HSI-DeNet: Hyperspectral Image Restoration via Convolutional Neural Network

Yi Chang, *Student Member, IEEE*, Luxin Yan, *Member, IEEE*, Houzhang Fang, Sheng Zhong, and Wenshan Liao

***Abstract*— The spectral and the spatial information in hyperspectral images (HSIs) are the two sides of the same coin. How to jointly model them is the key issue for HSIs' noise removal, including random noise, structural stripe noise, and dead pixels/lines. In this paper, we introduce the deep convolutional neural network (CNN) to achieve this goal. The learned filters can well extract the spatial information within their local receptive filed. Meanwhile, the spectral correlation can be depicted by the multiple channels of the learned 2-D filters, namely, the number of filters in each layer. The consequent advantages of our CNN-based HSI denoising method (HSI-DeNet) over previous methods are threefold. First, the proposed HSI-DeNet can be regarded as a tensor-based method by directly learning the filters in each layer without damaging the spectral-spatial structures. Second, the HSI-DeNet can simultaneously accommodate various kinds of noise in HSIs. Moreover, our method is flexible for both single image and multiple images by slightly modifying the channels of the filters in the first and last layers. Last but not least, our method is extremely fast in the testing phase, which makes it more practical for real application. The proposed HSI-DeNet is extensively evaluated on several HSIs, and outperforms the state-of-the-art HSI-DeNets in terms of both speed and performance.**

***Index Terms*— Convolutional neural network (CNN), denoising, destriping, hyperspectral image (HSI) restoration.**

I. Introduction

HYPERSPECTRAL image (HSI) restoration has been a hot topic in the past few years, which includes a variety of classical low-level vision tasks, such as HSIs' denoising, destriping, super-resolution, and so on. In this paper, we focus on various noise removal methods in the HSIs. Numerous works have been proposed for advancing these fields and tremendous progress has been achieved in recent years [1]–[8]. However, there still exist several dark clouds over the HSI restoration filed, which are urgently needed to be solved.

Manuscript received March 29, 2018; revised June 4, 2018; accepted July 14, 2018. Date of publication August 14, 2018; date of current version January 21, 2019. This work was supported in part by the project of the National Natural Science Foundation of China under Grant 61571207, Grant 61433007, and Grant 41501371, and in part by the Hubei Provincial Natural Science Foundation of China under Grant 2018CFA089. *(Corresponding author: Luxin Yan.)*

Y. Chang, L. Yan, S. Zhong, and W. Liao are with the National Key Laboratory of Science and Technology on Multispectral Information Processing, School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: yichang@hust.edu.cn; yanluxin@hust.edu.cn; zhongsheng@hust.edu.cn; wenshanliao@hust.edu.cn).

H. Fang is with the School of Software, Xidian University, Xi'an 710071, China (e-mail: houzhangfang@xidian.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2018.2859203

First, different from the natural 2-D image, the HSIs could deliver additional spectral information of the scenes, resulting in 3-D images. Due to the inherent 3-D tensor format of the HSIs, the previous vector-/matrix-based methods cannot fully exploit the spectral-spatial structural correlation. The way of modeling the spectral-spatial structural correlation in the HSIs contributes much to the final restoration performance. Thus, we raise the first question: *how to jointly utilize both the spatial and the spectral information, and at the same time preserve the spectral-spatial structural correlation intactly for HSI 3-D modeling*?

Second, the noise characteristic in HSIs is complicated, which mainly includes both the random noise and the structural stripe noise, due to the multidetector imaging systems. Such a mixed-noise situation makes the classical Gaussian assumption invalid. Although the elaborated mixture of Gaussian (MoG) methods have been proposed [9], [10], the restoration results are still somewhat unsatisfactory for the stripe noise. Hence, we ask the second question: *should we choose the statistical modeling method directly to the end for HSI noise modeling*?

Third, the size of the HSIs is quite large due to the additional spectral dimension. Most of the previous methods suffer from very long running time, due to the large data size and the complicated operations on HSIs, such as nonlocal patch/cubic searching [11], [12]. This makes the state-of-the-art HSI restoration methods unpractical for real application. Consequently, we present the third question: *could we just use simple and effective models for fast speed*?

In recent years, for the three mentioned problems (data modeling, noise characteristic modeling, and speed), several works have made great progress.

For HSIs' data modeling, existing HSIs' restoration methods can be roughly classified into three categories: 1-D vector-based sparse representation methods, 2-D matrix-based low-rank matrix recovery (LRMR) methods, and 3-D tensor-based approximation methods. The previous two kinds of methods are easy to break the spectral-spatial structural correlation, due to vectorization. The tensor-based methods are naturally proposed to preserve the intrinsic structure correlation with better restoration results, especially the low-rank tensor-based methods in recent two years [11]–[15]. However, the rank of the tensor in the Tucker decomposition-based method is usually defined as the sum of the rank along each of its modes, and still needs to resort to the matrix. The problem has been alleviated, but still exists.

For noise characteristic modeling, there are two kinds of works: image denoising and image decomposition. The first

kind of methods [9], [10] introduced the MoG model to accommodate the distribution of the mixed noise in HSIs. However, for complex mixed noise in HSIs, especially the stripe noise with a similar structure as the image content, these methods cannot well differ the noise from the HSIs. As for the latter method, Chang *et al.* [16] introduced the robust principal component analysis (RPCA) [17] model by regarding the stripe noise as the structural error component to be estimated with Gaussian assumption. Although they are suitable for the Gaussian and the stripe noise, they are less effective for other kinds of mixed noise.

For the running time, the researchers of HSIs have paid less attention to this problem. Most of the optimization-based methods are time-consuming. The most representative method with fast processing time is block-matching and 4-D filtering (BM4D) [18], since it employs the 3-D transform $\rightarrow$ filtering $\rightarrow$ inverse 3-D transform pipeline with very simple operations. However, it is not data-adaptive, and only works well for the random noise but less effective for the stripe noise. In [11], we put forward the idea "making it shorter" by utilizing the key low-rank property in the HSI and discarding the weaker correlation so as to relieve the computation burden. Nevertheless, the running time is still intolerant.

In this paper, we propose a convolutional neural network (CNN)-based HSI restoration method to practically resolve these three problems.

A. HSI Data Modeling

Instead of optimizing a cost function with various unsupervised priors for the HSIs, we treat the HSI restoration task as a discriminative mapping (learning the mapping parameters) problem and present a CNN-based HSIs' restoration model. Naturally, the learned multichannel filters in each layer can simultaneously model the spectral-spatial information explicitly. The spatial structural pattern in HSIs can be well represented by the local receptive filed of the learned filters via the spatial convolution operation, and the spectral correlation in HSIs can be depicted by the learned multichannel filters along the third axis, which in HSIs' restoration we call it a spectral filter. The spectral filter can be regarded as a spectral average operator. Moreover, the degraded HSIs can be directly imported to the CNN model without any further vectorization or matrixization.

B. Noise Modeling

As for the mixed noises, we do not apply any *explicit* expression to fit their distributions, since the noise in the HSIs is too much complicated. Instead, our philosophy is to bypass the difficulty of constructing sophisticated distribution, and resort to a large amount of data. We make use of noisy HSIs to *implicitly* fit the distribution of the noises via the CNN model. Thus, the highly nonlinear fitting capability of the CNN enables our HSI denoising method (HSI-DeNet) to handle arbitrary mixed noise for both the single image and multiple images easily.

C. Running Time

As for the training phase, we employ the residual learning strategy for fast convergence. As for the test phase, the CNN, which mainly contains several convolution and activate layers, is quite suitable for parallel computation on GPU. Given a $512 \times 512 \times 10$ noisy image, it only takes about 20 ms to be processed in the test phase.

As for the architecture of our HSI-DeNet, we mainly apply the residual learning strategy, dilated convolution, and multichannel filtering. We do not directly export the desired clean image, but the residual noisy image. Residual learning has been demonstrated to be very effective in speeding up the training and boosting the final performance both in terms of low-level vision [19]–[21] and mid-level vision [22], [23]. The dilated convolution [24] is used to enlarge the receptive field of the filters in the spatial domain; meanwhile, multichannel filtering is designed for comprehensively capturing the spectral information. Note that our concentration is not about the CNN design, but to demonstrate that the CNN is quite suitable for the HSIs' restoration task.

We further extend the HSI-DeNet into an adversarial framework. We introduce the adversarial subnetwork as a learnable prior. The L_2 loss-based generator subnetwork focuses on the pixel-level modeling benefiting for quantitative assessment [peak signal-to-noise ratio (PSNR) and structure similarity (SSIM)] of the restored image, while the adversarial loss-based discriminator subnetwork aims at the feature-level modeling benefiting for the qualitative assessment (visual appearance) of the restored image. The two terms compete with each other, which is very similar to the fidelity + prior/regularization framework in optimization-based methods. The contributions of the proposed work are summarized as follows.

1) To the best of our knowledge, this is the first work for HSIs' restoration with a fully CNN. The proposed HSI-DeNet can be regarded as the learning multiple channels' 2-D filters, and well preserve the spectral-spatial correlation.
2) We incorporate residual learning, dilated convolution, and multichannel filtering into the network for better modeling the HSIs. Moreover, we explore the adversarial network for HSIs' restoration with better balance in qualitative and quantitative terms.
3) The HSI-DeNet is robust and effective for mixed noise in HSIs, and very flexible for the arbitrary input. Even for the single image overwhelmed with mixed noise, we could also obtain satisfactory result.
4) The proposed method has been tested on extensive HSI data sets with impressive results. Compared with previous methods, our method has achieved faster testing speed and better restoration performance.

