MULTITASK SPARSE NEURAL NETWORK FOR HYPERSPECTRAL IMAGE DENOISING

Fengchao Xiong[1] *Minchao Ye*[2] *Jun Zhou*[3] *Jianfeng Lu*[1] *Yuntao Qian*[4]

[1] School of Computer Science and Engineering, Nanjing University of Science and Technology, China
[2] College of Information Engineering, China Jiliang University, China
[3] School of Information and Communication Technology, Griffith University, Australia
[4] College of Computer Science, Zhejiang University, China

ABSTRACT

Data-driven deep learning (DL)-based methods directly learn the nonlinear mapping between noisy hyperspectral images (HSIs) and corresponding clean ones. However, DL-based methods neglect the prior knowledge of HSIs embodied by physical models. Consequently, they require complex network architectures and a large number of training samples. To address the above issues, this paper introduces a multitask sparse neural network (MTSNN) which bridges the sparsity prior of HSIs with data-driven deep learning for HSI denoising. Specifically, we first build a multitask sparse (MTS) denoising model which shares sparse coefficients among bands to exploit the spectral-spatial correlation and learns a dictionary for each band to depict the distinct spatial structure among bands. The iterative optimization of the MTS model is then unfolded to yield our MTSNN by introducing some learnable parameters. MTSNN is a multi-branch network. Each branch performs a single denoising task for an individual band. All branches are connected by shared coefficients, forming multitask denoising for all bands. The hybrid advantages of the MTS model and data-driven learning equip MTSNN with strong denoising ability, preferable learning capability, superior interpretability, and higher generalization capacity. Experimental results demonstrate that our method achieves state-of-the-art denoising performance compared with several alternative approaches.

***Index Terms*—** Hyperspectral image denoising, model-based neural network, multitask learning, sparse coding

1. INTRODUCTION

Hyperspectral images (HSIs) have been applied to a variety of fields such as object detection [1] and tracking [2] thanks to the material identification ability enabled by hundreds of contiguous narrow bands. Because of the imaging condition, sensor defects, calibration error, etc., the captured HSIs are usually contaminated with noises. As an indispensable preprocessing step, HSI denoising is helpful to subsequent applications. HSI denoising suffers from three difficulties, i.e., spectral-spatial structure-preserving, high-dimensionality nature, and limited training samples.

Noise exists in both the spatial and spectral domains of HSIs. Intuitively, classic grayscale image denoising methods, for example, BM3D [3] and DnCNN [4] can be applied in a band-by-band manner. Bandwise denoising neglects the spectral-spatial structure of HSIs and can not fully remove all the noises. Spectral-spatial methods model HSIs with tensors or cubes to simultaneously consider the spatial and spectral information and achieve more desirable denoising [5, 6].

Among spectral-spatial methods, low-rank and sparse representation-based methods stand out thanks to their superior capability of depicting the low-dimensional structure of underlying clean HSIs. By coding HSIs in a low-dimensional subspace, the noises can be removed to a large extent. Low-rank representation (LRR)-based methods decompose the observed HSIs into the product of factor matrices/tensors [7, 8] or minimize the rank of matrix/tensors [9–11] so that the local spectral-spatial structure of HSIs is depicted. Sparse representation (SR)-based methods express the noisy HSIs over a dictionary [12, 13]. A combination of sparse and low-rank representation has also been applied for HSI denoising [14, 15]. No matter LRR or SR, they are both based on hand-crafted prior of HSIs, for example, local and non-local similarity prior and have preferable interpretability as well as high generalization ability. However, most LRR and SR methods formulate HSI denoising as an optimization problem which usually requires numerous iterations and exhaustive hyperparameter tuning to produce high-quality denoising performance.

