hand, their overparameterized architectures always demand large scale of noisy-clean HSI pairs to train on, which are costly to acquire. As a result,

the interpretability, flexibility, and reliability of DL-based HSI denoising are actually a matter of concern [22].

As mentioned above, both model-based and DL-based methods have their merits and also defects. As far as we know, how to effectively integrate the advantages of the two methods while circumventing their disadvantages for HSI denoising remains to be explored. To fill the gap, in this paper, we propose a Model Aided nonlocal neural Network (MAC-Net) to bridge model-based and DL-based methods for HSI denoising. Fig. 1 illustrates the framework of our MAC-Net. Specifically, motivated by the success of spectral low-rank representation in spectral correlation exploitation, we first establish a low-rank model which projects the observed HSI into a much lower subspace spanned by a few spectra. In this way, the spectral low-rank prior of HSIs is explicitly depicted. Moreover, the nonlocal similarity prior can capture the long-range dependencies among different spatial contents of an HSI and always improves HSI denoising performance [23], [24]. To this end, we then embed a nonlocal attention block into U-Net to construct a nonlocal convolutional neural network and then integrate it into the spectral low-rank model, formulating a hybrid model-based and DL-based method. The nonlocal attention block is mathematically equivalent to dot-product attention [25], [26] but has relatively high memory and computational efficiency, allowing for more flexible integration into the high-resolution part of the network. After that, we cast the optimization step of the hybrid method as a forward process of a neural network to yield our model aided neural network for HSI denoising. Finally, all the parameters of MAC-Net are learned by end-to-end training.

Our MAC-Net has the hybrid advantages of model-based and DL-based methods. On one hand, the spectral low-rank model reasonably allows our MAC-Net to more effectively capture the strong spectral correlation and exploit the established benefits of model-based methods such as high interpretability, superior generalization ability, and less requirement of training samples. On the other hand, the nonlocal U-Net can capture the local and nonlocal multi-scale spatial structure. Data-driven learning also greatly helps MAC-Net enhance the original spectral low-rank model and reduce the overall model error induced by inaccurate model construction. As a result, our MAC-Net has the virtue of strong representation ability, superior generalization ability, and high denoising capacity, especially in complex environments. Experimental results on synthetic data and real-world HSIs show that our MAC-Net achieves convincing improvements over many state-of-the-art model-based and DL-based methods.

The contributions of this paper can be summarized as follows: 1) We combine advantages of model-based and DL-based methods to develop a spectral low-rank model aided neural network with high denoising capacity. 2) We introduce a lightweight nonlocal attention block to more efficiently utilize the nonlocal dependencies for improved denoising.

This paper is organized as follows: Section II briefly reviews recent works on model-based and DL-based HSI denoising as well as model-based deep learning. In Section III, we derive the network architecture of MAC-Net and present its learning and implementation details. Section IV reports the

experimental results on both synthetic and real-world data and compare them against several other competing approaches. This paper concludes in Section V.

II. RELATED WORK

This section briefly introduces recent works concerning model-based and deep learning-based HSI denoising methods and also the model-based deep learning.

1) Model-based HSI Denoising: Each band of the hyperspectral image can be regarded as a gray-level image and state-of-the-art denoising methods such as block matching 3D filtering (BM3D) [27] and K singular value decomposition (K-SVD) [28] can be independently performed for HSI denoising in a band-wise manner. Band-wise methods ignore the spectral-spatial correlations among bands and thereby achieve inferior denoising. By modeling an HSI in a 3D cube [29] or with tensors [30], spectral-spatial methods jointly exploit the spatial and spectral information and are more capable of removing noise.

Natural HSI has many physical properties such as sparsity, low-rankness, and nonlocal self-similarity, which have induced many state-of-the-art methods in HSI denoising. Sparse representation-based models assume clean HSI resides in a subspace spanned by a few atoms in a dictionary. By sparsely encoding an HSI over fixed dictionaries such as 3D wavelet and 3D discrete cosine transformation (DCT) [31] or adaptively learned dictionaries [9], [10], the noise component can be removed to a large extent. In addition to sparse models, the high spatial similarity and strong spectral correlation also imply the low-rank characteristic of HSIs. Low-rank matrix approximation methods such as matrix factorization [5], [32] and matrix rank minimization [33] emerge as useful tools to recover the underlying low-rank structure of HSIs. The estimation of matrix rank is an NP-hard problem and is typically approximately estimated by the nuclear norm. L_1 norm-based nuclear norm is inclined to punish more on components with larger singular values, which however are more likely related to the predominant edges of HSIs, resulting in detail loss in denoised HSIs. In contrast, nonconvex variants of the nuclear norm, for example, weighted Schatten p -norm [34] and normalized ϵ -penalty [35] can better recover the low-rank structure of HSIs and thereby enhance HSI denoising performance.

Matrix-based methods unavoidably destroy the spectral-spatial structural correlation of HSIs when converting a 3D HSI into a 2D matrix. For this reason, tensor-based methods have attracted intensive attention these years [36]–[39] thanks to their superior advantages of modeling HSIs without information loss. Accordingly, Peng *et al.* constructed a nonlocal tensor dictionary learning model to depict the nonlocal spatial similarity and global spectral correlation of HSIs [40]. Moreover, many pieces of research focus on defining different types of tensor nuclear norm [36], [41] to measure the intrinsic correlations across various tensor modes so that the clean HSI can be more accurately extracted from the noisy observation.

Besides low-rankness and sparsity priors, other HSI priors such as local similarity [42], nonlocal similarity [43] and joint

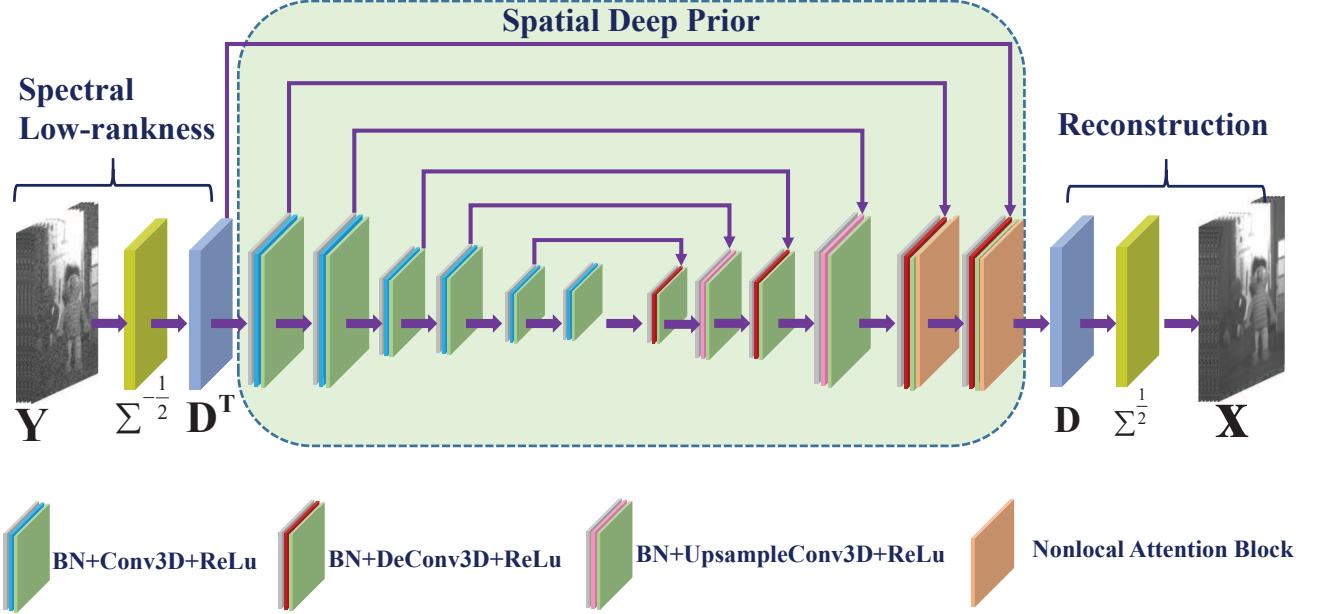


Fig. 1. Illustration of our MAC-Net. MAC-Net is a hybrid model-based and data-driven DL-based method. The spectral low-rankness module explicitly describes strong spectral correlations and the nonlocal 3D U-Net module implicitly characterizes the spatial structures. As a result, MAC-Net not only overcomes the difficulties of accurately modeling the true physical properties of HSIs because of the strong representation ability of deep learning but also has superior generalization ability thanks to the embedded physical model.

use of multiple priors [23], [24] provide stronger image models for HSI denoising. For example, [24] and [23] integrated the spatial self-similarity prior with spectral low-rank prior and achieved fast and impressive denoising. Wang *et al.* [44] embedded the spectral-spatial total variation regularization into low-rank Tucker decomposition to characterize the spectral-spatial piecewise-smooth and low-rank structures of HSIs.

Model-based methods are easily understandable and do not heavily rely on data to handle noise despite the requirement of a small number of data to estimate unknown model parameters. However, simplified hand-crafted priors possibly are not able to truly capture the actual diversity and complexity of practical HSI denoising, limiting the applicability of model-based methods.

2) *Deep Learning-based HSI Denoising:* Inspired by the noticeable success of deep learning in high-level tasks [45], DL-based methods have also been applied in HSI denoising. DL-based methods learn an explicit mapping between the noisy HSIs and clean HSIs from large-scale datasets, also known as deep prior. How to effectively preserve the spectral-spatial structure of HSIs is a key issue in DL-based HSI denoising. Early work [20] and [21] used multiple channels of 2D convolutions to exploit the spectral-spatial structure correlation of HSIs but has high computational complexity. Further, Shi *et al.* [18] took advantage of 3D atrous convolution to simultaneously exploit the spectral-spatial information of HSIs while increasing the receptive field. Dong *et al.* [19] decomposed 3D convolution into 2D spatial convolution and 1D spectral convolution to reduce the number of parameters and computational complexity. Alternatively, Wei *et al.* [46] combined 3D convolution with quasi-recurrent pooling to more effectively capture the global spectral correlation. In addition to spectral-spatial representation, multiscale feature

expression and residual learning were also employed in [21] and [18] to enhance denoising performance. Mafeijet *et al.* [47] took the noise-level map as an input to train a single network that is adaptive to multilevel noise. Yuan *et al.* [48] developed a noise intensity estimation block to estimate the noise and used it to guide the subsequent denoising, increasing the adaptivity of the network to different datasets.

Though data-driven DL-based methods can recover the clean HSIs by training the networks on a large number of training data, the domain knowledge of HSIs such as the observation model and underlying characteristics of HSIs are not fully utilized. As a result, complicated architectures and massive training data are demanded to cover the real-world complexity of HSI denoising. The “black-box” nature of data-driven DL-based methods makes it difficult to understand their denoising mechanism and provide performance reliability, too.