The remainder of this paper is organized as follows. The related HSIs' restoration methods are introduced in Section II. Section III presents the concrete architecture of the HSI-DeNet and its adversarial extension. Extensive experimental results are reported in Section IV. Section V concludes this paper.

II. Related Works

The HSI community is indeed highly associated with the development of the computer vision community. To date,

a variety of HSI-DeNets have been proposed in accordance with the most popular tools at that time. In this paper, we classify them into three main categories, and compare them with the proposed method, respectively.

A. Filter-Based Methods

At the beginning of the 21st century, the most powerful representation tool was the wavelet and its variations [25]–[27]. Unsurprisingly, followed by this direction, Othman and Qian [28] proposed a hybrid spatial-spectral derivative domain wavelet shrinkage model with a fixed wavelet dictionary to reduce the noise in HSIs. A generalized multidimensional Wiener filter for denoising is adapted to HSIs [29]. The interested readers could refer to the related works [30]–[32]. However, the main drawback of these methods is that they used the handcrafted and fixed wavelet basis for all HSIs. It has been verified that the learning representation is more powerful than the predefined handcraft representation [33]. Our idea to learn the filters is in line with these filtering-based methods. However, the HSI-DeNet is more adaptive to the HSIs, which facilitates us to learn a more overcomplete representation.

B. Optimization-Based Methods

Most of the HSIs' restoration methods are optimization-based, including 1-D vector-based sparse representation methods, 2-D matrix-based LRMR methods, and 3-D tensor-based approximation methods. The most representative methods for the 1-D method are the total variational (TV) [34] and the dictionary learning [35] method. In 2012, Yuan *et al.* [36] proposed an HSIs' denoising algorithm by employing a spectral-spatial adaptive total variation model. Zhao and Yang [37] introduced an HSI-DeNet by jointly utilizing the sparsity and low-rank property of HSIs in spatial and spectral domains.

With the development of the RPCA [17] and the fast optimization algorithm [38], the 2-D LRMR methods have shown its effectiveness for HSIs' restoration [9], [10], [37], [39]–[42]. By lexicographically ordering the 3-D cube into a 2-D matrix representation along the spectral dimension, Zhang *et al.* [40] and He *et al.* [41] proposed a low-rank matrix restoration model for mixed-noise removal in HSIs. Chang *et al.* [42] proposed a globally low-rank decomposition model for HSIs' destriping, since only parts of data vectors are corrupted by the stripes but the others are not. However, these vector/matrix-based methods inevitably cause damage to the spectral-spatial structural correlation for the 3-D tensor HSIs.

To alleviate this issue, the tensor-based HSI-DeNets have emerged [43]–[45]. In recent two years, when tensor decomposition meets the sparsity property, this direction has yielded the state-of-the-art HSIs' restoration works [11]–[15]. These tensor-based methods substantially improved the HSIs' denoising performance, at the cost of higher computational burden. Our starting point to better preserve the spectral-spatial correlation from the tensor perspective is the same as these methods. While our HSI-DeNet mainly relies on the fully convolutional operation, the previous tensor-based methods still need to resort to the matrix. Compared with the optimization-based methods, our HSI-DeNet shows better performance in running time. Moreover, the optimization-based methods are with the strong assumption to the Gaussian noise. Our work bypasses this assumption by using large training data set to implicitly fit the distribution of the arbitrary noisy input.

C. Learning-Based Methods

Deep learning has been widely used in HSIs' mid-level tasks, such as classification [46], pan-sharpening [47], [48], object detection [49], to name a few. In addition, the CNN has proved its effectiveness in natural image low-level vision tasks, such as denoising [19], super-resolution [50], and so on. Thus, it is natural for us to introduce the CNN to the HSIs' restoration task. Compared with the classical network, our HSI-DeNet has two distinct characteristics. On the one hand, we consider the spatial and the spectral property of the HSIs, and apply the dilated convolution and multichannel filters to model them, respectively. On the other hand, we output the residual noise, and not the clean image. Such a residual learning strategy shows faster training convergence speed.

III. Proposed HSI-DeNet

A. Preliminary for CNN

Assuming there are D layers in the designed network, for a given sample $\boldsymbol{Y} \in \mathbb{R}^{R\times C\times B}$, the output of the first layer is $\boldsymbol{X}^{(1)} = S(\boldsymbol{W}^{(1)} \otimes \boldsymbol{Y} + \boldsymbol{P}^{(1)}) \in R^{R\times C\times B_1}$, where $\boldsymbol{W}^{(1)}$ is the projection matrix to be learned from the first layer, $\boldsymbol{P}^{(1)}$ is the bias vector, $\otimes$ is the convolutional operator, B_1 is the channel number of the first layer, and $S : \mathbb{R} \mapsto \mathbb{R}$ is the nonlinear activation function, which handles each pixel individually, such as the *sigmoid* or rectified linear unit (RELU). Next, the output of the first layer $\boldsymbol{X}^{(1)}$ is treated as the input of the second layer. Consequently, the output of the dth layer can be expressed as

$$\boldsymbol{X}^{(d)} = S(\boldsymbol{W}^{(d)} \otimes \boldsymbol{X}^{(d-1)} + \boldsymbol{P}^{(d)}) \in R^{R\times C\times B_d}. \quad (1)$$

Equation (1) known as the forward procedure is to extract the features from the input data in a hierarchy manner. The visual appearance of the features can be seen in Fig. 1. The goal is to learn the mapping parameters by transforming the degraded data domain to the desired data domain. In conventional model-based methods, the restoration procedure can be formulated in the *maximum a posteriori* framework from the statistical perspective as follows:

$$\begin{aligned}\hat{\boldsymbol{X}} &= \arg\max_{\boldsymbol{X}} p(\boldsymbol{Y}|\boldsymbol{X})p(\boldsymbol{X}) \\ &\propto \arg\max_{\boldsymbol{X}} \{\log p(\boldsymbol{Y}|\boldsymbol{X}) + \log p(\boldsymbol{X})\}. \end{aligned} \quad (2)$$

Each component in (2), such as the noise (posterior term) and the clean data (prior term), is modeled with one specific and explicit distribution. The intuition to transform the degraded data to another domain behind the conventional methods and CNN is similar. For model-based methods, they make an assumption for the data distribution in the transformed domain, such as the gradient domain-based TV prior [34] and the wavelet domain-based Framelet prior [51], [52]. Usually, we can apply the fast optimization method, such

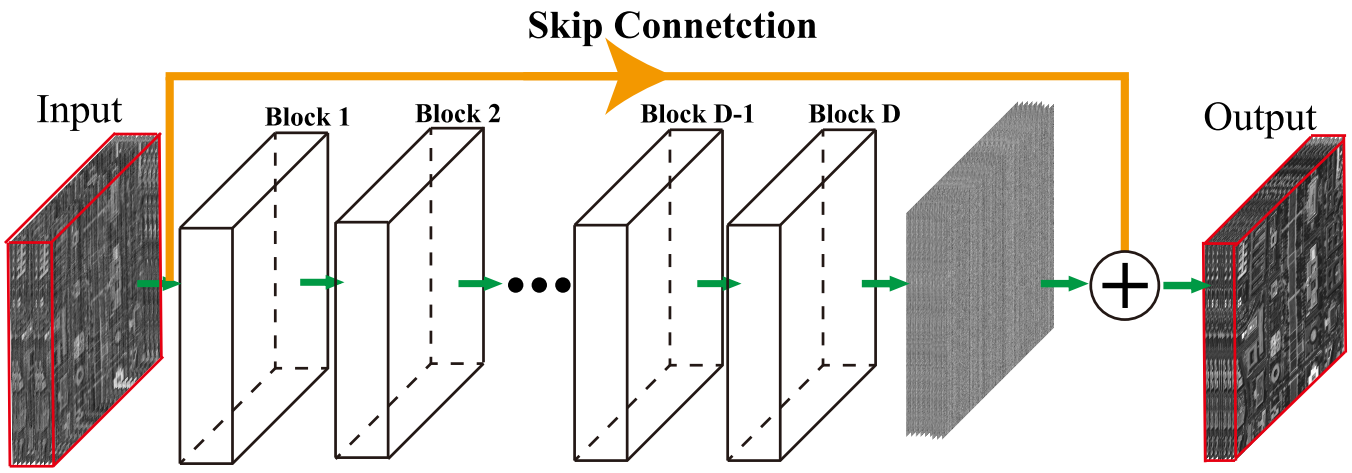


Fig. 1. Architecture of the proposed HSI-DeNet.

as ADMM [38], by introducing the auxiliary variable $\boldsymbol{A}$ to solve (2)

$$\boldsymbol{A}^{(k+1)} = \text{shrink}_{\alpha}(\boldsymbol{D}\boldsymbol{X}^{(k)} + \alpha \boldsymbol{J}^{(k)}) \tag{3}$$

where $\boldsymbol{D}$ is the sparse transformation operator, $\boldsymbol{A}$ is the auxiliary variable that can be approximately equivalent to $\boldsymbol{X}$, $\boldsymbol{J}$ can be regarded as the compensating variation, α is the regularization parameter, shrink_{α} is the soft shrinkage operator, and k is the iteration number.

We can observe that (1) and (3) are very similar to each other. Both of them obtain the desired solution gradually via a linear transformation and then nonlinear activation function. The number of the recursion depends on the depth of the deep model and the iteration of the optimization method. This intrinsic similarity can partially explain why the deep model is also suitable for the image restoration task. However, the transformation parameters in the CNN model are adaptively learned to implicitly fit the distribution of the training data set, which makes them more professional for a specific task.