Instead of explicit priors, deep learning (DL)-based methods directly learn the nonlinear mapping between observed HSIs and clean counterparts [16]. Yuan *et. al* [17] took advantage of residual learning and multiscale representation for HSI denoising. Wei *et. al* [18] adopted a quasi-recurrent pooling function to capture the spectral-spatial correlations

This work was supported in part by the National Natural Science Foundation of China under Grant 62002169 and 62071421, Jiangsu Provincial Natural Science Foundation of China under Grant BK20200466, Zhejiang Provincial Natural Science Foundation of China under Grant LY22F010010 and 111 Program (No. B13022). (Corresponding author: Fengchao Xiong; fcxiong@njust.edu.cn)

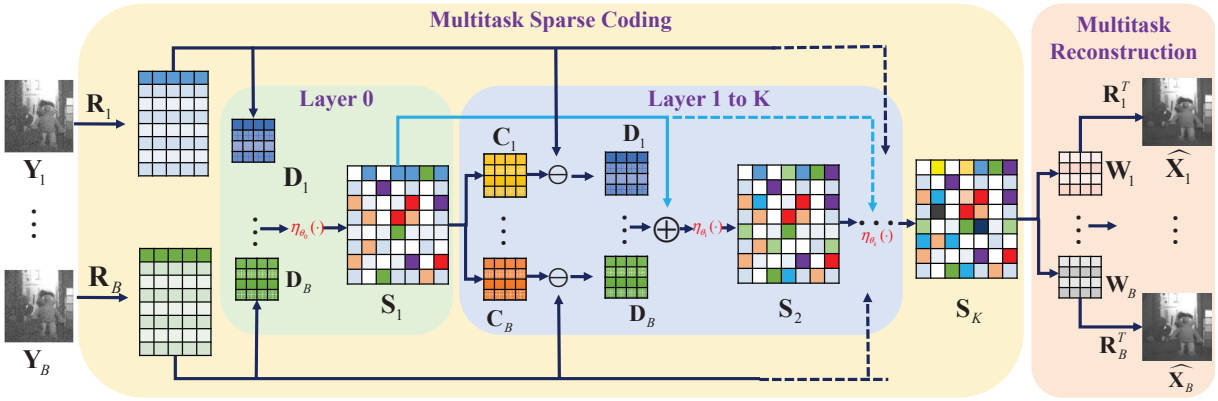


Fig. 1. Network architecture of MTSNN.

better and achieved better denoising performance. Noise intensities were employed as an input to handle a large range of noises with a single network [19]. The architecture of most DL-based methods is generic rather than specified for HSI denoising. This makes them lack interpretability and require a complex network architecture and many training samples to achieve desired HSI denoising. The high cost of HSI acquisition causes HSI denoising suffers from a small training sample problem, limiting the feasibility of DL-based methods.

This paper introduces a multitask sparse neural network that combines the achievements of SR-based and DL-based methods for improved HSI denoising. Regarding denoising one band as a single task and denoising all bands as a multitask problem, we introduce a multitask sparse (MTS) model that connects the denoising task of all the bands by sharing sparse coefficients. The learned dictionaries of all bands embody the spatial structure of HSIs, and the shared sparse matrix captures spectral-spatial correlation among bands. We then consider the iterative optimization of the MTS model as a forward process of a neural network and introduce some learnable weights to parameterize the optimization with a multitask sparse neural network shown in Fig. 1. MTSNN includes a multitask sparse coding module to induce shared coefficients among bands and a multitask sparse reconstruction module to reconstruct individual bands with distinct dictionaries. The prior knowledge embodied by the MTS model facilitates improving the generalization and interpretation ability of MTSNN. Moreover, data-driven training also compensates for the limitation of the MTS model, such as exhaustive parameter twisting. Extensive experiments show that our MTSNN outperforms SR and LR-based methods and DL-based methods. Throughout analysis also confirms its high generation capacity and learning ability.

2. PROPOSED MTSNN METHOD

2.1. Multitask Sparse Denoising Model

Given an HSI $\mathbf{Y}$ with N pixels and B bands, under the contamination of additive Gaussian noise, the degradation can be represented by