3) *Model-based Deep Learning:* Model-based deep learning integrates the domain knowledge into the network in the form of an established physical model while preserving the learning capacity of DL-based methods from data. Such a data-driven model-based scheme inherits the hybrid advantages of both methods such as superior interpretability, high generalization ability, and strong representation ability. With above virtue, model-based DL have been widely applied in hyperspectral unmixing [49], [50] and super-resolution [51], [52] and achieved impressive performance. Following this line, we aim to combine the spectral low-rank model with spatial deep prior of HSIs and construct a spectral low-rank model aided nonlocal network to achieve more effective HSI denoising.

III. MAC-NET FOR HSI DENOISING

In this section, we introduce the details of the proposed MAC-Net and address its implementation details.

A. Model Formulation

Let \mathbf{Y} be a noisy observed HSI containing N pixels and L bands. Under the assumption that the clean HSI is corrupted by additive Gaussian noise, we have the following observational model

$$\mathbf{Y} = \mathbf{X} + \mathbf{E} \quad (1)$$

where $\mathbf{X} \in \mathbb{R}^{L \times N}$ is the underlying clean HSI and $\mathbf{E} \in \mathbb{R}^{L \times N}$ measures the noise item.

Generally, Eq. (1) is an ill-posed inverse problem that separates underlying clean \mathbf{X} from noisy observation \mathbf{Y} . Model-based methods tackle this problem under the premise that there are a set of mathematical properties that the HSI is believed to follow. Because of the strong correlation among spectral bands, HSIs have spectral low-rankness. Mathematically, low-rankness implies there are many zero entries in the singular vector of a matrix. Fig. 2 presents the singular values of the unfolding matrix along the spectral dimension of Washington DC HSI. As can be seen from the figure, the singular values show an obvious decaying trend in the spectral domain and most values are zero or nearly zero. This means that high-dimensional HSI \mathbf{X} essentially resides in a K -dimensional subspace with $K \ll L$.

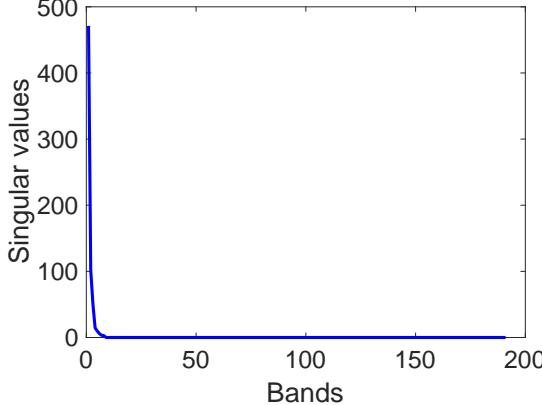


Fig. 2. Singular value curve of the unfolding matrix of a clean HSI along spectral dimension.

Following this line, a straightforward way to model the spectral low-rank structure of \mathbf{X} is to decompose it into the product of two factor matrices, with one being the basis matrix spanning the subspace and the other one being the coefficient matrix. Accordingly, the model in Eq. (1) can be reformulated as

$$\mathbf{Y} = \mathbf{DW} + \mathbf{E} \quad (2)$$

Here, $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_k] \in \mathbb{R}^{L \times K}$ denotes K basis in spectral domain with $K \ll L$ to capture the spectral redundancy and \mathbf{W} is the coefficients matrix of \mathbf{X} with respect to \mathbf{D} . K can be identified by hyperspectral signal subspace identification by minimum error (HySime) method [53]. In our implementation, we set \mathbf{D} as an orthogonal matrix, i.e., $\mathbf{D}^T \mathbf{D} = \mathbf{I}$ where \mathbf{I} is

an identity matrix as it can reduce the complexity of the model and accelerate the numerical optimization [54].

Assuming the noise follows independent and identically distribution (i.i.d), under the framework of the *maximum a posteriori* (MAP) estimation framework, the denoising model in Eq. (2) can be reformulated as the following constrained least square problem:

$$\min_{\mathbf{W}, \mathbf{D}} \frac{1}{2} \|\mathbf{Y} - \mathbf{DW}\|_F^2 \quad \text{s.t.} \quad \mathbf{D}^T \mathbf{D} = \mathbf{I} \quad (3)$$

The spectral low-rank prior is not sufficient by itself to capture the physical characteristics of clean HSIs, possibly leading to the loss of image details. In addition to spectral low-rank prior, there are also many models depicting the spatial properties of HSIs, such as sparsity, low-rankness, and local/nonlocal similarity. However, all of these priors are hand-crafted and a simplified approximation of the real spatial property of underlying clean HSIs. Compared with explicitly hand-crafted modeling, data-driven deep neural networks directly learn the nonlinear mapping from data. The network implicitly embodies the prior knowledge inside the data, for example, denoising prior. Incorporating the deep prior into the optimization objective function, the model in Eq. (3) can be rewritten as

$$\min_{\mathbf{W}, \mathbf{D}} \frac{1}{2} \|\mathbf{Y} - \mathbf{DW}\|_F^2 + \lambda f(\mathbf{W}) \quad \text{s.t.} \quad \mathbf{D}^T \mathbf{D} = \mathbf{I} \quad (4)$$

where $f(\mathbf{W})$ represents the implicit spatial priors embodied by deep neural network. Here, we add regularization on \mathbf{W} instead of \mathbf{X} . On one hand, the original noisy HSI is spectrally redundant. It is expected to obtain more accurate spatial structure expression from compact subspace so as to promote HSI denoising. On the other hand, \mathbf{W} is much smaller than \mathbf{X} and the network needs to estimate a small number of variables, reducing computational complexity during processing.

B. Algorithm Optimization

The objective function in Eq. (4) has two unknown variables and we solve them one by one as below. The optimization problem of \mathbf{D} is

$$\begin{aligned} \hat{\mathbf{D}} &= \arg \min_{\mathbf{D}^T \mathbf{D} = \mathbf{I}} \|\mathbf{Y} - \mathbf{DW}\|_F^2 \\ &= \arg \min_{\mathbf{D}^T \mathbf{D} = \mathbf{I}} \text{tr}[(\mathbf{Y} - \mathbf{DW})^T (\mathbf{Y} - \mathbf{DW})] \\ &= \arg \max_{\mathbf{D}^T \mathbf{D} = \mathbf{I}} \text{tr}(\mathbf{Y} \mathbf{W}^T \mathbf{D}^T) \\ &= \mathbf{U} \mathbf{V}^T \end{aligned} \quad (5)$$

where \mathbf{U} and \mathbf{V} respectively contains the left and right singular vectors of $\mathbf{Y} \mathbf{W}^T$. The subproblem of solving \mathbf{W} is

$$\min_{\mathbf{W}} \frac{1}{2} \|\mathbf{Y} - \mathbf{DW}\|_F^2 + \lambda f(\mathbf{W}) \quad (6)$$

As \mathbf{D} is an orthogonal matrix, solving Eq. (6) amounts to

$$\min_{\mathbf{W}} \frac{1}{2} \|\mathbf{D}^T \mathbf{Y} - \mathbf{W}\| + f(\mathbf{W}) \quad (7)$$

In Eq. (7), $\mathbf{D}^T \mathbf{Y}$ can be seen as the projected image of \mathbf{Y} with respect to the basis set \mathbf{D} . From this point of view, Eq. (7) can be considered as a denoising problem and $f(\cdot)$ is a denoiser.

The solution of \mathbf{W} is thus can be obtained by $\widehat{\mathbf{W}} = f(\widehat{\mathbf{D}}^T \mathbf{Y})$ where λ is absorbed into the network.

Iteratively running Eq. (5) and Eq. (7) for numbers of times yields the solution of \mathbf{D} and \mathbf{W} , which is time-consuming. Alternatively, as in [23] and [24], we propose to relax the optimization problem of \mathbf{D} as

$$\{\widehat{\mathbf{D}}, \widehat{\mathbf{W}}\} = \arg \min_{\mathbf{W}, \mathbf{D}^T \mathbf{D} = \mathbf{I}} \|\mathbf{Y} - \mathbf{DW}\|_F^2 \quad (8)$$

whose close form solution can be obtained by performing the truncated SVD of \mathbf{Y} , i.e., $\mathbf{Y}_K = \mathbf{U}_K \Sigma_K \mathbf{V}_K^T$ and setting $\widehat{\mathbf{D}} = \mathbf{U}_K$, $\widehat{\mathbf{W}} = \widehat{\mathbf{D}}^T \mathbf{Y}$. After that $\widehat{\mathbf{W}}$ is further denoised by $f(\cdot)$ to yield $\widehat{\mathbf{W}}$. In this way, the solution of \mathbf{D} only depends on \mathbf{Y} and there is no need to iteratively optimize \mathbf{W} and \mathbf{D} . Finally, the clean image can be recovered by $\widehat{\mathbf{X}} = \widehat{\mathbf{D}} \widehat{\mathbf{W}}$.

As pointed in previous research [9], in real-world HSIs, the noise intensity in the same band is identical but varies in different bands. That is to say the noises are spatially i.i.d but spectrally non-i.i.d and $\mathbf{E} \sim \mathcal{N}(0, \Sigma)$. Σ is a covariance matrix size of $L \times L$ and can be generally assumed positive definite and estimated by Hysime [53]. Provided with Σ , we can convert non-i.i.d noise to i.i.d noise with

$$\mathbf{Y}' = \Sigma^{-\frac{1}{2}} \mathbf{Y} \quad (9)$$

where $\Sigma^{-\frac{1}{2}}$ is the square root of Σ^{-1} and Σ^{-1} is the inverse matrix of Σ . Accordingly, the model in Eq. (1) can be reformulated as

$$\mathbf{Y}' = \Sigma^{-\frac{1}{2}} \mathbf{X} + \Sigma^{-\frac{1}{2}} \mathbf{E} \quad (10)$$

Then \mathbf{D} and \mathbf{W} can be computed by Eq. (8) and Eq. (7) and the clean image is recovered by $\widehat{\mathbf{X}} = \Sigma^{\frac{1}{2}} \widehat{\mathbf{D}} \widehat{\mathbf{W}}$.

Overall, the denoising procedure can be summarized in Algorithm 1. Algorithm 1 contains five steps and all the steps are carried out in sequence with no iterations, making it fast for denoising. However, as a drawback of relaxed optimization of \mathbf{D} , there is no guarantee that the overall procedure seeks to find an estimated HSI closest to the ground truth HSI with minimum mean square error because of no feedback information from the subsequent steps. From another point of view, steps 1-5 can be alternatively regarded as a forward process of a DNN. Thanks to end-to-end training, the reconstruction error between estimated HSIs and ground-truth HSIs can be backpropagated to guide the network learning. To this end, we construct a deep neural network based on the denoising procedure in Algorithm 1 as will be presented in the next subsection.