B. Problem Formulation

The noise degradation model in this paper is formulated as

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N} \tag{4}$$

where $\boldsymbol{Y} \in \mathbb{R}^{R \times C \times B}$ is the measured HSI, R, C, and B stand for the numbers of the row, column, and band respectively, $\boldsymbol{X}$ is the desired clear HSI, and $\boldsymbol{N}$ is the noise in HSI, which includes various noise components. Note that the goal of this paper is to estimate the residual noise component $\boldsymbol{N}$, not the clear image, from the degraded image $\boldsymbol{Y}$. The main reason is that we adopt the residual learning idea from [19] and [22] to train a residual mapping $\mathcal{F}(\boldsymbol{Y}) = \boldsymbol{N}$. The restoration problem is formulated as a regression task as follows:

$$\boldsymbol{J}_{\text{Recon}} = \frac{1}{2}\|\mathcal{F}(\boldsymbol{Y}) - \boldsymbol{N}\|^2 \tag{5}$$

where $\mathcal{F}$ is the composite network mapping function of S.

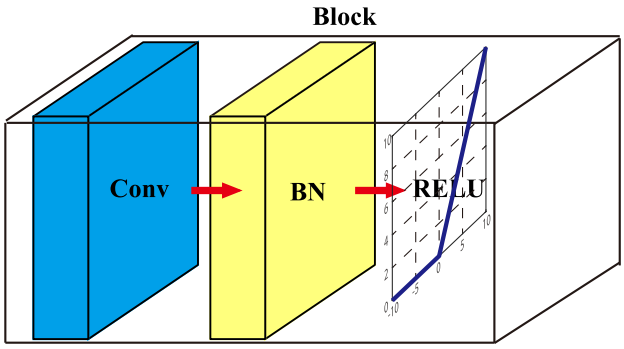


Fig. 2. Illustration of one block. Each block contains the convolution, BN, and nonlinear response.

C. Architecture of HSI-DeNet

In the proposed HSI-DeNet, we use a very deep convolutional network followed by [19], [50], and [53]. Each convolutional layer consists of M_d filters with the size of $3 \times 3 \times N$, except the first and the last output layer. The channel of the first and the last output layer has to match the spectral dimension of the input HSI. We use a 3×3 filters throughout the network with stride 1, which has been demonstrated that the decomposition of larger size filters into small-size filters with deeper layers would make the model more discriminative [19], [20], [22], [50], [53]. To avoid the boundary effect and preserve the spatial size, we pad each layer with the same size as the original image.

The architecture of the HSI-DeNet is shown in Table I. Each block contains three components: convolutional, batch normalization (BN), and RELU, as shown in Fig. 2. We denote the Convolutional(C) + Batch normalization(B) + RELU(R) block as CBR. The depth D of HSI-DeNet is 19 (including the $\mathcal{L}_2$ loss layer). The main reasons for us to choose the depth as 19 are threefold. On the one hand, the depth in the CNN model is similar to the iteration number in optimization-based methods. Many works [20], [54] have discussed their relationship and design their deep architecture based on the optimization solvers. Since the iteration number of the nonconvex problem

TABLE I
DETAILED ARCHITECTURE DESCRIPTION OF THE HSI-DENET

Layer	CR	CBR1	CBR2	CBR3	CBR4	CBR5	CBR6	CBR7	CBR8
Kernel Size	3*3	3*3	3*3	3*3	3*3	3*3	3*3	3*3	3*3
Filter Number	64	64	64	64	128	128	128	256	256
Dilation	1	1	1	1	1	1	1	2	2
Receptive Field	3	5	7	9	11	13	15	19	23
Image Size	40*40*64	40*40*64	40*40*64	40*40*64	40*40*128	40*40*128	40*40*128	40*40*256	40*40*256
Layer	CBR9	CBR10	CBR11	CBR12	CBR13	CBR14	CBR15	CBR16	C
Kernel Size	3*3	3*3	3*3	3*3	3*3	3*3	3*3	3*3	3*3
Filter Number	256	128	128	128	64	64	64	64	10
Dilation	2	1	1	1	1	1	1	1	1
Receptive Field	27	29	31	33	35	37	39	41	43
Image Size	40*40*256	40*40*128	40*40*128	40*40*128	40*40*64	40*40*64	40*40*64	40*40*64	40*40*10

is usually determined empirically, we also observe from the experimental results that the depth 19 is very robust for our HSI restoration task. On the other hand, we follow the widely used VGG-nets [53] to set the depth of our HSI-DeNet as 19. This depth can achieve well balance between training difficulty and representation ability. Last but not least, our training image size is 40×40, which is slightly smaller than the receptive field of our model. This means our model could utilize the whole contextual information of the given HSIs.

The BN layer is incorporated for avoiding the gradient vanishing or divergence issue. And the RELU layer is utilized for pursuing sparsity and also for its highly nonlinear ability. Note that we do not contain any pooling layer in our network. The main reason is that the image processing task is a regression task, while the high level vision tasks are always classification-based tasks. In image processing, we need to estimate the pixel-level information. However, the pooling layer would inevitably cause information loss. Therefore, in the HSI-DeNet, we do not apply any pooling layer.

1) Residual Learning: Previous deep learning-based image processing methods directly mapped the degraded image into the clean one [55]–[57]. However, the gradient vanishing issue restricts these methods to train a very deep model with powerful representation. In this paper, we added a skip connection between the input and the output, which means that the network actually learns the difference between the input and the output. This residual learning scheme proposed by He *et al.* [22] figured out creative way to learn the sparse residual image, not the image itself, since the sparser gradient of the residual image was easier to be propagated. The loss function of our network is $1/2\times\|\boldsymbol{N}-f(\boldsymbol{Y})\|$, where f is the network mapping function. Several pioneer works have demonstrated its effectiveness in various fields [19]–[21], [23], [50]. Therefore, it is natural for us to apply the residual learning in HSIs' restoration.

2) Dilated Convolution and Multichannel Filtering: It is known that the more the contextual information CNN models, the better the restoration results they obtain [50]. Modern networks integrate multiscale contextual information, namely, enlarge the receptive field of the network via designing deeper layers [19], [53]. However, this may increase the difficulty of the training procedure due to the gradient vanishing issue. In this paper, we introduce the dilated convolution [24] into the middle layers (Layer CBR7, 8, 9 with dilation 2), which aggregates multiscale contextual information without losing resolution or increasing the depth of the network. We can observe from Table I that the receptive field of our model is larger than the size of the image, which means we could completely make use of both the local and the nonlocal contextual information.

The classical networks, such as Alexnet [33], VGGnet [53], and U-net [58], all employed the multichannel filtering strategy, in which each layer contains multiple feature maps. These models increased the number of the filters at the first few layers, and then reduced the number of the filters gradually. Such a flexible manner greatly increases the representation ability of the network. For HSIs with multiple spectra (corresponding to multichannel in the network), multichannel filtering in each layer becomes much more important, which undoubtedly increases the representation ability for the spectral information. As shown in Table I, we gradually increase the channels from 10, 64, and 128 to 256, and then decrease it symmetrically from 256, 128, and 64 to 10.

3) Training Details: We initialize the convolutional filters with the Xavier method [22]. The learning rate is initially set as 0.001 and decreased to a small value 0.00001. The momentum and decay are fixed as 0.9 and 0, respectively. The ADAM solver [59] is introduced to optimize the model. We trained the model with 300 epochs with the batch size as 128. We obtained the training data from the ICVL [60], which includes 201 scenes. We cropped 500 subimages from them as the training data set and 50 subimages as the test data set. The training data were normalized to [0, 1]. Since the earlier image bands (mainly from 400 to 450 nm) in ICVL contain random noises, we just extracted the band 550–640 nm with the interval 10 (namely ten bands) as the input. Note that the training samples in our model are 500 (180×180) images. However, since the CNN does not require fixed inputs, we extracted 16 (40×40) subsamples via the sliding window with stride 40 from each sample. Then, we augmented each subsample eight times with flip and rotation. Therefore, the total training samples ($40\times40\times10$) in our experiment

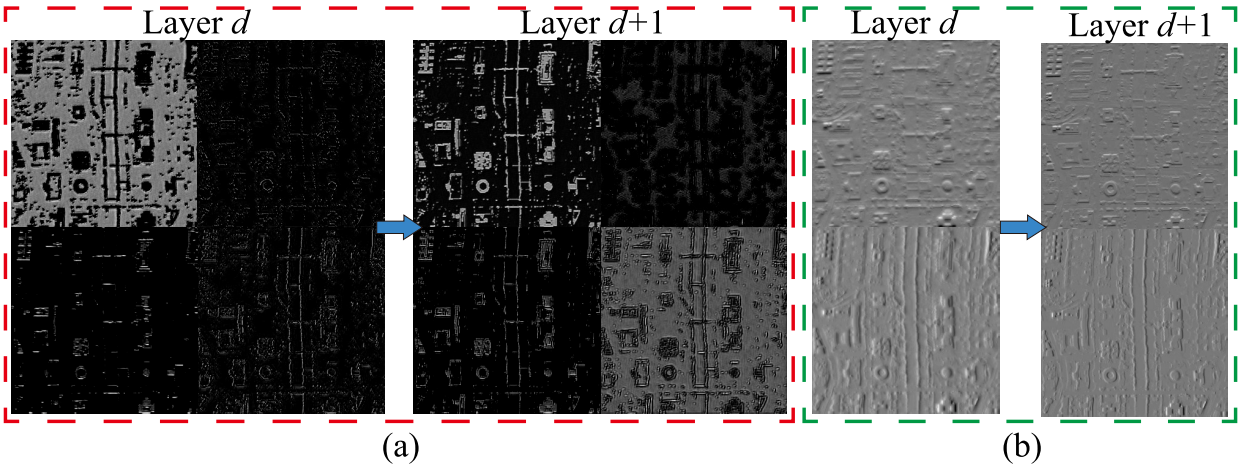


Fig. 3. Comparison between the features extracted by (a) HSI-DeNet and (b) TV methods. It is evident that in each iteration/layer, the sparsity-based shrinkage methods (here, we just give an example with TV) utilize the same features, while the HSI-DeNet could extract the hierarchical features.

are $500 \times 16 \times 8 = 64\,000$. If we further consider each band as an image (compared with single image-based CNN models), the final training samples would be regarded as $64\,000 \times 10 = 640\,000$. That is the main reason why our training model works well with only 500 samples. In our experiment, we did not observe obvious difference for the training data ranging from 300 to 700. The MatConvnet toolbox [61] is employed to train the HSI-DeNet.