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

where $\mathbf{X} \in \mathbb{R}^{N\times B}$ and $\mathbf{N} \in \mathbb{R}^{N\times B}$ represent underlying clean HSI and noises, respectively. For a local patch $\mathbf{y}_b$ extracted from b-th band containing $p \times p$ pixels, SR-based denoising for a specified band can be expressed by

$$\min_{\mathbf{s}_b} \frac{1}{2}\|\mathbf{y}_b - \mathbf{A}_b\mathbf{s}_b\|_F^2 + \lambda\|\mathbf{s}_b\|_1 \tag{2}$$

where $\mathbf{A}_b \in \mathbb{R}^{p^2\times M}$ is the dictionary containing M atoms. For all the patches in the b-th band image, SR-based denoising problem can be written as

$$\min_{\{\mathbf{s}_b^{(n)}\}_n} \sum_n \frac{1}{2}\|\mathbf{R}_b^{(n)}\mathbf{Y}_b - \mathbf{A}_b\mathbf{s}_b^{(n)}\|_F^2 + \lambda\|\mathbf{s}_b^{(n)}\|_1 \tag{3}$$

where $\mathbf{R}_b^{(n)}$ denotes extracting n-th patch from b-th band image $\mathbf{Y}_b$ and L_1 norm of $\mathbf{s}_b$ is used to promote sparse solutions.

Treating the denoising problem for an individual band as a single task, the denoising problem for all the bands is a multitask task linked by shared sparse coefficients $\mathbf{S} = [\mathbf{s}^{(1)}, \cdots, \mathbf{s}^{(n)}, \cdots]$, leading to the multitask sparse denoising model. By sharing the coefficients $\mathbf{s}^{(n)}$ among all the bands, the spectral-spatial correlations are exploited. Formally, the MTS model can be expressed as

$$\min_{\mathbf{s}^{(n)}} \sum_{b=1}^{B}\sum_n \left(\frac{1}{2}\|\mathbf{R}_b^{(n)}\mathbf{Y}_b - \mathbf{A}_b\mathbf{s}^{(n)}\|_F^2 + \lambda\|\mathbf{s}^{(n)}\|_1\right) \tag{4}$$

where different dictionaries are used to describe different spatial structures of different bands.

2.2. Multitask Sparse Neural Network

Iterative shrinkage-thresholding algorithm is a privilege tool to solve Eq. (4) with the following update rule:

$$\mathbf{s}_{k+1}^{(n)} = \mathbf{soft}_\theta\left(\mathbf{s}_k^{(n)} + \sum_{b=1}^{B} t\mathbf{A}_b^{\mathrm{T}}\left(\mathbf{R}_b^{(n)}\mathbf{Y}_b - \mathbf{A}_b\mathbf{s}_k^{(n)}\right)\right) \tag{5}$$

Table 1. Comparison of different methods on 50 testing HSIs of ICVL Dataset. The top two values are marked *Red* and *Blue*.

σ	Index	Noisy	Sparse methods			Low-rank methods				DL methods				
			BM 4D	TDL	MTS NMF	3DLog TNN	LLRT	NG Meet	Fast HyDe	Dn-CNN	SDe CNN	QRN N3D	QRN N3D-R	**MTS NN**
[0-15]	PSNR↑	33.47	45.12	38.84	45.62	43.89	46.17	39.42	**47.85**	41.92	41.06	43.36	45.91	**47.86**
	SSIM↑	.6295	.9753	.8481	.9562	.9902	.9649	.8647	**.9918**	.9596	.9565	.9884	.9901	**.9948**
	SAM↓	.1622	.0254	.0829	.0409	**.0150**	.0319	.0891	.0163	.0274	.0329	.0336	.0200	**.0131**
[0-55]	PSNR↑	21.67	38.21	29.45	37.57	33.37	37.19	30.48	41.73	37.90	35.66	40.15	**41.91**	**42.63**
	SSIM↑	.2361	.6211	.5238	.8569	.6892	.8389	.6748	**.9845**	.9293	.8736	.9729	.9733	**.9796**
	SAM↓	.1861	.0552	.2409	.1372	.2766	.1020	.2830	.0344	.0504	**.0309**	.0602	.0382	**.0215**
[0-95]	PSNR↑	16.97	35.27	25.40	34.20	24.53	32.49	27.20	**39.30**	34.65	32.45	37.66	38.41	**40.47**
	SSIM↑	.1442	.8764	.3735	.7998	.4251	.7191	.5506	**.9695**	.8442	.7875	.9479	.9267	**.9774**
	SAM↓	.7199	.0799	.3641	.2128	.6087	.1687	.4275	**.0447**	.1094	.0876	.0468	.0493	**.0270**