C. MAC-Net

Based on the denoising procedure in Algorithm 1, we construct a model aided nonlocal convolutional neural network (MAC-Net). Fig. 1 illustrates our MAC-Net. MAC-Net is an end-to-end network composed of three parts, i.e., spectral low-rank module, spatial deep prior module (SDPM), and reconstruction module. Guided by the spectral low-rank prior, the spectral low-rankness module is related to steps 1-3 in Algorithm 1 and aims to model the strong correlations among bands. SDPM takes advantage of nonlocal 3D residual U-Net to exploit the spatial denoising prior of HSIs and can be

Algorithm 1 Our denoising procedure

Input: Noisy hyperspectral image \mathbf{Y} .

Output: Recovered clean hyperspectral image \mathbf{X} .

- 1: Estimate noise correlation matrix Σ with Hysime.
 - 2: Convert non-i.i.d noise to i.i.d noise with $\mathbf{Y}' = \Sigma^{-\frac{1}{2}} \mathbf{Y}$.
 - 3: Estimate the subspace number K and perform trunctated SVD of \mathbf{Y}' , i.e., $\mathbf{Y}'_K = \mathbf{U}_K \Sigma_K \mathbf{V}_K^T$ and set $\widehat{\mathbf{D}} = \mathbf{U}_K$.
 - 4: Denoise the projected image with a deep neural network to produce $\widehat{\mathbf{W}}$, i.e., $\widehat{\mathbf{W}} = f(\widehat{\mathbf{D}}^T \mathbf{Y}')$.
 - 5: Output the estimated clean hyperspectral image by $\widehat{\mathbf{X}} = \Sigma^{\frac{1}{2}} \widehat{\mathbf{D}} \widehat{\mathbf{W}}$.
-

connected with step 4. According to step 5, the reconstruction module maps the denoised projected image into the original spectral space to yield denoised HSIs.

Generally speaking, our MAC-Net is a hybrid model-based and DL-based method. On one hand, our network is induced from the model in Eq. (4), making it convey more interpretability and provable performance guarantees. For example, all the modules can be connected with the optimization steps in Algorithm 1. On the other hand, our MAC-Net also supports discriminatively learning the spatial priors from data via deep learning techniques, overcoming the limitation of predefined hand-crafted priors. Moreover, the forward process of our MAC-Net can be alternatively interpreted as minimizing the following objective function

$$\min_{\mathbf{D}, \Theta} \frac{1}{2} \|\Sigma^{-\frac{1}{2}} \mathbf{Y} - \mathbf{D} f(\cdot)\|_F^2 \quad (11)$$

where Θ is the parameter set of $f(\cdot)$. $\Sigma^{-\frac{1}{2}}$ transforms the observed HSI \mathbf{Y} into a new spectral space so that the noise follows Gaussian i.i.d distribution and \mathbf{D} encapsulates the spectral low-rank structure of HSIs. Fed with $\mathbf{D}^T \Sigma^{-\frac{1}{2}} \mathbf{Y}$, $f(\cdot)$ tries to generate an image that matches the underlying clean HSI in the latent spectral subspace.

As we know, without the regularization term of $f(\cdot)$, the solution of \mathbf{W} is $\mathbf{D}^T \Sigma^{-\frac{1}{2}} \mathbf{Y}$ based on the spectral low-rank model in Eq. (3). From this point of view, the spectral low-rank model can be explained as a partially known model. SDPM module performs denoising on the projected image $\mathbf{D}^T \Sigma^{-\frac{1}{2}} \mathbf{Y}$ and serves as a complementary to the model. Thanks to the advantages of learning complex mappings from data, the representation ability of the model are strengthened, facilitating improving the denoising capability in complex environments. As a reward from the physical model, the joint working with the spectral low-rank model also helps our network less depend on massive amounts of training data and increase the interpretability.

Both spectral low-rankness and reconstruction modules can be considered as a sequence of 1×1 convolutions along the spectral dimension. In the following, we introduce the detailed structure of our SDPM. We adopt the U-Net architecture [55] to implement SDPM considering its multi-scale representation ability and noticeable effectiveness in image denoising [56]. As presented in Table I, SDPM consists of one feature

¹“BN”, “UpConv3D” and “NAB” represent batch normalization, upsampleconv3d and efficient nonlocal attention block, respectively.

TABLE I
NETWORK CONFIGURATION OF OUR SPATIAL DEEP PRIOR MODULE¹

Layer	Configuration	Stride	Output size
Extractor	BN+Conv3D+ReLU	1, 1, 1	$C \times H \times W \times K$
Encoder	BN+Conv3D+ReLU	1, 1, 1	$C \times H \times W \times K$
	BN+Conv3D+ReLU	2, 2, 1	$2C \times \frac{H}{2} \times \frac{W}{2} \times K$
	BN+Conv3D+ReLU	1, 1, 1	$2C \times \frac{H}{2} \times \frac{W}{2} \times K$
	BN+Conv3D+ReLU	2, 2, 1	$4C \times \frac{H}{4} \times \frac{W}{4} \times K$
	BN+Conv3D+ReLU	1, 1, 1	$4C \times \frac{H}{4} \times \frac{W}{4} \times K$
Decoder	BN+DeConv3D+ReLU	1, 1, 1	$4C \times \frac{H}{4} \times \frac{W}{4} \times K$
	BN+UpConv3D+ReLU	$\frac{1}{2}, \frac{1}{2}, 1$	$2C \times \frac{H}{2} \times \frac{W}{2} \times K$
	BN+DeConv3D+ReLU	1, 1, 1	$2C \times \frac{H}{2} \times \frac{W}{2} \times K$
	BN+UpConv3D+ReLU	$\frac{1}{2}, \frac{1}{2}, 1$	$C \times H \times W \times K$
	BN+DeConv3D+ReLU+NAB	1, 1, 1	$C \times H \times W \times K$
Reconstruction	BN+DeConv3D+ReLU+NAB	1, 1, 1	$C \times H \times W \times K$

extraction layer, five pairs of encoder-decoder layers, and one reconstruction layer.

The feature extraction layer aims to extract shallow features from input projected HSI with 3D convolutions. Although orthogonal projection tends to uncorrelate the HSI in the subspace, this does not mean that there are no dependencies between channels projected HSI [13]. The main reason is that orthogonal projections are global for all the pixels while channel correlations are distributed in local cubes. Therefore, we employ 3D convolutions to simultaneously capture the local spatial and channel statistical dependency among $\mathbf{D}^T \mathbf{Y}'$. In addition, it is known that different HSIs may reside in different spectral subspaces with varied dimensions. 3D convolution allows our network to adaptively process the projected HSIs with arbitrary channels as it is independent of the channel number of input images. By cascading multiple down-scale blocks, the encoder gradually decreases the spatial resolution in half and doubles the feature dimension, yielding multi-scale spatial representation. Each down-scale block has a configuration of two $3 \times 3 \times 3$ 3D convolutions whose stride are respectively set as 2 and 1.

The decoder recovers the resolution of the feature maps through a sequence of up-scale and deconvolutional blocks. Each up-scale block doubles the spatial resolution of the feature map and reduces the number of channels by half. Moreover, a residual connection is established between the encoder and decoder with the same resolution to propagate the features in shallow layers to deeper layers, facilitating preserving information and avoiding gradient vanishing during training. The reconstruction layer employs the residual connection and 3D deconvolution to recover underlying clean projected HSIs. Additionally, as in [57], batch normalization and ReLU activation are also adopted in all the layers.

3D convolutions can capture the local correlation. Apart from local correlation, many patches at different locations or scales in the images also have many similar patterns, i.e., nonlocal similarity. Nonlocal similarity can simultaneously capture the local and long-range spatial similarity of HSIs so that local and nonlocal patches can be jointly denoised. For this reason, considering the nonlocal similarity property of HSIs

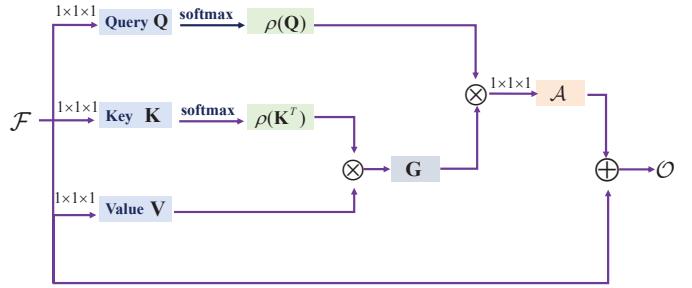


Fig. 3. Illustration of the nonlocal attention block. The nonlocal attention can effectively capture the long-range dependencies among pixels and has dramatically less computational complexity than commonly used nonlocal block in [26].

always improves HSI denoising performance. For example, He *et. al* [23] grouped similar patches along the spatial dimension and employed the low-rank matrix approximation to exploit the non-local similarity of HSIs. Similarly, Zhuang *et. al* [24] took advantage of joint sparse representation to capture the non-local similarity. All these methods have achieved the state-of-the-art denoising performance. To this end, an efficient nonlocal attention block is set in the last layer of the encoder and the reconstruction layer to explore the long-range dependencies between different locations. Based on dot-product attention, nonlocal block [26] is a typical choice for computer vision tasks. Given an input feature map $\mathcal{F} \in \mathbb{R}^{c \times h \times w \times k}$, nonlocal block requires the calculation of the pairwise similarities between pixels before yielding per-pixel attention maps and occupies $O(n^2)$ memory and $O(d_k n^2)$ computation where $n = chw$ and d_k is the dimensionality of the keys. This threatens challenges for memory and computational load, especially for images with large size.

Instead, we adopt an efficient nonlocal attention block (NAB) [58] to capture nonlocal dependencies among HSIs. NAB first forms the global contexts by aggregating the value matrix with the key matrix, which is then multiplied with the query matrix to yield the attention map. Specifically, as shown in Fig. 3, the input feature maps $\mathcal{F} \in \mathbb{R}^{c \times w \times h \times d}$ are firstly filtered by three $1 \times 1 \times 1$ 3D convolutions and reshaped into matrix form to produce the query ($\mathbf{Q} \in \mathbb{R}^{chw \times d_k}$), key ($\mathbf{K} \in \mathbb{R}^{chw \times d_k}$) and value ($\mathbf{V} \in \mathbb{R}^{chw \times v_k}$) matrices. After that the query matrix and key matrix are normalized by a softmax function, resulting in scaled key and query matrix, i.e., $\rho(\mathbf{Q})$ and $\rho(\mathbf{K})$. The attention is then yielded by

$$\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \rho(\mathbf{Q})(\rho(\mathbf{K})^T \mathbf{V}) \quad (12)$$

We further apply a $1 \times 1 \times 1$ 3D convolution on $\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V})$ to restore the dimensionality to d and add with input \mathcal{F} to form a residual structure, i.e.,

$$\mathcal{O} = \mathcal{F} + \text{cov3d}(\text{Att}(\mathbf{Q}, \mathbf{K}, \mathbf{V})) \quad (13)$$

Unlike the nonlocal block, NAB has substantial efficiency as it greatly reduces the computational complexity to $O(d^2n)$ and memory load to $O(dn + d^2)$. This makes it highly applicable for high-resolution images and yields very comparable performance even higher performance in many computer vision tasks [58]. Therefore, we set the NAB in the last encoding layer and the reconstruction layer whose feature maps have

the highest spatial resolution. As observed in [58], d_k and d_v has little impact on performance within a reasonable range. For simplicity, we set their values as $d_k = d_v = \lfloor \frac{d}{2} \rfloor$.