4) Relationship With Previous Methods: Comparing with the filtering-based methods, such as the wavelet, the learned filters in the HSI-DeNet could be regarded as its data adaptive version. The previous handcraft wavelets could only capture the specific image structures, such as horizontal, vertical, and diagonal information. While hundreds of various filters in HSI-DeNet are more representative for the HSI structure, the additional nonlinear response function further increases its generalization ability.

Comparing with the sparsity-based optimization methods, such as dictionary learning [35], TV [62], and low rank [40], their relationships with the CNN have been discussed in [20] and [54]. For example, the iteration of the shrinkage can be unfolded and regarded as the nonlinear response in the network. As shown in Fig. 3, in each phase (iteration in optimization methods and layers in CNN, respectively), the optimization-based methods employ the same transformation for the image, while the CNN extracts the different scale and directional information gradually. Thus, we can roughly regard the conventional methods as the shallow model, while the CNN is the deep model with stronger representation.

D. Extension to Adversarial Network

The generator networks optimized based on the pixelwise loss function, such as the mean squared error or $\mathcal{L}_1$, tend to produce the oversmooth results [63], [64], since they mainly focus on the pixelwise-level image differences. To exploit priors from the image level, we introduce the adversarial discriminators for the generator. The generative network (described in Section III-C) can be further incorporated in an adversarial framework, by additionally incorporating with a discriminator. An adversarial network [65] is a recent approach that has shown remarkable performances to generate synthetic photorealistic images. Generative adversarial networks (GANs) are composed of two models that are alternatively trained to compete with each other. The generator is trained to produce the true data distribution p_{data}, so that the generated images are difficult for the discriminator to differentiate from real images. Meanwhile, the discriminator served as a classifier to distinguish fake images generated by G from real images. The objective function of the adversarial networks is expressed as follows:

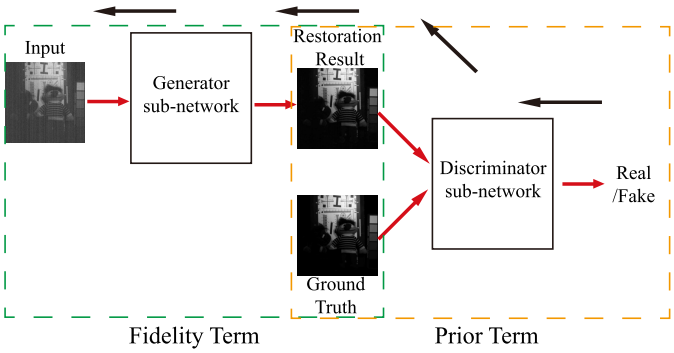


Fig. 4. Flowchart of the proposed HSI-DeGAN.

$$\boldsymbol{J}_{\text{adver}} = \mathbb{E}_{X\sim P_X}[\log D(\boldsymbol{X})] + \mathbb{E}_{\boldsymbol{Y}\sim P_Y}[\log(1 - D(G(\boldsymbol{Y})))] \tag{6}$$

where $P_{\boldsymbol{Y}}$ and $P_{\boldsymbol{X}}$ denote the distributions of noise input $\boldsymbol{Y}$ and real data $\boldsymbol{X}$, G denotes the generator, and D represents the discriminator.

We show the general flowchart of the proposed method in Fig. 4. The proposed HSI-DeGAN mainly contains two parts: the generator and the discriminator subnetwork. The red arrow denotes the forward procedure, and the black arrow indicates the backward propagation. The left green rectangular box is the L_2 reconstruction-based generator (Section III-C), which

TABLE II
DETAILED DESCRIPTION OF OUR DISCRIMINATOR

Layer	CR	CBR1	CBR2	C
Kernel Size	4*4	4*4	4*4	5*5
Filter Number	64	128	256	1
Stride	2	2	2	2
Image Size	20*20*64	10*10*128	5*5*256	1*1*1

can be regarded as the fidelity term in the optimization method, since both of them are used to constrain the restored result that should be consistent with the observation. The right orange rectangular box is the adversarial perceptual loss-based discriminator, which can be equivalent to the prior/regularization term for further solution refine. Here, the discriminator makes the distribution between the restored image and the ground true indistinguishable, which can be regarded as the adaptively learned prior for the data.

The input to the discriminator is a pair of images: a ground truth image and a restored image by the generator. The output of the discriminator is a binary value for the ground truth image to be one and the restored image to be zero. The detailed architecture of the discriminator is shown in Table II. Here, the dilation of each layer is set as 1. We gradually decrease the image size by introducing the stride. The whole loss function of our network is defined as

$$\boldsymbol{J}_f = \lambda_1 \boldsymbol{J}_{\text{Recon}} + \lambda_2 \boldsymbol{J}_{\text{Adver}} \tag{7}$$

where λ_1 and λ_2 are the weights to balance the effects of different losses. The generator and the discriminator can be trained adversarially/alternatively. The discriminator aims to maximize the above objective, while the generator tries to minimize the objective. Note that, in the testing phase, only the generator network is required to obtain the clear image that is as natural as the real image.

IV. EXPERIMENTAL RESULTS

A. Experimental Setting

We use the ICVL [60] as our training data set for all tasks. The test data sets include the CAVE and ICVL. The ICVL database images were acquired using a Specim PS Kappa DX4 hyperspectral camera and a rotary stage for spatial scanning. At this time, it contains 201 images and will continue to grow progressively. Images were collected at 1392 × 1300 spatial resolution over 519 spectral bands (400–1000 nm at roughly 1.25-nm increments). The CAVE database images were acquired using a Cooled CCD camera (Apogee Alta U260). The database consists of 32 scenes at 512 × 512 spatial resolution from 400 to 700 nm at 10-nm steps (31 bands total).

For the HSI-DeNets, we compare with block-matching and 3-D filtering (BM3D) [66], parallel factor analysis (PARAFAC) [44], LRMR [40], BM4D [18], tensor dictionary learning (TDL) [14], intrinsic tensor sparsity regularization (ITSReg) [15], and Laplacian regularized low-rank tensor recovery (LLRT) [11]. We use the codes provided by the authors downloaded from their homepages, and fine tune the parameters by default or following the rules in their papers to achieve the best performance. And once our manuscript has been accepted, the training and testing code of our methods can be downloaded from the homepage of the author.[1]

The PSNR, SSIM [67], erreur relative globale adimensionnelle de synthese (ERGAS [68]), and spectral angle map (SAM [69]) are employed for the quantitative assessment. The PSNR and SSIM evaluate the spatial quality, and the ERGAS and SAM assess the spectral quality. The larger the PSNR and SSIM values are and the smaller the ERGAS and SAM values are, the better the restored images are.

B. Simulated Noise Removal

We test four noisy cases: the random noise, the stripe noise, the mixed random and stripe noise, and the single image.

1) Random Noise Removal: Here, we give two examples degraded with the additive Gaussian random noise with zero mean and different variances. The BM3D and BM4D [Figs. 5 and 6(c) and (f)] tended to introduce the unexpected ringing artifacts. The LRMR [Figs. 5 and 6(e)] suffered from the residual noise, since it has strong assumption on the low-rank constraints of the multispectral inputs. The low-rank tensor-based TDL, ITSReg, and LLRT worked well in light noise case [Fig. 5(g)–(i)], while for heavy noise case [Fig. 6(g)–(i)], they either left residual noise or oversmoothed the HSI. In Figs. 5 and 6, the HSI-DeNet could remove the noise satisfactorily and preserve clear details, such as the ellipse region in Fig. 5 and rectangle regions in Fig. 6.

Moreover, with the increase in the noise level, the HSI-DeNet obtained much more advantageous over other methods in terms of the quantitative assessments. The main reason is that, when the noise is overwhelmed in the HSIs, the local, nonlocal self-similarity or spectral correlation has been severely damaged. Thus, the performance of previous methods inevitably decreased rapidly. On the contrary, the learning-based HSI-DeNet could infer the missing information from the external data set, also benefitting from its intrinsic tensor-based structural preserving ability. The HSI-DeNet consistently obtained the best results for both the spectral and spatial assessments. This demonstrated that our HSI-DeNet could better preserve spectral integrity due to the tensor-level operation, and also better preserve spatial structural details due to the learned information from the external data set.