where t is the step size. $\mathbf{soft}_\theta$ is a soft-shrinkage function defined as $\mathbf{soft}_\theta(\mathbf{s}) = \text{sign}(\mathbf{s})\max(|\mathbf{s}| - \theta, 0)$ where θ is the shrinkage parameter related with λ and t. After numbers of iterative optimization, $\widehat{\mathbf{s}^{(n)}}$ can be obtained. Then the denoised HSIs is yield by concatenating the denoising results of each band, i.e., $\widehat{\mathbf{X}} = [\widehat{\mathbf{X}_1}, \cdots, \widehat{\mathbf{X}_B}]$. Each $\mathbf{X}_b$ is obtained by summarising the overlapped patches, i.e.,

$$\widehat{\mathbf{X}_b} = \left(\sum_{n=1}(\mathbf{R}_b^{(n)})^{\mathrm{T}}\mathbf{R}_b^{(n)}\right)^{-1}\sum_{n=1}(\mathbf{R}_b^{(n)})^{\mathrm{T}}\mathbf{A}_b\widehat{\mathbf{s}^{(n)}} \tag{6}$$

where $(\mathbf{R}_b^{(n)})^{\mathrm{T}}$ puts back the patches to the original locations.

The physical property of HSIs enables the interpretability of the MTS model. However, the MTS model is based on the hand-crafted prior and can not enjoy the advantages of data-driven leaned priors. Hyperparameter tuning is also troublesome but important for desired denoising performance.

For another point of view, the iterative optimization of $\mathbf{s}^{(n)}$ can be considered as the forward process of a neural network consisting of K layers. To this end, we introduce two set of parameters $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_B]$ with $\mathbf{D}_b = t\mathbf{A}_b^{\mathrm{T}}$ and $\mathbf{C} = [\mathbf{C}_1, \cdots, \mathbf{C}_B]$ with $\mathbf{C}_b = t\mathbf{A}_b$, Eq. (5) can be unfolded into a neural network with consecutive layers being

$$\mathbf{s}_{k+1}^{(n)} = \eta_{\theta_k}\left(\mathbf{s}_k^{(n)} + \sum_{b=1}^{B}\mathbf{D}_b\left(\mathbf{R}_b^{(n)}\mathbf{Y}_b - \mathbf{C}_b\mathbf{s}_k^{(n)}\right)\right) \tag{7}$$

where $\eta_{\theta_k}(\cdot)$ acts as $\mathbf{soft}_\theta(\cdot)$ in Eq. (5) and can be realized by $\text{ReLu}(\mathbf{s} - \theta_k) - \text{ReLu}(-\mathbf{s} - \theta_k)$. Similarly, introducing $\mathbf{W} = [\mathbf{W}_1, \cdots, \mathbf{W}_B]$ with $\mathbf{W}_b = t\mathbf{A}_b$, Eq. (6) is converted into a multitask reconstruction module:

$$\widehat{\mathbf{X}_b} = \left(\sum_{n=1}(\mathbf{R}_b^{(n)})^{\mathrm{T}}\mathbf{R}_b^{(n)}\right)^{-1}\sum_{n=1}(\mathbf{R}_b^{(n)})^{\mathrm{T}}\mathbf{W}_b\widehat{\mathbf{s}^{(n)}} \tag{8}$$

From Eqs. (5) and (8), $\mathbf{C}$ and $\mathbf{W}$ are decoupled from $\mathbf{D}$ and do not restrictedly obey the relationship in the original optimization. Moreover, different θ_k are used between layers. The main reason is that such settings can accelerate the network convergence [20].

Stacking Eq. (7) for K times formulates a multitask sparse coding module to achieve the sparse representation of observed HSIs. Connecting the multitask sparse coding module with the reconstruction module defined in Eq. (8), we can obtain a multitask sparse neural network (MTSNN) shown in Fig. 1. MTSNN includes B branches and each branch can be regarded as a single denoising task for an individual band. All the branches formulate multitask denoising for all the bands. All the branches interact with each other by producing the same sparse coefficient so as to capture the spectral-spatial correlation among bands. Unlike most "black box" networks whose architectures are obtained by heuristic design requiring exhaustive trials, our MTSNN is derived under the framework of the MTS model and is therefore interpretable. Moreover, discriminatively learning the parameters and image priors form data also bypasses the troublesome hyperparameter tuning and limitations of hand-crafted priors.