D. Network Implementation

Loss function: Given a training set of noisy-clean HSI pairs, the training loss of our MAC-Net is defined as the Euclidean distance between the predicted HSI and ground truth, formulated as

$$\mathcal{L} = \|\text{MAC-Net}(\mathbf{Y}, \Theta) - \mathbf{X}\|_F^2 \quad (14)$$

where Θ is the parameter set of MAC-Net and \mathbf{X} is the ground truth HSI. Considering different HSIs have distinct spectral subspace and noise correlation, we directly use Σ and \mathbf{D} to assign the parameters of spectral low-rankness and reconstruction layers, leaving only SDPM to be learned. With such a setting, our network can simultaneously employ the internal data, i.e., the noisy HSI for adaptive learning of Σ and \mathbf{D} and external data, i.e., training set, to better capture the structure of individual HSIs. As a result, the denoising flexibility and capability of our network are enhanced.

Training details: We implemented our MAC-Net on the PyTorch platform and used two NVIDIA GeForce RTX 3090 GPUs to train the network through the ADAM algorithm. We set the number of epochs as 300. The initial learning rate was assigned as 5×10^{-3} and then reduced by a factor of 0.35 in every 80 epochs. nonlocal means was performed on transformed \mathbf{Y} , i.e., $\Sigma^{-\frac{1}{2}}\mathbf{Y}$ to better estimate the subspace.

IV. EXPERIMENTS

In this section, we comprehensively evaluate the denoising ability of the proposed MAC-Net on both synthetic data and real-world data. Moreover, throughout analysis and ablation study are also presented to show the generalization and learning capacity of our method.

A. Experiment Settings

Training dataset: Following the experimental setting in [46], the training set contains 100 HSIs selected from ICVL hyperspectral dataset². Each image is size of $1392 \times 1300 \times 31$. We carried out random flipping, cropping, and resizing to enrich the training set. After that, subimages size of $64 \times 64 \times 31$ are used to train the network.

Testing dataset: The testing dataset includes two parts, i.e., synthetic data and real-world data. The synthetic data contains 50 close-range HSIs selected from ICVL hyperspectral dataset and remote sensing Washington DC Mall HSI. The real-world data are remote sensing HSIs including HYDICE Urban HSI and AVIRIS Indian Pines HSI.

Compared methods: Twelve methods are selected as alternative methods to be compared with our method, including 8 model-based methods, i.e., BM4D [29], TDL [40], MTSNMF [9], LLRT [30], NGMeet [23], LRMR [59], LRTDTV [44] and FastHyDe [24] and 4 DL-based methods, i.e., DnCNN [57], HSI-SDeCNN [47], HSID-CNN [47],

QRNN3D [46]. DnCNN is performed in a band by band manner. We tried our best to manually tune the parameters of model-based methods to the optimal.

Evaluation indexes: We use peak signal-to-noise ratio (PSNR), structure similarity (SSIM), and spectral angle mapper (SAM) to quantitatively assess the denoising ability of all the methods. PSNR and SSIM respectively measure the quality of reconstruction of lossy compression and perceived changes in structural information. SAM depicts the spectral differences between clean and denoised HSIs. Smaller SAM and larger PSNR and SSIM imply better denoising. Moreover, we compare the restored HSI with clean HSI and noisy HSI to illustrate their differences for qualitative assessment.

Noise generation: As mentioned earlier, there exists spectrally non-i.i.d noise. To this end, different distributions of zero-mean Gaussian noise were added to each band to produce the noisy HSIs. The noise variances σ are randomly selected in three ranges, i.e., [0-15], [0-55], [0-95]. For these three cases, the exact variance is given for methods that require σ as an input, for example, DnCNN and BM4D. Moreover, we also design a blind case in which the noise variances are unknown but estimated by [60] under the observation of its superior accuracy and stability in noise level estimation.

B. Evaluation on ICVL HSIs

1) *Quantitative Evaluation:* Table II presents the denoising performance of all the competing methods regarding PSNR, SSIM and SAM indexes on 50 testing HSIs of the ICVL dataset with each size of $512 \times 512 \times 31$. The top two results are respectively highlighted with bold red and bold blue. In general, those methods that consider the spectrally non-i.i.d noises into consideration yield significantly better denoising, for example, BM4D, MTSNMF, LRMR, LRTDTV, FastHyDe, and QRNN3D, especially in the case of obvious noise differences among bands, i.e., $\sigma \in [0 - 55], [0 - 95]$. Among model-based methods, FastHyDe can better handle noise as it simultaneously considers the global spectral low-rankness and takes more advantage of the spatial sparsity with the state-of-the-art BM3D denoiser. In comparison to HSI-SDeCNN and HSID-CNN, QRNN3D is more capable of handling noises thanks to better depicting local spectral-spatial structure with 3D convolution and global correlation along the spectrum with quasi-recurrent pooling function. The ignorance of spectral correlation and spectral-spatial structure leads to poor denoising performance of band-wise Dn-CNN. Compared with the best model-based method, FastHyDe, the embedded spatial nonlocal convolutional subnetwork, i.e., SDPM, has more capacity of capturing the spatial structure of HSIs than hand-crafted priors thanks to data-driven training. Unlike “black-box” QRNN3D, the integration of the spectral low-rank model enhances the interpretability and generalization ability of MAC-Net and also reduces the number of parameters to be learned. Moreover, the nonlocal attention block is also helpful to capture the long-range dependencies among different contents of the HSI. For the above reasons, MAC-Net can better filter out the noises and obtain the highest denoising effectiveness for all the indexes and cases.

²Link:<http://icvl.cs.bgu.ac.il/hyperspectral/>

TABLE II
COMPARISON OF DIFFERENT METHODS ON 50 TESTING HSIS FROM ICVL DATASET. THE TOP TWO VALUES ARE MARKED **RED** AND **BLUE**.

σ	Index	Noisy	Sparse methods			Low-rank methods				DL methods					
			BM4D [29]	TDL [40]	MTSNMF [9]	LLRT [30]	NGMeet [23]	LRMR [59]	LRTDTV [44]	FastHyDe [24]	Dn-CNN [57]	HSI-SDe CNN [47]	HSID-CNN [21]	QRNN3D [46]	MAC-Net (Ours)
[0-15]	PSNR	33.47	45.12	38.84	45.62	46.17	39.42	41.49	42.17	47.85	41.92	41.06	38.08	43.36	49.19
	SSIM	0.6295	0.9753	0.8481	0.9562	0.9649	0.8647	0.9630	0.9717	0.9941	0.9596	0.9565	0.9669	0.9884	0.9980
	SAM	0.1622	0.0254	0.0829	0.0409	0.0319	0.0891	0.0268	0.0229	0.0163	0.0274	0.0329	0.0511	0.0336	0.0127
[0-55]	PSNR	21.67	38.21	29.45	37.57	37.19	30.48	34.34	39.69	41.74	37.90	35.66	32.55	40.15	43.02
	SSIM	0.2361	0.9217	0.5238	0.8569	0.8389	0.6748	0.8005	0.9625	0.9845	0.9293	0.8736	0.8421	0.9729	0.9925
	SAM	0.5014	0.0552	0.2409	0.1372	0.1020	0.2830	0.0408	0.0335	0.0344	0.0504	0.0602	0.0918	0.0382	0.0280
[0-95]	PSNR	16.97	35.27	25.40	34.20	32.49	27.20	30.71	38.12	39.31	34.65	32.45	29.12	37.66	40.17
	SSIM	0.1442	0.8764	0.3735	0.7998	0.7191	0.5506	0.6302	0.9539	0.9773	0.8442	0.7875	0.7049	0.9479	0.9837
	SAM	0.7199	0.0799	0.3641	0.2128	0.1687	0.4275	0.0589	0.0402	0.0447	0.1094	0.0876	0.1316	0.0468	0.0377
Blind	PSNR	19.99	37.00	27.71	36.25	34.42	28.89	33.14	39.14	40.95	36.82	34.20	31.31	39.32	41.62
	SSIM	0.2054	0.8970	0.4614	0.8340	0.7518	0.6074	0.7445	0.9598	0.9820	0.9094	0.8284	0.7804	0.9877	0.9884
	SAM	0.5829	0.0662	0.2953	0.1649	0.1467	0.3481	0.0489	0.0365	0.0393	0.0648	0.0723	0.1020	0.0409	0.0358

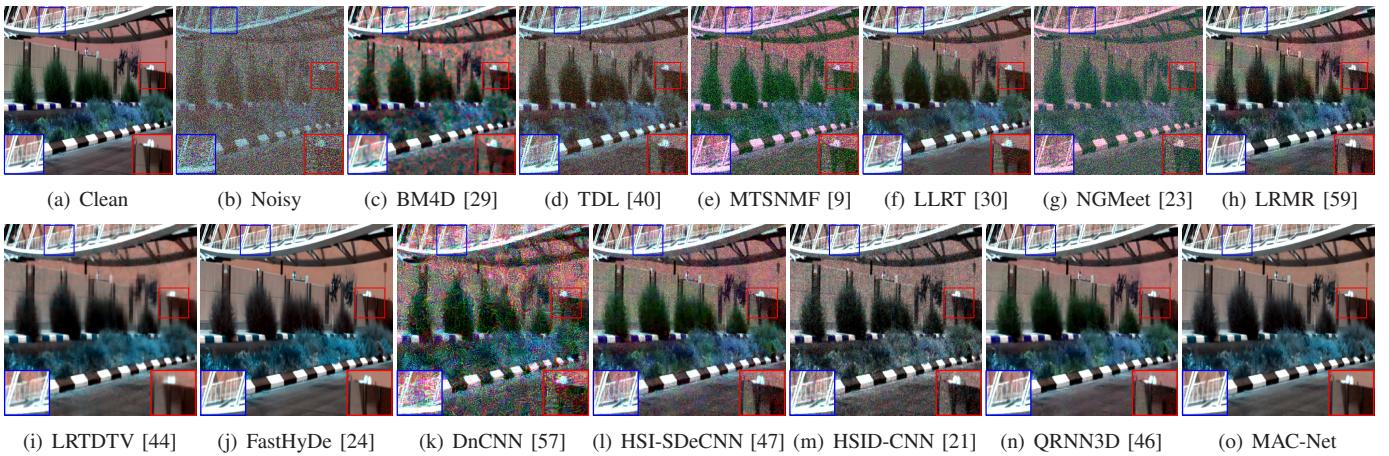


Fig. 4. Denoising results on the *gavyam_0823-0933* image with the noise variance in [0-95]. The false-color images were generated by combining bands 5, 18, 25. MAC-Net achieves the best visual results with less artifacts.