2) Stripe Noise Removal: As far as we know, most of the aforementioned HSI denoising cannot handle the stripe noise. We compared the HSI-DeNet with the state-of-the-art destriping methods: TV [34], unidirectional variational model (UTV) [62], wavelet-Fourier adaptive filtering (WFAF) [70], statistical linear destriping (SLD) [71], low-rank multiple image decomposition (LRMID) [42], and anisotropic spectral-spatial total variational (ASSTV) [3]. From Fig. 7(c)–(e) and (g), we can see obvious unexpected artifacts. For the marked eclipse region of the unidirectional model [Fig. 7(f) and (h)], the original linear pattern with

[1] http://www.escience.cn/people/changyi/index.html

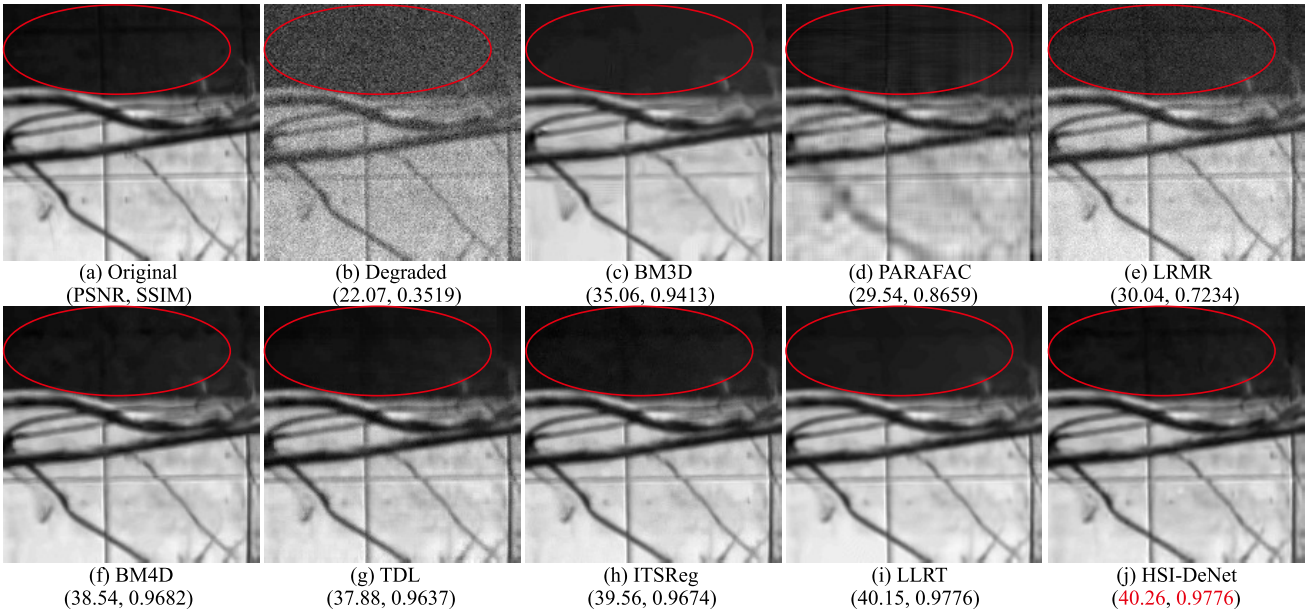


Fig. 5. Simulated random noise removal results at 550-nm band under the noise level $\sigma = 20$ on the CVIL data set. We show the PSNR value and SSIM results. (a) Original image. (b) Degraded image. Denoising results by (c) BM3D, (d) PARAFAC, (e) LRMR, (f) BM4D, (g) TDL, (h) ITSReg, (i) LLRT, and (j) HSI-DeNet.

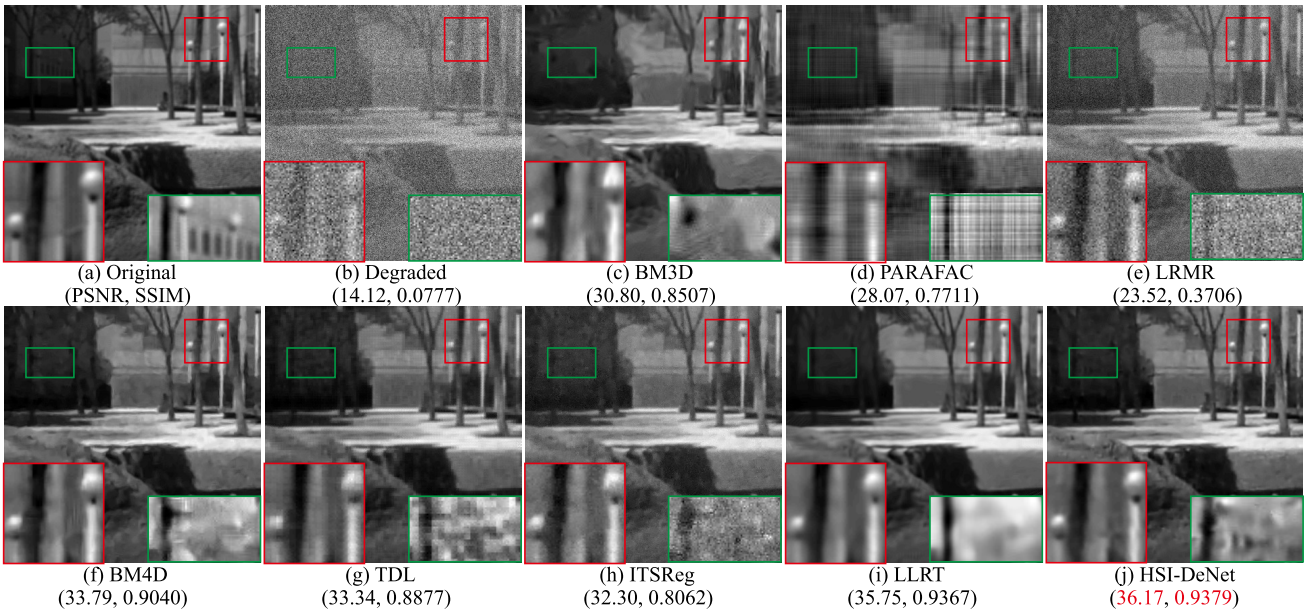


Fig. 6. Simulated random noise removal results at 550-nm band under the noise level $\sigma = 50$ on the CVIL data set. (a) Original image. (b) Degraded image. Denoising results by (c) BM3D, (d) PARAFAC, (e) LRMR, (f) BM4D, (g) TDL, (h) ITSReg, (i) LLRT, and (j) HSI-DeNet.

the same direction as the stripe has also been removed. The result of our HSI-DeNet is with satisfactory visual appearance. Moreover, the image structure has been preserved intactly. This result strongly demonstrates that the CNN has more powerful representation, which could better distinguish the stripe from the image structure.

3) Mixed-Noise Removal: The random noise and the stripe noise always co-exist in the real HSI. In this section, we test the performance of our HSI-DeNet under the mixed-noise case. The results are shown in Fig. 8. We observe an interesting phenomenon that the state-of-the-art HSI restoration method could only handle the random noise and alleviate the stripe noise. However, they cannot totally remove the stripe noise, as shown in Fig. 8(c)–(i). On the contrary, in Fig. 8(j), the HSI-DeNet could simultaneously remove both the random and the stripe noise with better visual appearance. This demonstrates that explicit modeling of the mixed noise is inappropriate. Implicit modeling via the CNN offers a new perspective for this complex problem. Moreover, their weak effect to the stripe noise heavily depends on the spectral

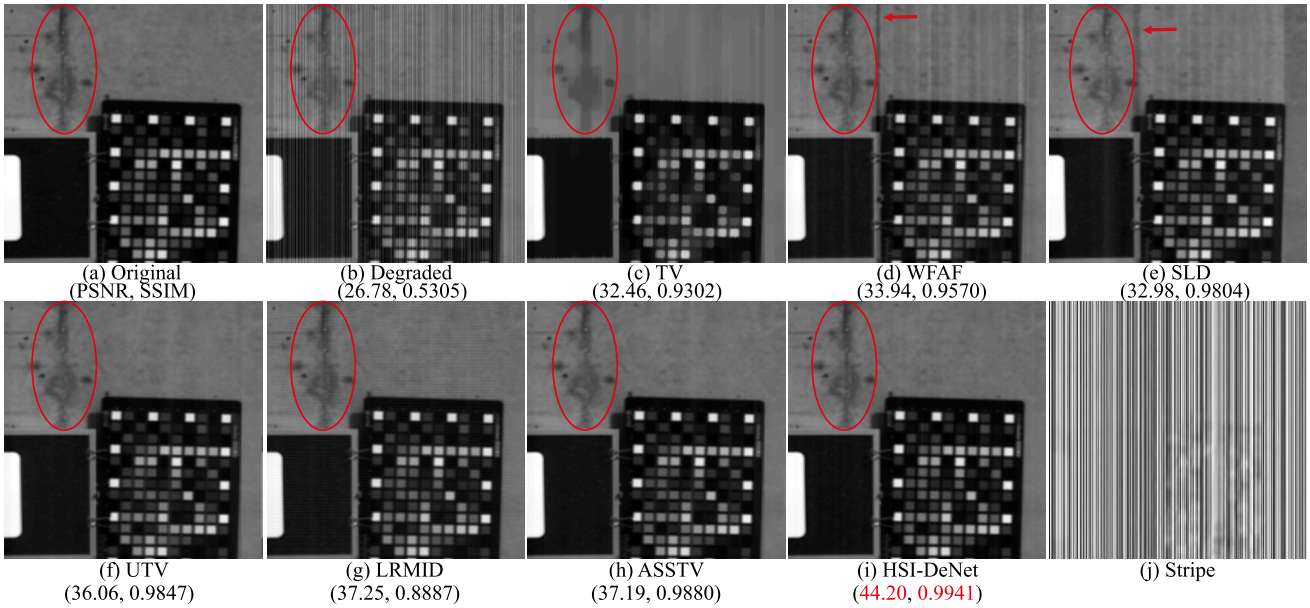


Fig. 7. Simulated nonperiodical stripe noise removal results at 550-nm band under the stripe intensity $S = 20$ on the CVIL data set. (a) Original image. (b) Degraded image. Destriping results by (c) TV, (d) WFAF, (e) SLD, (f) UTV, (g) LRMID, (h) ASSTV, and (i) HSI-DeNet. (j) Estimated stripe component.