2.3. Training and Denoising

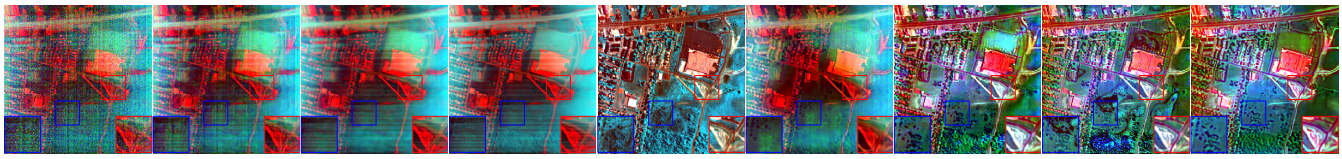
Provided with N pairs of training images $\{\mathbf{Y}^{(n)}, \mathbf{X}^{(n)}\}_{n=1,\cdots,N}$, the loss function is defined as the L_2 loss between estimated HSI and ground-turuth HSI as follows:

$$\mathcal{L}_\Theta = \frac{1}{2N}\|\text{MTSNN}(\mathbf{Y}^{(n)}, \mathbf{\Theta}) - \mathbf{X}^{(n)}\|_F^2 \tag{9}$$

where $\mathbf{\Theta} = \{\mathbf{D}, \mathbf{C}, \mathbf{W}, \theta_1, \cdots, \theta_K\}$ are the network parameters. Once the network is obtained, the denoising problem is a feed-forward process with fixed numbers of iterations, greatly reducing the computational load.

3. EXPERIMENTAL RESULTS

In this section we compared MTSNN with ten methods in the literature, including 7 model-based methods, i.e., BM4D [21], TDL [12], MTSNMF [13], LLRT [10], 3DLogTNN [11], NGMeet [22], and FastHyDe [14] and 3 DL-based methods, i.e., DnCNN [4], SDeCNN [19], QRNN3D [18]. DnCNN is performed in a band by band manner. We tried our best to tune the parameters of model-based methods to the optimal. We could not retrain all the deep learning-based approaches in the ICVL benchmark because of the computational cost, which was too large for our budget. As such, we used the trained models provided by the authors to test all the deep learning-based methods. We only retrained QRNN3D and named the trained model as QRNN3D-R as it achieved state-of-the-art performance. Peak signal-to-noise ratio (PSNR), structure



(a) Noisy (b) BM4D (c) MTSNMF (d) NGMeet (e) FastHyDe (f) SDeCNN (g) QRNN3D (h) QRNN3D-R (i) **MTSNN**

Fig. 2. Denoising results on real-world Urban HSI. The false-color images were generated by combining bands 1, 106, 148.

similarity (SSIM), and spectral angle mapper (SAM) are used to measure the denoising performance quantitively.

3.1. Comparison with the State-of-the-arts

Training Dataset: The training dataset was constructed by selecting 100 HSIs from the ICVL HSI dataset [1]. We extracted subimages size of $64 \times 64 \times 31$ from the original images size of $1392 \times 1300 \times 31$ to train the network. Random flipping, cropping, and resizing were performed to augment the training set. As the noise intensities are different between bands [13, 14], we added different levels of noise to the clean HSIs to generate the noisy-clean HSI pairs. Specifically, the standard noise variances σ are in three ranges, i.e., [0-15], [0-55] and [0-95].

Implementation Details: We first ran SPAMS software [2] on the clean HSIs by setting patch size and the number of atoms respectively as 9×9 and 256, yielding a set of dictionaries $\mathbf{A} = [\mathbf{A}_1, \cdots, \mathbf{A}_B]$ with each size of $\mathbb{R}^{81\times 256}$. $\mathbf{A}$ was then used to initialize the $\mathbf{D}$, $\mathbf{W}$, and $\mathbf{C}$. The initial θ_k was given as 0.02. MTSNN was implemented on the PyTorch platform. We trained the network with two NVIDIA GeForce RTX 3090 GPUs via Adam optimizer whose batch size was set as 8 and the number of epochs was set as 200. The initial learning rate was assigned as 5×10^{-4} and reduced by a factor of 0.35 every 80 epochs. Our code will be accessed via https://github.com/bearshng/MTSNN.