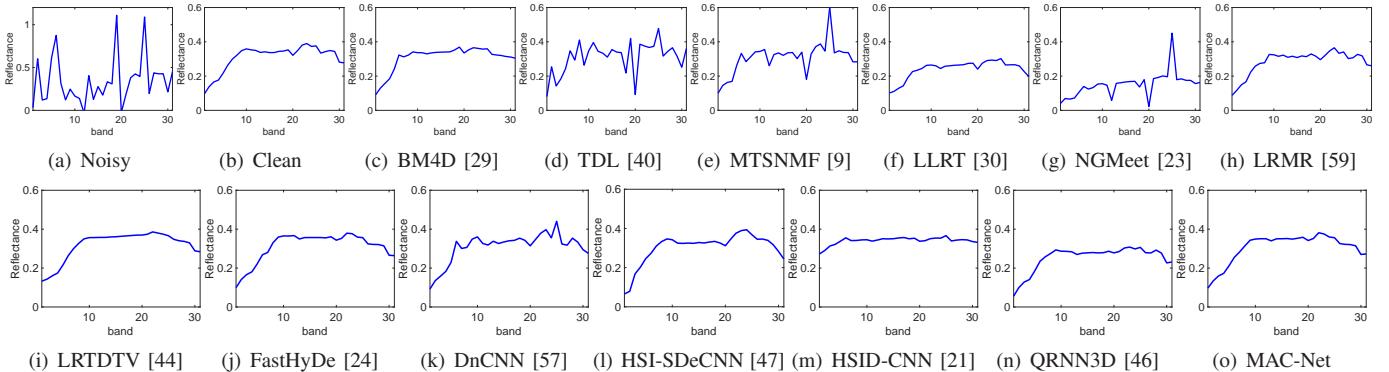


Fig. 5. Denoising results of pixel (451, 135) in *gavyam_0823-0933* HSI.

2) *Qualitative Evaluation:* In order to more comprehensively demonstrate the denoising effectiveness of all the methods, we further show a visual comparison on the false-color images generated by combining bands 5, 18, and 25 before and after denoising in Fig. 4. As a visual comparison shown in Fig. 4 (a) and Fig. 4 (b), heavy noise makes it very difficult to tell the content of the image, threatening a great challenge for a denoiser. Except for LRTDTV and FastHyDe, other model-based methods provide lower denoising quality, either failing to reduce all the noises, see, Fig. 4 (d)-(h), or

removing noise at the sacrifice of important details, see Fig. 4 (c). As 3D data, the hyperspectral image has characteristics of the spectral-spatial structure, global spectral correlation, and local/nonlocal spatial interactions, and more. This prior information is very conducive to robust denoising and thereby enhances QRNN3D, LRTDTV, FastHyDe, and QRNN3D with more promising denoising performance. Band-wise DnCNN introduces noticeable artifacts because of the ignorance of spectral-spatial correlations of HSI. Unlike QRNN3D which integrates the prior information of HSIs in a “black-box”

TABLE III
COMPARISON OF DIFFERENT METHODS ON WASHINGTON DC MALL. THE TOP TWO VALUES ARE MARKED **RED** AND **BLUE**.

σ	Index	Noisy	Sparse methods			Low-rank methods					DL methods				
			BM4D [29]	TDL [40]	MTSNMF [9]	LLRT [30]	NGMeet [23]	LRMR [59]	LRTDTV [44]	FastHyDe [24]	Dn-CNN [57]	HSI-SDe CNN [47]	HSID-CNN [21]	QRNN3D [46]	MAC-Net (Ours)
[0-15]	PSNR	32.38	43.89	35.31	48.39	33.57	47.83	44.80	46.73	56.40	39.83	44.30	41.85	36.92	57.43
	SSIM	0.6507	0.9621	0.7796	0.9878	0.7043	0.9681	0.9684	0.9792	0.9980	0.9034	0.9689	0.9479	0.8545	0.9983
	SAM	0.2978	0.0741	0.2003	0.0415	0.2536	0.0833	0.0609	0.0472	0.0199	0.1022	0.0684	0.0795	0.1327	0.0167
[0-55]	PSNR	22.52	37.18	27.88	42.40	23.43	40.73	36.98	42.39	48.48	35.04	38.34	36.93	35.86	50.59
	SSIM	0.3054	0.8554	0.4991	0.9478	0.3495	0.8920	0.8518	0.9384	0.9875	0.7582	0.8944	0.8769	0.8381	0.9908
	SAM	0.7883	0.1459	0.4679	0.0802	0.6870	0.2439	0.1514	0.0981	0.0460	0.1762	0.1268	0.1361	0.1475	0.0382
[0-95]	PSNR	16.80	34.51	24.29	38.68	18.08	37.03	32.69	40.14	46.70	32.37	35.34	32.11	33.85	45.52
	SSIM	0.1642	0.7717	0.3542	0.8790	0.1783	0.8278	0.7166	0.9026	0.9807	0.6632	0.8293	0.7344	0.7681	0.9746
	SAM	1.054	0.1719	0.6086	0.1219	0.9260	0.3802	0.2303	0.1174	0.0494	0.2399	0.1676	0.2319	0.1726	0.0676
Blind	PSNR	21.90	37.22	27.79	42.21	23.07	40.40	36.73	41.93	48.13	34.87	38.02	36.82	35.77	49.40
	SSIM	0.2891	0.8544	0.4986	0.9460	0.3360	0.8866	0.8416	0.9332	0.9875	0.7549	0.8900	0.8731	0.8362	0.9883
	SAM	0.8075	0.1496	0.4527	0.0801	0.7069	0.2441	0.1537	0.0917	0.0430	0.1801	0.1282	0.1278	0.1478	0.0427

TABLE IV
COMPUTATIONAL TIME (SECONDS) OF DIFFERENT DENOISING ALGORITHMS ON WASHINGTON DC MALL HSI.

Method	Sparse methods			Low-rank methods					DL methods				
	BM4D	TDL	MTSNMF	LLRT	NGMeet	LRMR	LRTDTV	FastHyDe	Dn-CNN	HSI-SDe CNN	HSID-CNN	QRNN3D	MAC-Net
Time	528.1	45.76	122.7	2059	131.5	113.5	1146	24.71	217.7	113.3	115.1	104.9	8.613

manner, MAC-Net is a spectral low-rank model aided network and is more interpretable. The combination of spectral low-rank model for and data-driven learning of spatial structures augments the denoising capacity of MAC-Net. For this reason, MAC-Net surpasses QRNN3D with better visual quality, as can be seen in the enlarged areas where the denoised HSI produced by QRNN3D contains noisy artifacts. Overall, FastHyDe and MAC-Net are visually closer to the clean one than other methods.

Fig. 5 depicts the spectral reflectance of pixel (451, 135) before and after noise reduction. In consistence with the phenomenon shown in Fig. 4, our MAC-Net provides the best visual quality by best matching with the ground-truth spectrum. To summarize, the abovementioned observations confirm the powerful denoising ability of the proposed MAC-Net and also the benefits of combining physical models with deep learning techniques for noise reduction over both model-based and DL-based methods.

C. Evaluation on Remote Sensing HSIs

Besides experiments with close-range HSIs, i.e., ICVL HSIs, we also ran all the competing methods on remote sensing HSI, i.e., Washington DC Mall HSI (WDCM HSI). WDCM HSI originally contains 1208×307 pixels and covers 191 spectral bands, and was acquired by the hyperspectral digital imagery collection experiment (HYDICE) sensor. We cropped a subimage with $256 \times 256 \times 191$ sizes to conduct the experiments. Compared with ICVL HSIs, WDCM HSI has a higher spectral resolution but much lower spatial resolution. Because of the comparatively smaller number of remote sensing HSIs available for fine-tuning, we directly use the model trained on the ICVL dataset to test all the DL-based methods so as to show the generalization ability of all the methods.

Table III compares the PSNR, SSIM, SAM of all the methods in different cases. Except for $\sigma \in [0 - 95]$, our MAC-Net achieves the best denoising performance. Even $\sigma \in [0 - 95]$, MAC-Net still ranks the second among all the methods. MAC-Net learns a universal network for all the HSIs and fails to consider the uniqueness of an individual HSI. This is particularly important for HSI denoising in the case of high noise and very large range of noise. Moreover, there is a huge gap between remote sensing HSIs and close-range HSIs in terms of spectral and spatial resolution. Differently, FastHyDe is a model-based method which does not depend on the training set and can adaptively capture the characteristic of the observed HSI. For this reason, MAC-Net is inferior to FastHyDe when $\sigma \in [0 - 95]$.

Fig. 6 and 7 respectively depicts the band images and spectrum before and after denoising. FastHyDe and MAC-Net obviously outperform alternative denoisers by removing most noises with preserved dominant details. MAC-Net and FastHyDe show very similar visual results. Overall, this experiment further shows the strong denoising ability and generalization ability of MAC-Net over purely data-driven DL-based methods.

We further record the running time of all the methods on Washington DC Mall HSI. All the methods run on a machine with an Intel(R) Xeon Silver 4214R CPU at 2.40 GHz and 126 GB RAM. The running time is shown in Table IV. It can be seen that DL-based methods generally require less running time than traditional model-based methods because of no time-consuming numerical iteration. Among DL-based methods, our MAC-Net spends the least running time as the subspace projection is very fast and can greatly reduce the size of image required for convolution operations. To summarize, this experiment further shows the computational efficiency of our MAC-Net over alternative methods.