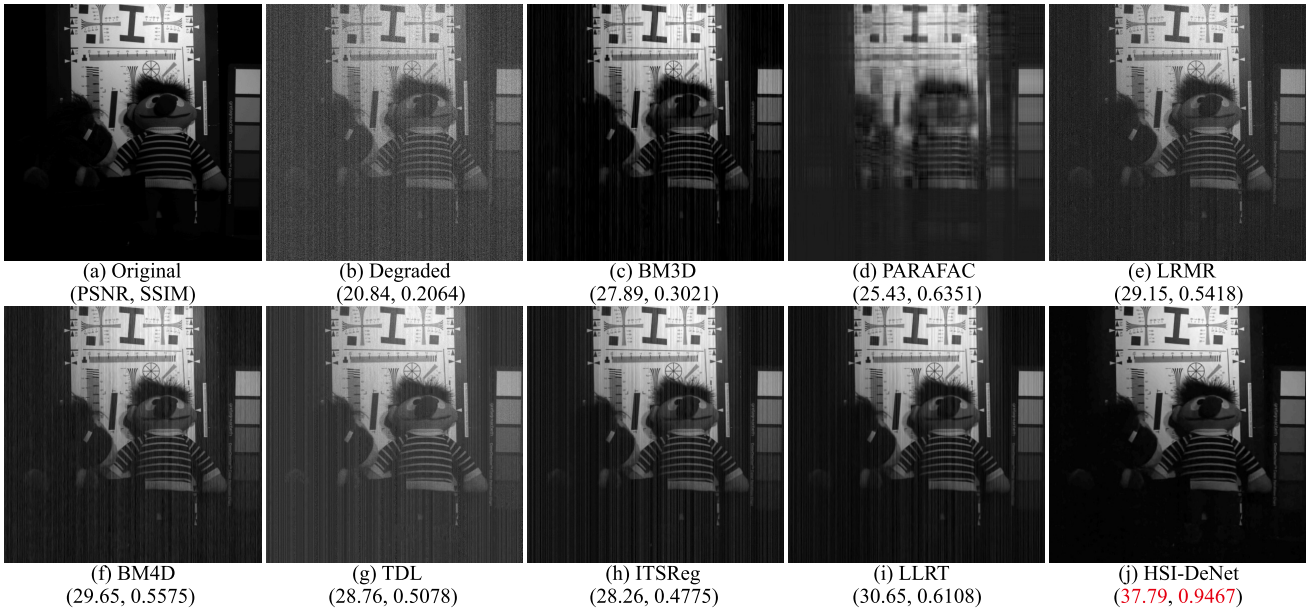


Fig. 8. Simulated mixed-noise removal results at 550-nm band of image *Toy* under the noise level $\sigma = 20$ and $S = 20$ on the CAVE data set. (a) Original image. (b) Degraded with mixed noise. Restoration results by (c) BM3D, (d) PARAFAC, (e) LRMR, (f) BM4D, (g) TDL, (h) ITSReg, (i) LLRT, and (j) HSI-DeNet.

correlation in HSIs. For the mixed noise in the single image, previous methods even do not consider this situation. We show in Section IV-B4 that the HSI-DeNet could squeeze more spatial information and perfectly handle it.

4) Single-Image Mixed-Noise Removal: To the best of our knowledge, few works have considered the single-image mixed-stripe and random noise removal issue, except the work [72]. Such a mixed-noise removal problem is extremely difficult. For one thing, only the spatial information is provided. For another, both the random noise and the stripe noise should be modeled simultaneously. For the conventional optimization-based method, complex and elaborate designing of the cost function is usually required, which makes transformed low-rank (TLR) [72] hard to be optimized with very long testing time. In this paper, we show that the HSI-DeNet could easily remove the mixed random and stripe noise in a single image, as shown in Fig. 9. The result of HSI-DeNet is more visual pleasure with higher PSNR and SSIM values. More importantly, the TLR took several minutes for this result, while the HSI-DeNet obtained the result within less than 1 s.

5) Comparison With Deep Method: Since there are few deep models for HSIs' restoration, we compare our HSI-DeNet

TABLE III

SPATIAL AND SPECTRAL QUANTITATIVE ASSESSMENTS OF DIFFERENT METHODS UNDER DIFFERENT NOISE LEVELS. THE R20_S20 MEANS THE SIMULTANEOUS RANDOM NOISE WITH VARIANCE 20 AND STRIPE NOISE WITH INTENSITY 20

Sigma	Index	Methods								
		Noisy	BM3D	PARAFAC	LRMR	BM4D	TDL	ITSReg	LLRT	HSI-DeNet
20	PSNR	22.07	35.06	29.54	30.04	38.54	37.88	39.56	40.15	**40.26**
	SSIM	0.3519	0.9413	0.8658	0.7234	0.9682	0.9637	0.9674	**0.9776**	**0.9776**
	SAM	0.1726	0.0258	0.0156	0.0390	0.0157	0.0100	0.0112	0.0093	**0.0088**
	ERGAS	176.25	39.48	74.52	72.19	26.52	28.68	23.57	**21.84**	22.02
50	PSNR	14.12	30.80	28.07	23.52	33.79	33.34	32.30	35.75	**36.17**
	SSIM	0.0777	0.8507	0.7711	0.3706	0.9040	0.8877	0.8062	0.9367	**0.9379**
	SAM	0.5604	0.0681	0.0200	0.0620	0.0350	0.0191	0.0514	0.0203	**0.0168**
	ERGAS	539.87	79.04	108.33	183.29	56.31	59.14	66.58	45.00	**42.66**
R20_S20	PSNR	20.84	27.84	25.43	29.15	29.65	28.76	28.26	30.65	**37.79**
	SSIM	0.2064	0.3021	0.6351	0.5418	0.5575	0.5078	0.4775	0.6108	**0.9467**
	SAM	1.0120	0.6017	0.7734	0.7625	0.7874	0.8028	0.8322	0.8029	**0.2212**
	ERGAS	375.62	163.80	260.09	149.66	139.77	159.15	161.04	148.49	**65.05**

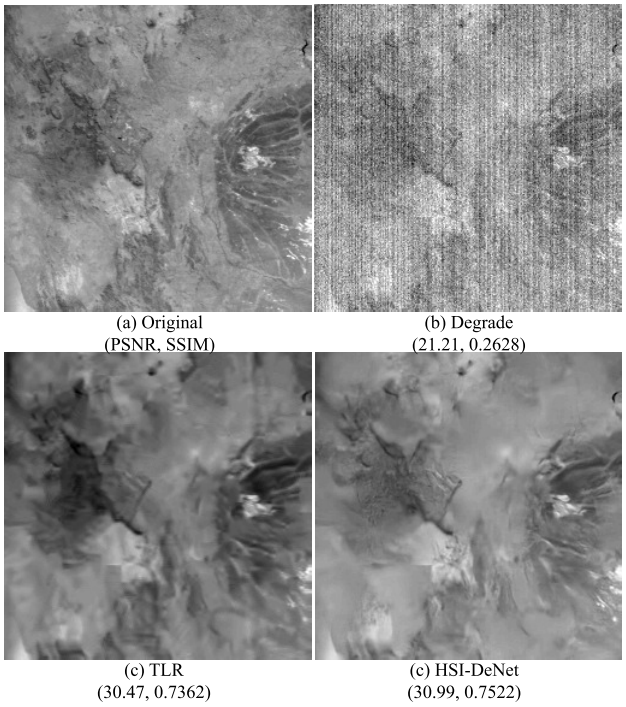
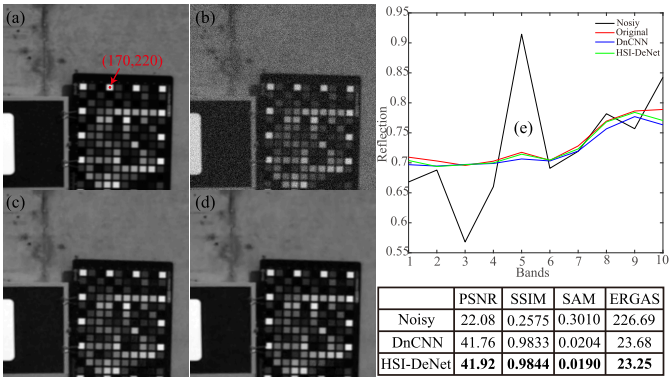


Fig. 9. Simulated mixed-noise removal results in single image under noise level $\sigma = 20$ and $S = 20$. (a) Original image. (b) Degraded with mixed noise. Restoration results by (c) TLR and (d) HSI-DeNet.



	PSNR	SSIM	SAM	ERGAS
Noisy	22.08	0.2575	0.3010	226.69
DnCNN	41.76	0.9833	0.0204	23.68
HSI-DeNet	**41.92**	**0.9844**	**0.0190**	**23.25**

Fig. 10. Quantitative and qualitative comparison between DnCNN and HSI-DeNet. (a) Original image. (b) Noisy image. (c) DnCNN. (d) HSI-DeNet. (e) Spectral reflection of one pixel at (170, 220). The table shows the quantitative results.

with DnCNN [19], which is a single image-based method for the natural image. We slightly change the input and output channels of the model to accommodate the HSIs. For the fair comparison, both of our method and modified DnCNN are trained with the same data set on ICVL. The hyperparameters are set as the rule in the original paper. Here, we give both the visual and quantitative comparison.