Comparison on Synthetic Data: Table 1 reports the denoising results of all the competing methods on ICVL testing dataset including 50 HSIs. On the whole, our MTSNN ranks the first among all the methods in all the cases with respect to all the indexes, confirming its strong denoising ability. Compared with model-based methods for example MTSNMF, our MTSNN provides dramatically higher denoising effectiveness as MTSNN combines the advantages of model-based and DL-based methods. Moreover, the embedded multitask sparse denoising model also helps MTSNN outperform alternative DL-based methods to a larger margin.

Comparison on Real-world Data: Besides synthetic data, we also tested all the methods on real-wold remote sensing HSI. Despite the differences in spatial and spectral resolution between close-range and remote sensing HSIs, we directly utilize the model trained on the ICVL dataset to test all the DL-based methods so as to examine their generalization ability. We dropped out some methods such as TDL and DnCNN as they produced relatively poor denoising performance. It can be seen from Fig. 2 that our MTSNN provides the best visual denoising performance by removing most of the noises with preserved dominant details, further showing its strong denoising capability and high generalization ability.

[1] http://icvl.cs.bgu.ac.il/hyperspectral/

[2] http://thoth.inrialpes.fr/people/mairal/spams/

3.2. Network Analysis

Here we investigate the key parameters of MTSNN on HSI denoising performance by setting σ in [0-95].

Impact of the Number of Training Samples: Model-based methods are based on the prior knowledge of HSIs and don't heavily rely on the number of training samples. As our MTSNN is based on the MTS model, it should also not require a large number of training samples. To verify this, we change the number of training samples from 20 to 100 and present the denoising performance in Table 2. The more training samples, the better denoising performance. Moreover, changing the number of training samples from 40 to 100 only provides an increase of 0.42 in the PSNR index, showing the strong learning ability and high generalization ability of MTSNN.

Table 2. The impact of the number of training samples.

Samples	20	40	60	80	100
PSNR	38.77	40.05	40.36	40.38	**40.47**

Impact of Network Depth: K is connected with the network depth. A larger K indicates a deeper network. As shown in Table 3, network depth has little impact on the denoising performance of MTSNN. The possible reason is that our MTSNN is obtained by unfolding the iterative optimization of the MTS model. Data-driven training and error backpropagation build the relationship between current and historical optimization, making the denoising performance less dependent on the unfolding times, i.e., K.

Table 3. The impact of network depth.

K	5	7	9	11	13
PSNR	40.49	40.50	40.47	40.53	**40.56**

4. CONCLUSION

This paper introduces a multitask sparse neural network (MTSNN) which is obtained by unfolding the iterative optimization of multitask sparse denoising model. MTSNN includes multiple branches and each branch performs a denoising task for a single band. All the branches are connected by producing the same coefficients. Experimental results and throughout analysis confirm its preferable learning capabality, high generalization ability, and strong denoising capability. We will embed the nonlocal similarity prior into the network to strengthen the denoising ability in our future work.

5. REFERENCES

[1] Longbin Yan, Min Zhao, Xiuheng Wang, Yuge Zhang, and Jie Chen, "Object detection in hyperspectral images," *IEEE SPL*, vol. 28, pp. 508–512, 2021.

[2] Fengchao Xiong, Jun Zhou, and Yuntao Qian, "Material based object tracking in hyperspectral videos," *IEEE TIP*, vol. 29, pp. 3719–3733, 2020.

[3] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE TIP*, vol. 16, no. 8, pp. 2080–2095, 2007.

[4] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE TIP*, vol. 26, no. 7, pp. 3142–3155, 2017.

[5] Yong Chen, Ting-Zhu Huang, Wei He, Xi-Le Zhao, Hongyan Zhang, and Jinshan Zeng, "Hyperspectral image denoising using factor group sparsity-regularized nonconvex low-rank approximation," *IEEE TGRS*, vol. 60, pp. 1–16, 2022.

[6] Fengchao Xiong, Jun Zhou, and Yuntao Qian, "Hyperspectral restoration via l_0 gradient regularized low-rank tensor factorization," *IEEE TGRS*, vol. 57, no. 12, pp. 10410–10425, 2019.