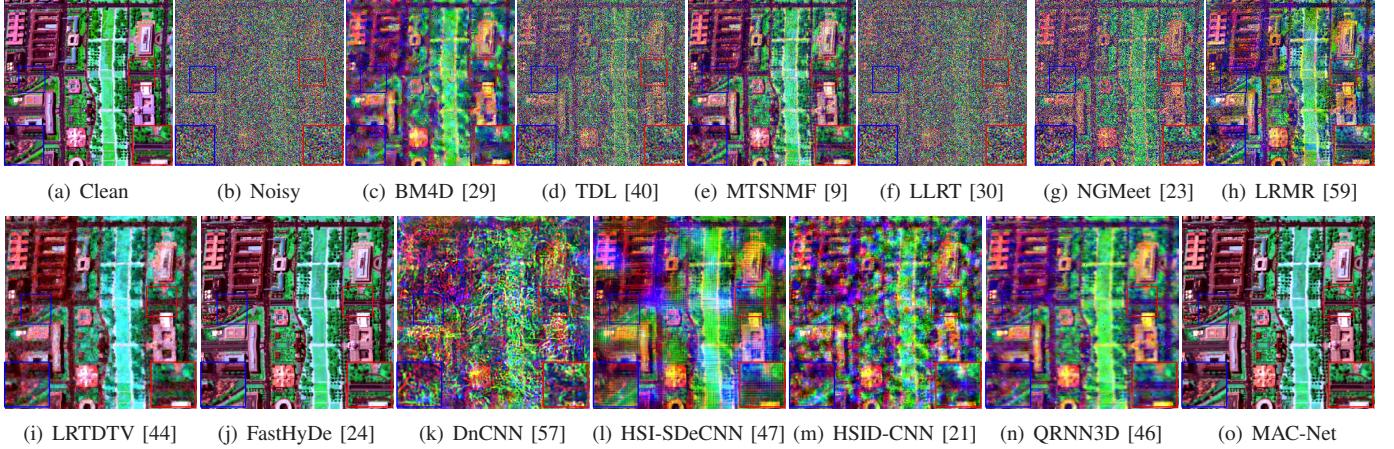


Fig. 6. Denoising results on the *Washington DC Mall* HSI with the noise variance in [0-95]. The false-color images were generated by combining bands 17, 94, 184. MAC-Net achieves the best visual results with less artifacts.

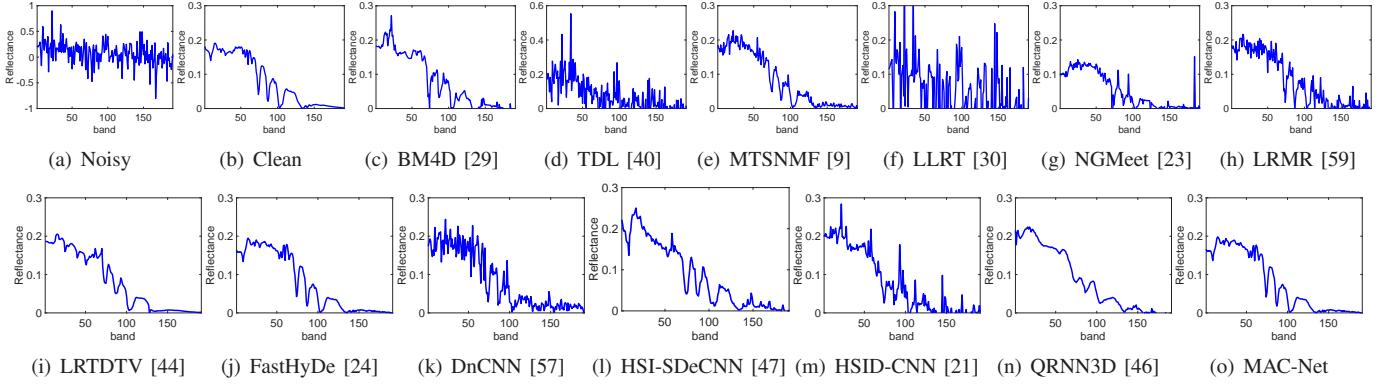


Fig. 7. Denoising results of pixel (100, 100) in *Washington DC Mall* HSI.

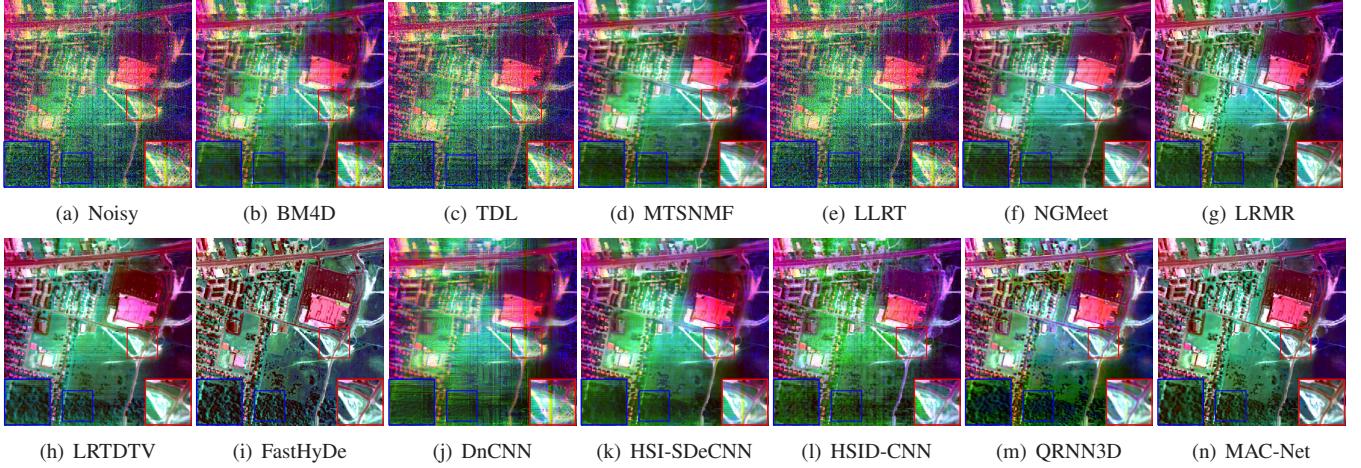


Fig. 8. Denoising results on real-world Urban dataset. The false-color images were generated by combining bands 1, 108, 208. The proposed MAC-Net provides the most appealing visual results.

D. Evaluation on Real-world Remote Sensing HSIs

We further carried out experiments on real-world remote sensing HSIs to thoroughly evaluate the denoising capacity of all the methods. Since there is no clean counterpart of real-world remote sensing HSI, we chose to qualitatively compare the denoising results.

1) *HYDICE Urban HSI*: HYDICE Urban HSI was collected by the HYDICE sensor and contains 307×307 pixels and 210 spectral bands with wavelength covering from 0.4

to $2.5 \mu\text{m}$. As can be seen in Fig. 8 (a), the HSI is heavily corrupted by complex noises, making it very suitable to examine the denoising ability of all the methods. Among model-based methods, FastHyDe and LRTDTV achieve better effects because of the simultaneous integration of the nonlocal/local spatial structure and global spectral low-rankness. The ignorance of the spectral-spatial structure of HSIs leads to serious details loss in the denoised HSIs produced by band-wise DnCNN. HSID-CNN and HSI-SDeCNN show inferior denois-

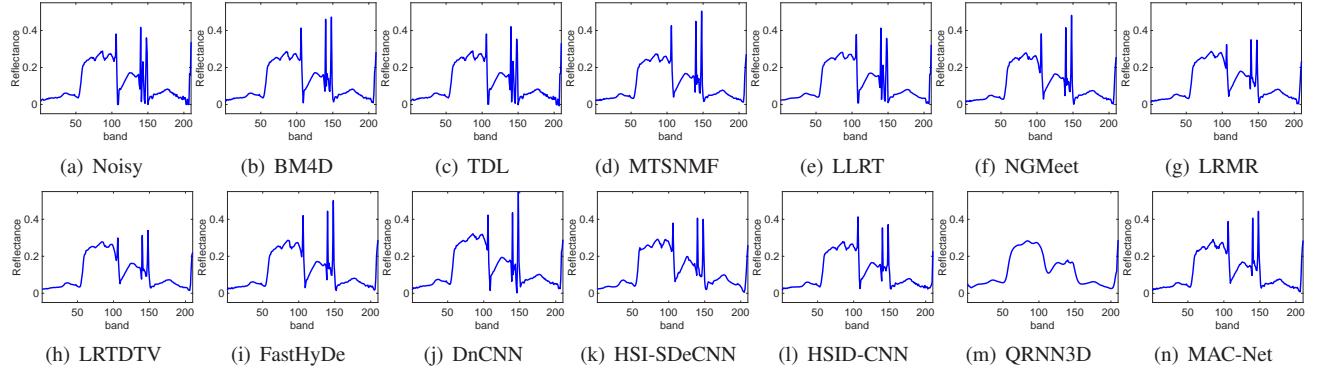


Fig. 9. Denoising results of pixel (249, 216) in real-world Urban data set.

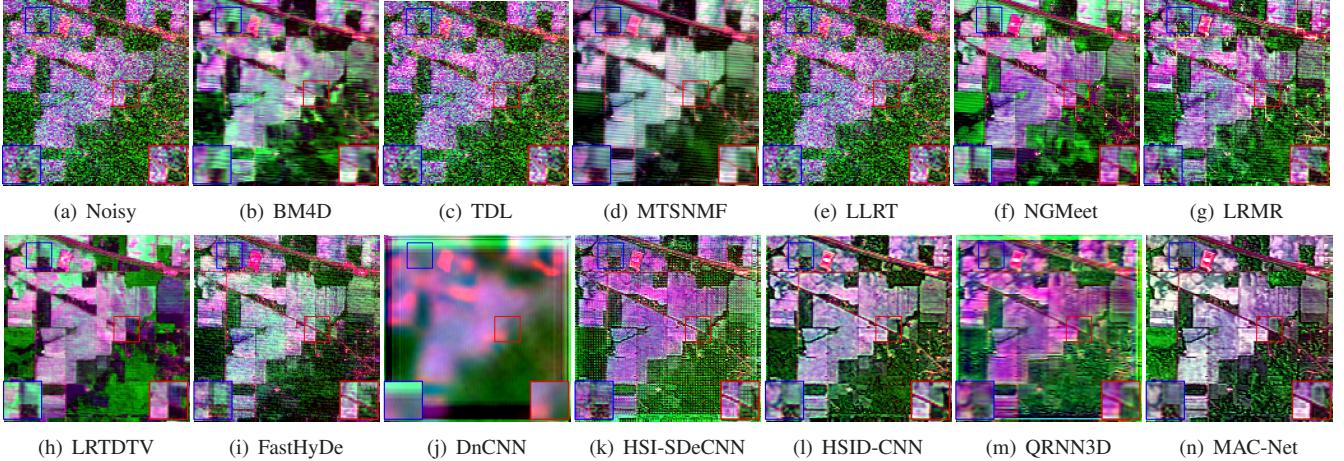


Fig. 10. Denoising results on real-world Indian Pines dataset. The false-color images were generated by combining bands 3, 108, 210. The proposed MAC-Net provides the best visual results.