We compare the denoising results on ten ICVL subimages with the size of $380 \times 380 \times 10$. We show one image as an example in Fig. 10. From the spatial visual performance, our method and DnCNN are quite similar to each other. From the spectral features, the spectral curve of our method is obviously closer to that of the original one. We could also obtain the same conclusion from the quantitative assessments. It is worth noting that, due to the additional multichannel strategy, our model is more representative for the spectral features. That is the main reason why our model is more suitable for HSIs' restoration task. And along with the dilated convolution, our method also obtains better performance in the spatial structural reconstruction.

6) Quantitative Comparison: We show the quantitative comparison results of Figs. 5 and 6 in Table III. We can observe that the HSI-DeNet obtains the best results under different noise levels. Moreover, with the increase in the noise level, the advantages of HSI-DeNet over other methods are more. The main reason is that the other methods all utilize the local or nonlocal information from the degraded image itself. When the noise level increases, the internal information has been heavily damaged, resulting in the degeneration of their performances. On the contrary, our HSI-DeNet could

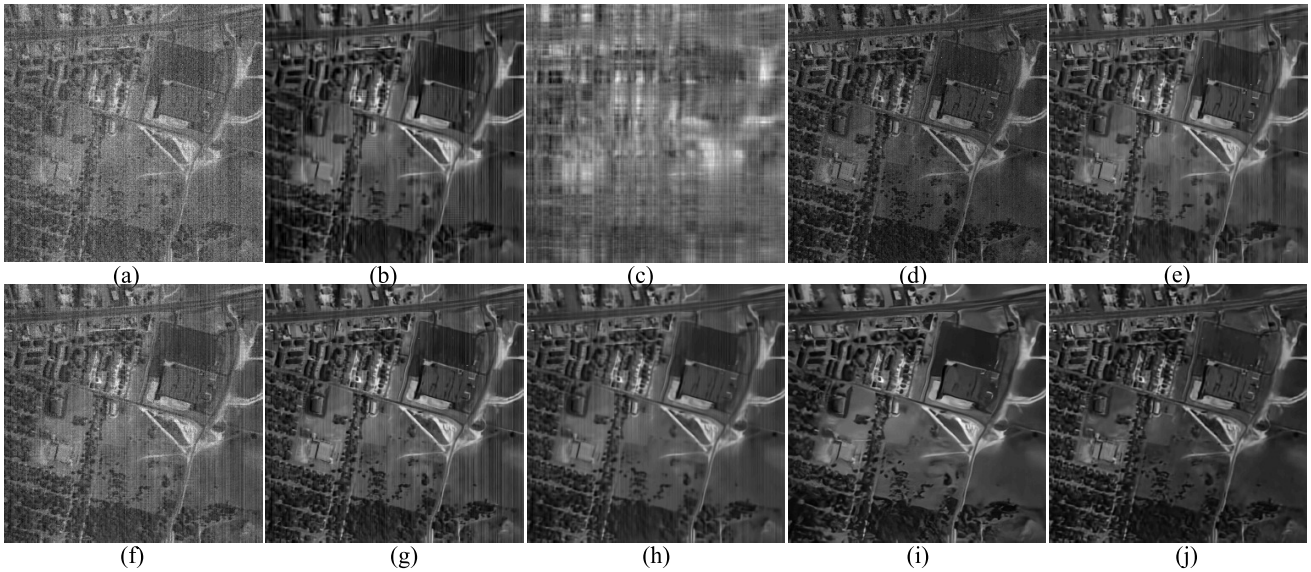


Fig. 11. Real HSI image *urban* noise removal results. (a) Band 138. Restoration results by (b) BM3D, (c) PARAFAC, (d) LRMR, (e) BM4D, (f) TDL, (g) ITSReg, (h) LLRT, (i) HSI-DeNet with single band as input, and (j) HSI-DeNet with multiple bands as input.

benefit from the external data set for better restoration performance.

C. Real Noise Removal

Here, we select the widely used *urban* data set to test the performance of the HSI-NeDet in real HSIs. *Urban* is one of the most widely used hyperspectral data. There are 307×307 pixels, each of which corresponds to a 2×2 m^2 area. In this image, there are 210 wavelengths ranging from 400 to 2500 nm, resulting in a spectral resolution of 10 nm. We extract ten bands (bands from 129 to 138 as the input). Only a few bands in the *urban* data set are affected by slightly random and stripe noise (mostly horizontal). To better show the results, we increased the noise level by adding both the random and stripe noises (vertical) in all the bands, as shown in Fig. 11(a). It is worth noting that the result of Fig. 11(i) is corresponding to the HSI-DeNet with only the single image band 138 as the input. The other methods are with 10 image bands as the input. From Fig. 11, we have two main observations. First, our single image-based [Fig. 11(i)] and multiple image-based [Fig. 11(j)] HSI-DeNet all obtain the best visual appearance, while the results of other methods are with unexpected artifacts, such as the stripe residual in Fig. 11(b), (f), (g), and (h). Second, the result of Fig. 11(i) is a bit oversmoothed, compared with Fig. 11(j). This is a powerful proof that our HSI-DeNet could benefit from the spectral correlation in HSIs.

Furthermore, in Fig. 12, we show that our HSI-DeNet could also be applied to other remote sensing images, such as moderate resolution imaging spectroradiometer (MODIS). Here, we use the MODIS Terra image band 30 as the test image [Fig. 12(a)]. Moreover, from the satisfactory result in Fig. 12(b) and also in Fig. 11(i), we could conclude that the HSI-DeNet fully utilizes the spatial information and benefits from the learned pattern from the external data set. This might give a new insight for conventional model-based HSI restoration methods.

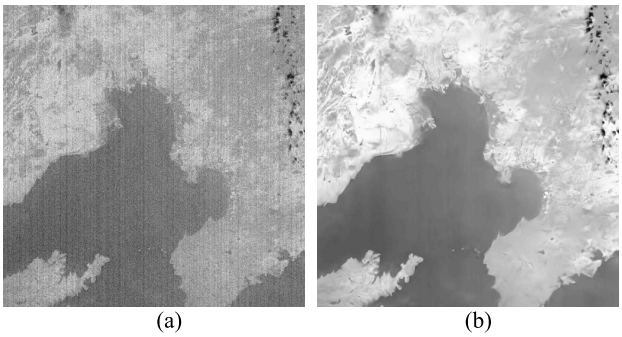


Fig. 12. Extension to the MODIS image mixed-noise removal. (a) Terra MODIS image band 30. (b) HSI-DeNet.

D. Study of HSI-DeNet

1) Effectiveness of the Discriminator: We performed an experiment to validate the effectiveness of the discriminator. As shown in Fig. 13, we compared the denoising result of HSI-DeNet [Fig. 13(d)] with its corresponding GAN [Fig. 13(e)]. Compared with other methods, the HSI-DeGAN obtained the image with the sharper edge and fewer artifacts from the visual appearance. However, the quantitative assessments of HSI-DeGAN are a little inferior to other methods. This is the main characteristic of the adversarial network that it could obtain a photorealistic image but a poor quantitative result [63], [64].

2) Effectiveness for Postprocessing: We use the unsupervised k-mean classification method to demonstrate that the proposed method can facilitate the subsequent processing. The number of the classification is five. The maximum iteration is set five times. Fig. 14 shows the denoising and classification results. The top row is the degraded *Salinas* image

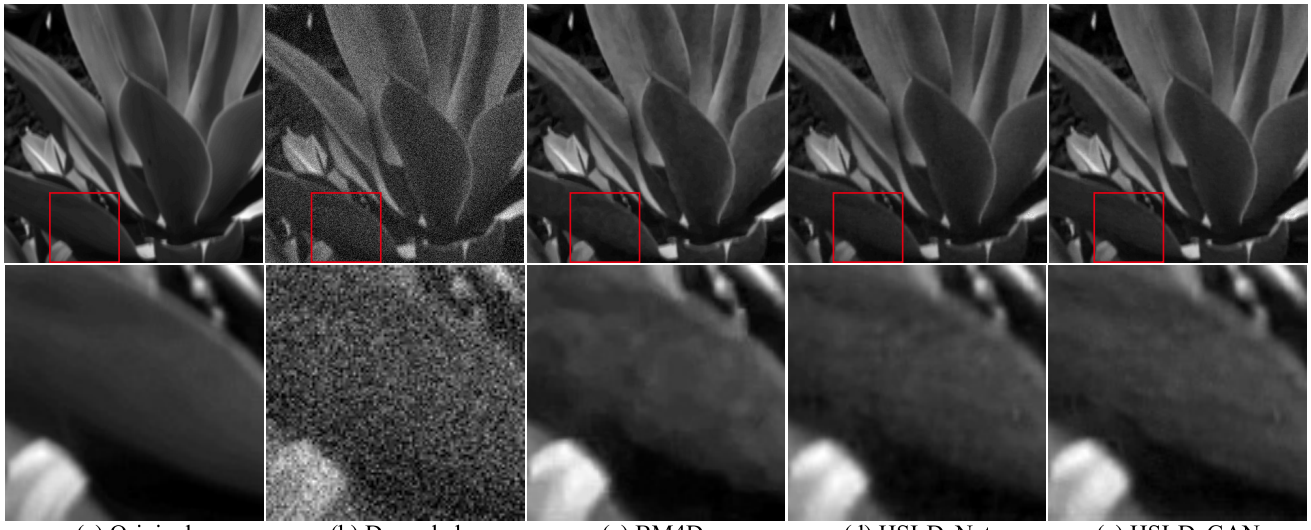


Fig. 13. Effectiveness of the discriminator. (a) Original image. (b) Degraded image. Denoising results by (c) BM4D, (d) HSI-DeNet, and (e) HSI-DeGAN. (First row) Whole image. (Second row) Zoomed-in-view results.