[7] Hongyan Zhang, Wei He, Liangpei Zhang, Huanfeng Shen, and Qiangqiang Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE TGRS*, vol. 52, no. 8, pp. 4729–4743, 2014.

[8] Yao Wang, Jiangjun Peng, Qian Zhao, Yee Leung, Xi-Le Zhao, and Deyu Meng, "Hyperspectral image restoration via total variation regularized low-rank tensor decomposition," *IEEE JSTARS*, vol. 11, no. 4, pp. 1227–1243, 2018.

[9] Qi Xie, Qian Zhao, Deyu Meng, and Zongben Xu, "Kronecker-basis-representation based tensor sparsity and its applications to tensor recovery," *IEEE TPAMI*, vol. 40, no. 8, pp. 1888–1902, 2018.

[10] Yi Chang, Luxin Yan, and Sheng Zhong, "Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *IEEE CVPR*, 2017, pp. 5901–5909.

[11] Yu-Bang Zheng, Ting-Zhu Huang, Xi-Le Zhao, Tai-Xiang Jiang, Tian-Hui Ma, and Teng-Yu Ji, "Mixed noise removal in hyperspectral image via low-fibered-rank regularization," *IEEE TGRS*, vol. 58, no. 1, pp. 734–749, 2020.

[12] Yi Peng, Deyu Meng, Zongben Xu, Chenqiang Gao, Yi Yang, and Biao Zhang, "Decomposable nonlocal tensor dictionary learning for multispectral image denoising," in *IEEE CVPR*, 2014, pp. 2949–2956.

[13] Minchao Ye, Yuntao Qian, and Jun Zhou, "Multitask sparse nonnegative matrix factorization for joint spectralspatial hyperspectral imagery denoising," *IEEE TGRS*, vol. 53, no. 5, pp. 2621–2639, 2015.

[14] Lina Zhuang and Jos M. Bioucas-Dias, "Fast hyperspectral image denoising and inpainting based on low-rank and sparse representations," *IEEE JSTARS*, vol. 11, no. 3, pp. 730–742, 2018.

[15] Fengchao Xiong, Jun Zhou, and Yuntao Qian, "Hyperspectral imagery denoising via reweighed sparse low-rank nonnegative tensor factorization," in *IEEE ICIP*, 2018, pp. 3219–3223.

[16] Guanyiman Fu, Fengchao Xiong, Shuyin Tao, Jianfeng Lu, Jun Zhou, and Yuntao Qian, "Learning a model-based deep hyperspectral denoiser from a single noisy hyperspectral image," in *Proc. IGARSS*, 2021, pp. 4131–4134.

[17] Qiangqiang Yuan, Qiang Zhang, Jie Li, Huanfeng Shen, and Liangpei Zhang, "Hyperspectral image denoising employing a spatial–spectral deep residual convolutional neural network," *IEEE TGRS*, vol. 57, no. 2, pp. 1205–1218, 2018.

[18] Kaixuan Wei, Ying Fu, and Hua Huang, "3-D quasi-recurrent neural network for hyperspectral image denoising," *IEEE TNNLS*, vol. 32, no. 1, pp. 363–375, 2021.

[19] Alessandro Maffei, Juan M. Haut, Mercedes Eugenia Paoletti, Javier Plaza, Lorenzo Bruzzone, and Antonio Plaza, "A single model CNN for hyperspectral image denoising," *IEEE TGRS*, vol. 58, no. 4, pp. 2516–2529, 2020.

[20] Bruno Lecouat, Jean Ponce, and Julien Mairal, "Fully trainable and interpretable non-local sparse models for image restoration," in *ECCV*, 2020, pp. 238–254.

[21] Matteo Maggioni, Vladimir Katkovnik, Karen Egiazarian, and Alessandro Foi, "Nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE TIP*, vol. 22, no. 1, pp. 119–133, 2013.

[22] Wei He, Quanming Yao, Chao Li, Naoto Yokoya, and Qibin Zhao, "Non-local meets global: An integrated paradigm for hyperspectral denoising," in *IEEE CVPR*, 2019, pp. 6861–6870.