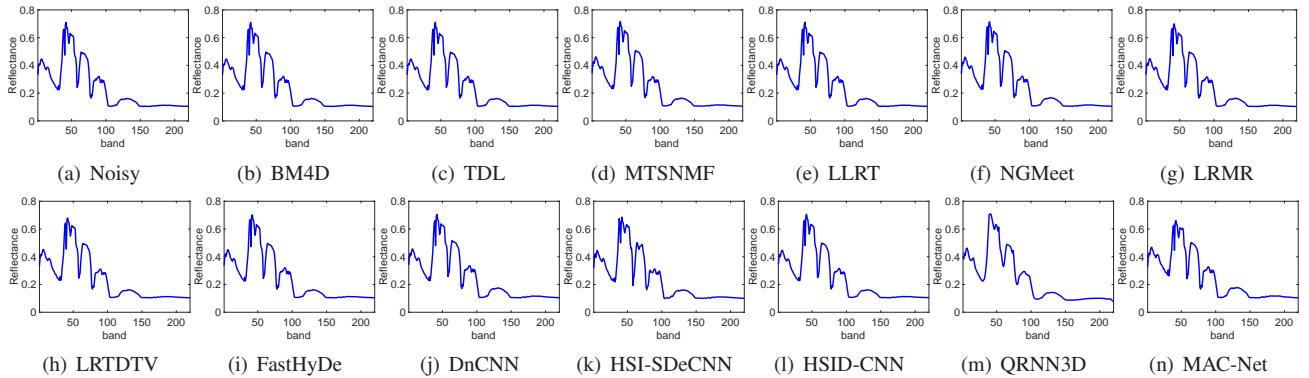


Fig. 11. Denoising results of pixel (100, 100) in real-world remote sensing Indian Pines dataset.

ing than QRNN3D and MAC-Net as they fail to capture the global spectral correlation. The hybrid advantages of model-based and DL-based methods allow MAC-Net to outperform QRNN3D, LRTDTV, and FastHyDe by removing most noises while preserving important textures. Fig. 9 depicts the spectrum of denoised and noisy HSI at the location of (216, 249). As the phenomena revealed in Fig. 8, LRTDTV, FastHyDe, QRNN3D, and MAC-Net provide the most desirable denoising thanks to higher ability to depict the global spectral correlation and local/nonlocal spatial dependencies. Benefiting from the embedded spectral low-rank model, MAC-Net can better preserve the local spectral details than QRNN3D. Overall, this experiment evidently shows the strong denoising ability and

generalization ability of our MAC-Net to remote sensing HSIs.

2) *AVIRIS Indian Pines HSI*: The Indian Pines HSI was imaged by NASA AVIRIS sensor and includes $145 \times 145 \times 220$ bands with a spatial resolution of 20m per pixel. Fig. 10 illustrates the denoising performance of all the compared methods as well as the original noisy HSI. Compared with alternative methods that either produce blur images or fail to thoroughly remove the noises, LRTDTV, FastHyDe and MAC-Net achieve dominant denoising effectiveness by recovering local details, especially edges, as they can better capture the spectral low-rankness of HSIs. Fig. 11 plots the spectral reflectance of pixel (100,100). As can be seen, all the methods provide very similar spectrum in real visual perception.

E. Ablation Study

Here, we conduct a comprehensive study on our MAC-Net with respect to the nonlocal block, network depth, and the number of training samples. All the experiments are conducted on ICVL training and testing datasets with noise variances in the range of [0-55].

TABLE V
THE DENOISING PERFORMANCE WITH/WITHOUT NAB.

Index	PSNR	SSIM	SAM
With	43.02	0.9925	0.0280
Without	42.49	0.9912	0.0292

1) *Effectiveness of the Nonlocal Attention Block:* In this experiment, we examine the effectiveness of NAB on the denoising performance. Table V presents the denoising performance with/without NAB. NAB encourages to capture of the nonlocal dependencies among different contents of the HSI and thereby improves denoising performance with a gain of 0.53 in terms of PSNR and 0.0013 in terms of SSIM, verifying its effectiveness.

2) *Impact of the Number of Training Samples:* The advantage of model-based methods includes less requirement of training samples and high generalization ability. In this experiment, we changed the number of training samples from 20 to 100 with an interval of 20. As shown in Table VI, on one hand, with the increase of training samples, PSNR improves from 42.44 to 43.02, SSIM grows from 0.9760 to 0.9925 and SAM drops from 0.0286 to 0.0280, indicating the denoising performance gradually gets better. On the other hand, compared with the training with 20 HSIs, the changes of training samples notice an insignificant gain of 0.58 and 0.0165 with respect to PSNR and SSIM respectively when 5 times of the training samples were used. Even with 20 HSIs, the denoising performance still outperforms alternative methods for example QRNN3D whose PSNR is 40.15. This phenomenon means that our MAC-Net doesn't heavily rely on training samples as alternative black-box methods such as QRNN3D, showing the practicality of our method in real-world HSI denoising where training samples are very rare. The main reason lies in the hybrid advantages of model-based methods in generalization ability and learning-based methods in strong representation capacity.

TABLE VI
THE IMPACT OF THE NUMBER OF TRAINING SAMPLES.

Index	20	40	60	80	100
PSNR	42.44	42.49	42.17	42.67	43.02
SSIM	0.9760	0.9767	0.9708	0.9754	0.9925
SAM	0.0286	0.0288	0.0311	0.0281	0.0280

3) *Impact of the Network Depth:* Apart from the number of training samples, we also test how the network depth affects the denoising performance. Σ , D , Σ^{-1} , D^T respectively occupy fixed layers, only leaving our SDPM impacting the network depth. As presented in Fig. 12, more layers contribute to a more powerful representation and thereby higher denoising ability. When the number of layers is more than 12, the

denoising performance tends to be stable. Based on the above observation, we set the number of layers as 12 to balance the denoising performance and network parameters.

4) *Sensitivity Analysis of K:* Accurately estimating the spectrum subspace dimensionality K is actually a hard task. Denoting the subspace dimensionality estimated by the Hysime algorithm as \hat{R} , K varies around \hat{R} to show the changes of denoising performance. As can be seen from Table VII, an overestimated K leads to lower denoising effectiveness because the spectral low-rankness of HSI cannot be fully exploited. An underestimation of K also results in poor denoising and the best denoising performance is obtained by setting $K = \hat{R}$. From the experiment, we think that Hysime is a reasonable tool to estimate K , which is also in line with many previous works such as LRTV [33] and LRTDTV [44].

TABLE VII
DENOISING PERFORMANCE CHANGES WITH RESPECT TO K .

K	\hat{R} -2	\hat{R} -1	\hat{R}	\hat{R} +1	\hat{R} +2
PSNR	42.88	42.88	43.02	42.59	41.85
SSIM	0.9801	0.9801	0.9925	0.9799	0.9925
SAM	0.0284	0.0284	0.0280	0.0294	0.0280

V. CONCLUSION

In this paper, we introduce a model aided nonlocal neural network (MAC-Net) for hyperspectral image denoising. The embedded spectral low-rank model encourages MAC-Net to absorb the generalization capability of model-based approaches for reasonable depiction of the spectral correlation among bands. The introduced nonlocal U-Net enables to inherit the strong representation capability of deep learning-based approaches for effective exploitation of local and nonlocal multi-scale spatial dependencies. Experiment comparisons on both close-range and remote sensing HSIs show that our MAC-Net outperforms both model-based and deep learning-based methods in terms of both subjective visual effects and objective quantitative measurements. Additional ablation study also confirms its strong learning ability, superior generalization capability, and less requirement of training samples. In our future work, we will extend the network to cope with mixed noises by incorporating the noise models.

REFERENCES

- [1] Q. Huang, W. Li, B. Zhang, Q. Li, R. Tao, and N. H. Lovell, “Blood cell classification based on hyperspectral imaging with modulated gabor and cnn,” *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 1, pp. 160–170, 2020.
- [2] R. Hang, Z. Li, P. Ghamisi, D. Hong, G. Xia, and Q. Liu, “Classification of hyperspectral and lidar data using coupled cnns,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4939–4950, 2020.
- [3] F. Xiong, J. Zhou, and Y. Qian, “Material based object tracking in hyperspectral videos,” *IEEE Trans. Image Process.*, vol. 29, pp. 3719–3733, 2020.
- [4] B. Rasti, P. Scheunders, P. Ghamisi, G. Licciardi, and J. Chanussot, “Noise reduction in hyperspectral imagery: Overview and application,” *Remote Sensing*, vol. 10, no. 3, 2018.
- [5] H. Zhang, J. Cai, W. He, H. Shen, and L. Zhang, “Double low-rank matrix decomposition for hyperspectral image denoising and destriping,” *IEEE Trans. Geosci. Remote Sens.*, pp. 1–19, 2021.
- [6] T.-X. Jiang, L. Zhuang, T.-Z. Huang, X.-L. Zhao, and J. M. Bioucas-Dias, “Adaptive hyperspectral mixed noise removal,” *IEEE Trans. Geosci. Remote Sens.*, pp. 1–13, 2021.

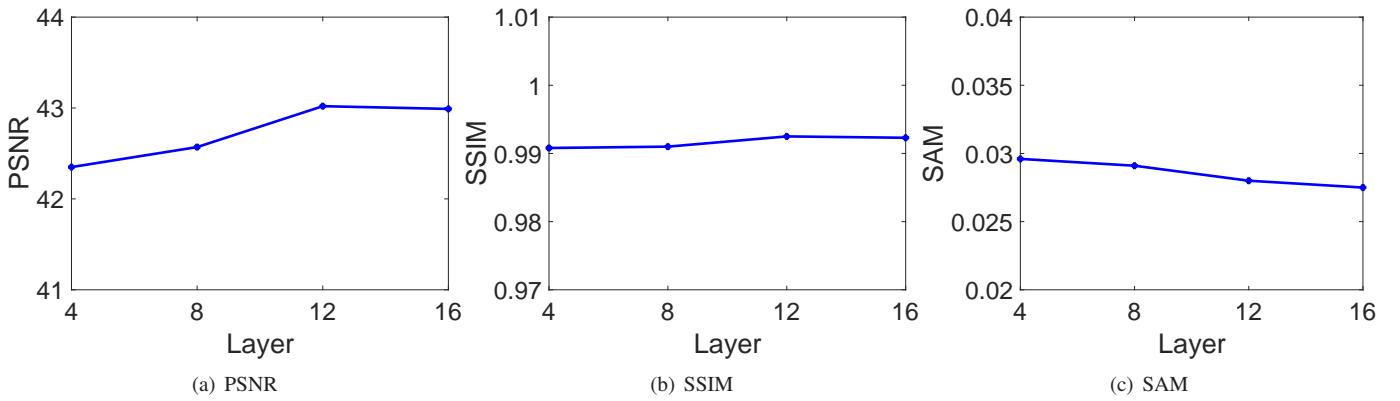
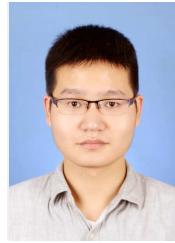


Fig. 12. PSNR, SSIM and SAM changes with respect to network depth.

- [7] P. Azimpour, T. Bahraini, and H. S. Yazdi, "Hyperspectral image denoising via clustering-based latent variable in variational bayesian framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3266–3276, 2021.
- [8] Y. Fu, A. Lam, I. Sato, and Y. Sato, "Adaptive spatial-spectral dictionary learning for hyperspectral image denoising," in *Proc. IEEE Int. Cof. Comput. Vis. (ICCV)*, December 2015.
- [9] M. Ye, Y. Qian, and J. Zhou, "Multitask sparse nonnegative matrix factorization for joint spectral-spatial hyperspectral imagery denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2621–2639, 2015.
- [10] T. Lu, S. Li, L. Fang, Y. Ma, and J. A. Benediktsson, "Spectral-spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, 2016.
- [11] J. Li, Q. Yuan, H. Shen, and L. Zhang, "Noise removal from hyperspectral image with joint spectral-spatial distributed sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 9, pp. 5425–5439, 2016.
- [12] X. Gong, W. Chen, and J. Chen, "A low-rank tensor dictionary learning method for hyperspectral image denoising," *IEEE Trans. Signal Process.*, vol. 68, pp. 1168–1180, 2020.
- [13] L. Zhuang, X. Fu, M. K. Ng, and J. M. Bioucas-Dias, "Hyperspectral image denoising based on global and nonlocal low-rank factorizations," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–17, 2021.
- [14] Y. Chang, L. Yan, B. Chen, S. Zhong, and Y. Tian, "Hyperspectral image restoration: Where does the low-rank property exist," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6869–6884, 2021.
- [15] J. Peng, Q. Xie, Q. Zhao, Y. Wang, L. Yee, and D. Meng, "Enhanced 3DTV regularization and its applications on hsi denoising and compressed sensing," *IEEE Trans. Image Process.*, vol. 29, pp. 7889–7903, 2020.
- [16] H. Wang, Y. Cen, Z. He, Z. He, R. Zhao, and F. Zhang, "Reweighted low-rank matrix analysis with structural smoothness for image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 4, pp. 1777–1792, 2018.
- [17] X. Bai, F. Xu, L. Zhou, Y. Xing, L. Bai, and J. Zhou, "Nonlocal similarity based nonnegative tucker decomposition for hyperspectral image denoising," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 11, no. 3, pp. 701–712, 2018.
- [18] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang, "Hyperspectral image denoising using a 3-D attention denoising network," *IEEE Trans. Geosci. Remote Sens.*, pp. 1–16, 2021.
- [19] W. Dong, H. Wang, F. Wu, G. Shi, and X. Li, "Deep spatial-spectral representation learning for hyperspectral image denoising," *IEEE Trans. Comput. Imaging*, vol. 5, no. 4, pp. 635–648, 2019.
- [20] Y. Chang, L. Yan, H. Fang, S. Zhong, and W. Liao, "HSI-DeNet: Hyperspectral image restoration via convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 667–682, 2019.
- [21] Q. Yuan, Q. Zhang, J. Li, H. Shen, and L. Zhang, "Hyperspectral image denoising employing a spatspectral deep residual convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1205–1218, 2019.
- [22] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing," *IEEE Signal Process. Mag.*, vol. 38, no. 2, pp. 18–44, 2021.
- [23] W. He, Q. Yao, C. Li, N. Yokoya, Q. Zhao, H. Zhang, and L. Zhang, "Non-local meets global: An integrated paradigm for hyperspectral image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2020.
- [24] L. Zhuang and J. M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 11, no. 3, pp. 730–742, 2018.
- [25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems (NeurIPS)*, 2017, pp. 5998–6008.
- [26] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, June 2018.
- [27] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.
- [28] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.
- [29] 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, 2013.
- [30] 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)*, 2017, pp. 5901–5909.
- [31] Y. Qian and M. Ye, "Hyperspectral imagery restoration using nonlocal spectral-spatial structured sparse representation with noise estimation," *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 6, no. 2, pp. 499–515, 2013.
- [32] Y. Chen, X. Cao, Q. Zhao, D. Meng, and Z. Xu, "Denoising hyperspectral image with non-i.i.d. noise structure," *IEEE Transactions on Cybernetics*, vol. 48, no. 3, pp. 1054–1066, 2018.
- [33] 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, 2016.
- [34] Y. Xie, Y. Qu, D. Tao, W. Wu, Q. Yuan, and W. Zhang, "Hyperspectral image restoration via iteratively regularized weighted schatten p -norm minimization," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4642–4659, 2016.
- [35] T. Xie, S. Li, and B. Sun, "Hyperspectral images denoising via non-convex regularized low-rank and sparse matrix decomposition," *IEEE Trans. Image Process.*, vol. 29, pp. 44–56, 2020.
- [36] Q. Xie, Q. Zhao, D. Meng, and Z. Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1888–1902, 2018.
- [37] F. Xiong, J. Zhou, and Y. Qian, "Hyperspectral imagery denoising via reweighed sparse low-rank nonnegative tensor factorization," in *Proc. IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3219–3223.
- [38] ———, "Hyperspectral restoration via l_0 gradient regularized low-rank tensor factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10410–10425, 2019.
- [39] X. Liu, S. Bourennane, and C. Fossati, "Denoising of hyperspectral images using the parafac model and statistical performance analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3717–3724, 2012.
- [40] 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. (CVPR)*, 2014, pp. 2949–2956.
- [41] Y.-B. Zheng, T.-Z. Huang, X.-L. Zhao, T.-X. Jiang, T.-H. Ma, and T.-Y. Ji, “Mixed noise removal in hyperspectral image via low-fibered-rank regularization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 734–749, 2020.
- [42] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, “Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition,” vol. 50, no. 8, pp. 3556–3570, 2020.
- [43] Y. Chen, W. He, N. Yokoya, T.-Z. Huang, and X.-L. Zhao, “Nonlocal tensor-ring decomposition for hyperspectral image denoising,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1348–1362, 2020.
- [44] 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 Observations Remote Sens.*, vol. 11, no. 4, pp. 1227–1243, 2018.
- [45] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, “Deep feature extraction and classification of hyperspectral images based on convolutional neural networks,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, 2016.
- [46] K. Wei, Y. Fu, and H. Huang, “3-D quasi-recurrent neural network for hyperspectral image denoising,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 363–375, 2021.
- [47] 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, 2020.
- [48] Y. Yuan, H. Ma, and G. Liu, “Partial-DNet: A novel blind denoising model with noise intensity estimation for HSI,” *IEEE Trans. Geosci. Remote Sens.*, pp. 1–13, 2021.
- [49] Y. Qian, F. Xiong, Q. Qian, and J. Zhou, “Spectral mixture model inspired network architectures for hyperspectral unmixing,” *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7418–7434, 2020.
- [50] F. Xiong, J. Zhou, M. Ye, J. Lu, and Y. Qian, “NMF-SAE: An interpretable sparse autoencoder for hyperspectral unmixing,” in *Proc. IEEE International Conference on Acoustics & Speech and Signal Processing (ICASSP)*, 2021, pp. 1865–1869.
- [51] W. Dong, C. Zhou, F. Wu, J. Wu, G. Shi, and X. Li, “Model-guided deep hyperspectral image super-resolution,” *IEEE Trans. Image Process.*, vol. 30, pp. 5754–5768, 2021.
- [52] Q. Xie, M. Zhou, Q. Zhao, Z. Xu, and D. Meng, “MHF-Net: An interpretable deep network for multispectral and hyperspectral image fusion,” *IEEE Trans. Pattern Anal. Mach. Intell.*, pp. 1–1, 2020.
- [53] J. M. Bioucas-Dias and J. M. P. Nascimento, “Hyperspectral subspace identification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 8, pp. 2435–2445, 2008.
- [54] C. Cao, J. Yu, C. Zhou, K. Hu, F. Xiao, and X. Gao, “Hyperspectral image denoising via subspace-based nonlocal low-rank and sparse factorization,” *IEEE J. Sel. Topics Appl. Earth Observations Remote Sens.*, vol. 12, no. 3, pp. 973–988, 2019.
- [55] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Proc. International Conference on Medical image computing and computer-assisted intervention (MICCAI)*, 2015, pp. 234–241.
- [56] W. Dong, P. Wang, W. Yin, G. Shi, F. Wu, and X. Lu, “Denoising prior driven deep neural network for image restoration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2305–2318, 2019.
- [57] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising,” *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, 2017.
- [58] Z. Shen, M. Zhang, H. Zhao, S. Yi, and H. Li, “Efficient attention: Attention with linear complexities,” in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, pp. 3531–3539.
- [59] 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, 2014.
- [60] X. Liu, M. Tanaka, and M. Okutomi, “Single-image noise level estimation for blind denoising,” *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 5226–5237, 2013.



Fengchao Xiong (S’18-M’20) received the B.E. degree in software engineering from Shandong University, Jinan, China, in 2014 and the Ph.D. degree from the College of Computer Science, Zhejiang University, Hangzhou, China, in 2019.

He visited Wuhan University from 2011 to 2012 and Griffith University from 2017 to 2018. He is currently a Lecturer with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. His research interests include hyperspectral image processing, machine learning, and pattern recognition.



Jun Zhou (M’06-SM’12) received the B.S. degree in computer science and the B.E. degree in international business from Nanjing University of Science and Technology, Nanjing, China, in 1996 and 1998, respectively. He received the M.S. degree in computer science from Concordia University, Montreal, Canada, in 2002, and the Ph.D. degree from the University of Alberta, Edmonton, Canada, in 2006.

In June 2012, he joined the School of Information and Communication Technology, Griffith University, Nathan, Australia, where he is currently an Associate Professor. Prior to this appointment, he was a Research Fellow with the Research School of Computer Science, Australian National University, Canberra, Australia, and a Researcher with the Canberra Research Laboratory, NICTA, Canberra. His research interests include pattern recognitions, computer vision, and spectral imaging with their applications in remote sensing and environmental informatics.



Qinling Zhao received the Ph.D. degree from Zhejiang University, Hangzhou, China in 2015. She is currently an Associate Professor in the school of computer science and engineering at the Nanjing University of Science and Technology. Her research interests include Intelligent embedded systems, operation system scheduling optimization and machine learning.



Jianfeng Lu (Member, IEEE) received his M.S. and Ph.D. degrees in Pattern Recognition and Intelligent System from the School of Computer Science and Engineering, Nanjing University of Science and Technology. He is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology. His research interests include image analysis, machine learning, data mining, and intelligent robot.



Yuntao Qian (M’04-SM’21) received the B.E. and M.E. degrees in automatic control from Xi’an Jiaotong University, Xi’an, China, in 1989 and 1992, respectively, and the Ph.D. degree in signal processing from Xidian University, Xi’an, China, in 1996.

During 1996–1998, he was a Postdoctoral Fellow with the Northwestern Polytechnical University, Xi’an, China. Since 1998, he has been with the College of Computer Science, Zhejiang University, Hangzhou, China, where he became a Professor in 2002. During 1999–2001, 2006, 2010, 2013, 2015–2016, and 2018 he was a Visiting Professor at Concordia University, Hong Kong Baptist University, Carnegie Mellon University, the Canberra Research Laboratory of NICTA, Macau University, and Griffith University. His current research interests include machine learning, signal and image processing, pattern recognition, and hyperspectral imaging. He is currently an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.