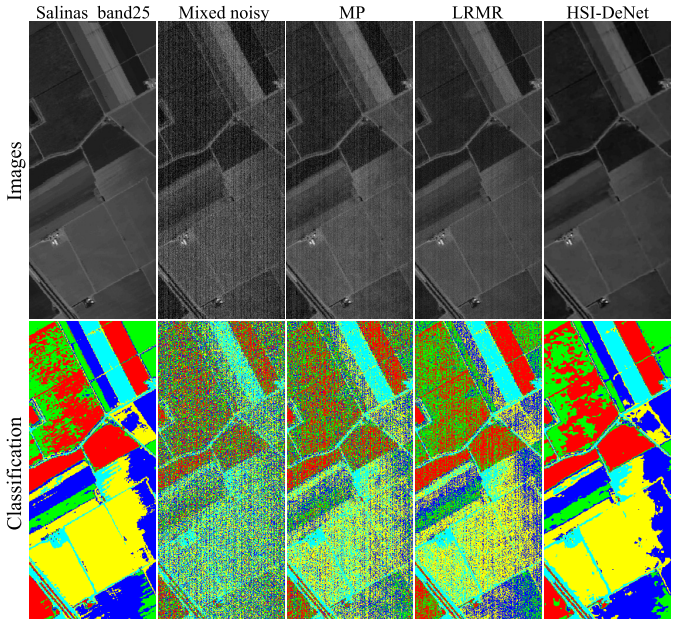


Fig. 14. Classification results by the k-mean classification method. (First row) Images. (Second row) Corresponding classification results.

and the recovery results by multihypothesis prediction [73], LRMR [40], and the proposed method. The bottom row is the corresponding classification results. In Fig. 14(c) and (d), obvious classification errors can be seen in the original classification result. On the contrary, the restored classification image in Fig. 14(e) does not contain any random or line artifacts. It can be seen that the classification results are significantly improved after the destriping and denoising process. This demonstrates that the restoration process is successfully applied.

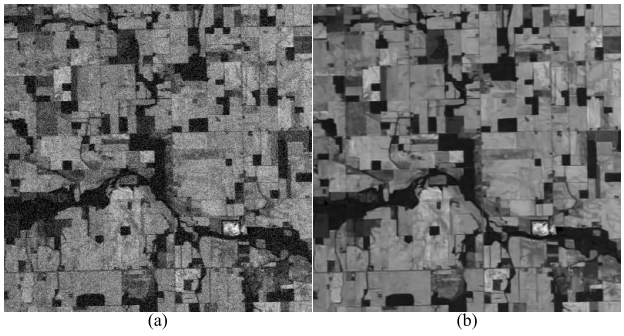


Fig. 15. Real AVIRIS image denoising results. Although our model is trained on the ICVL data set, it works well for numerous simulated and real HSI data sets with different spectral bands ranging. These experiments could demonstrate the generalization ability of our model to some extent.

3) Generalization Analysis: Note that our model is trained on the ICVL data set. We have tested our trained model for various data sets, including the ICVL, CAVE, and *urban*. The trained model worked well for the test images, such as the *urban* data, with different spectral bands ranging from 1690 to 1780 nm. And our method could also be well applied to other real HSI data, such as the Indian Pine shown in Fig. 15. These experiments demonstrate the generalization ability of our model for various HSIs. In fact, our model is sensitive to the noise level and category. We will discuss the limitations in Section IV-D.4.

Although the trained model on ICVL is robust to various HSIs, we perform an experiment to illustrate that the fine-tuning strategy could further boost the final restoration performance. We trained our model on ICVL and then fine-tuned it on the CAVE (32 scenes) where 90% is used as training and 10% as test. We show the quantitative results *Toy* under noise level $\sigma = 20$ and $S = 20$ between without and with fine-tuning in Table IV. We can observe that fine-tune

TABLE IV
EFFECTIVENESS OF FINE-TUNING STRATEGY

Model	PSNR	SSIM	SAM	ERGAS
Without Fine-tune	**37.79**	0.9467	0.2212	65.05
Fine-tune	37.54	**0.9496**	**0.2084**	**53.53**

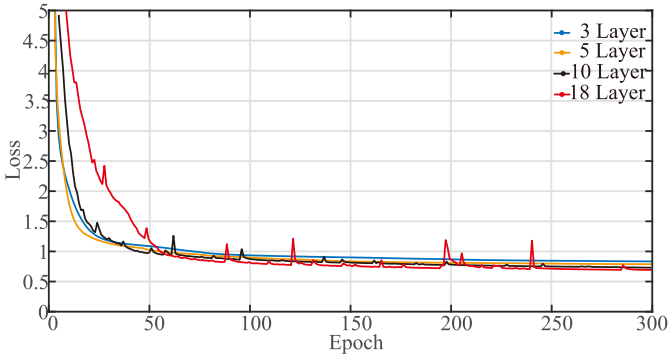


Fig. 16. Relationship between the training loss and depth of our network.

TABLE V
QUANTITATIVE ANALYSIS OF THE RELATIONSHIP BETWEEN THE TRAINING LOSS AND DEPTH OF OUR NETWORK

Model Depth	PSNR	SSIM	SAM	ERGAS
3 Layer	40.80	0.9787	0.0237	26.72
5 Layer	41.05	0.9800	0.0221	25.79
10 Layer	41.30	0.9820	0.0211	25.12
18 Layer	**41.58**	**0.9831**	**0.0209**	**24.24**

obviously improves the restoration performance in terms of the spectral information. That is to say the fine-tuned model is more adaptive to the specific imaging system. However, it is a little surprising that the PSNR value is even slightly inferior to that of the without fine-tuning, and the SSIM has been slightly improved. We speculate that this is due to the lack of training samples in CAVE. Moreover, a large part of the images in CAVE is all dark with zero values. Thus, the fine-tuned model may be underfitting due to the insufficient spatial information. The experiment can demonstrate that the fine-tuning strategy is a very effective way for the HSIs data set with slightly different distribution.

4) Depth of the Network: We also analyze the influence of the number of the convolutional layers. In Fig. 16, we show the training loss of the model with different depths. Here, we just choose the model with 3, 5, 10, and 18 layers as a representation. We have two observations here. First, at the first 50 epochs, the training loss of the model with fewer layers is usually lower. We guess that the shallow models are easier to be trained. Second, with sufficient training, the deeper the model is, the lower the training loss is. Moreover, we give the quantitative assessment in Table V. That is to say the depth of the model does facilitate the training of the HSIs, and improve the restoration performance.

5) Running Time: We test the running times of various methods for different HSI sizes, including both the CPU and GPU versions of the HSI-DeNet. The experiments are performed with the personal computer with MATLAB 2014a, one

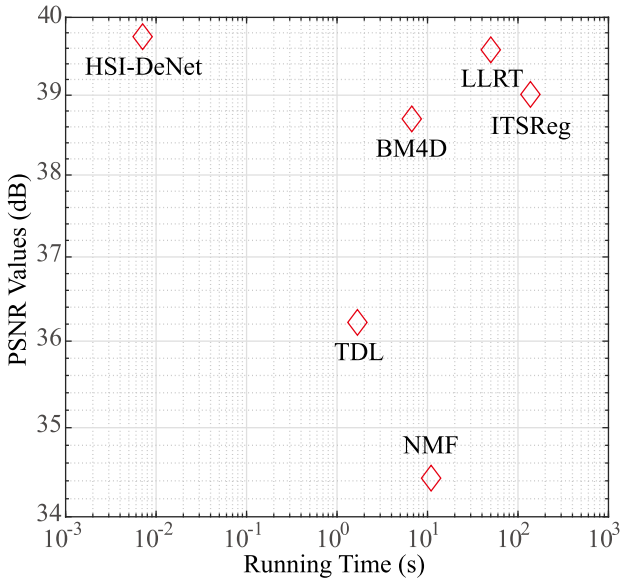


Fig. 17. Comparison with the state-of-the-art HSI-DeNets in terms of both speed and performance for image with size 180×180×10.

Titan X GPU, an Intel i7 CPU at 3.6 GHz, and 32-GB memory. As shown in Table VI, even the CPU version of the HSI-DeNet is much faster than that of the other methods. Moreover, the GPU version of HSI-DeNet is almost 100 times faster than the CPU version. More importantly, with the increase in the HSIs' size, the running time of the GPU version of HSI-DeNet does not increase too much, while the running times of the others are more and more unacceptable. This property makes our method extremely suitable for HSIs' processing, since the data size of HSIs is always very large. We jointly compare the running time and restoration performance of all methods, as shown in Fig. 17. Our method achieves the best performance in both the speed and the PSNR value.

6) Regularization Parameter: Equation (7) contains two terms: the reconstruction term and the adversarial term. Looking back at the framework flowchart in Fig. 4, the physical meaning of the first term is to guarantee the intensity similarity between the restored image and the ground truth. The physical meaning of the second term is to work as a learnable discriminative prior, which desires the distribution of restored data to be adaptively similar to that of the ground truth. Keeping this in mind, it is a guidance for us to set the parameters in (7) and control the importance of each term. Normally, we set $\lambda_1 = 1$ and adjust the regularization parameter λ_2.

Here, we explore how the regularization parameter λ_2 influence the reconstruction performance. We show the training loss of the reconstruction term by changing λ_2 in [0.001, 0.01, 0.1, 1]. In Fig. 18, with the increase in λ_2, the training loss of the reconstruction term gradually increases. This is reasonable. With the increase in λ_2, the model will pay less attention to the reconstruction error. That is to say, the reconstruction error becomes larger, and the PSNR value gradually decreases with the increase in λ_2. For the extreme case, when $\lambda_2 = 0$,

TABLE VI
RUNNING TIME (SECONDS) COMPARISON UNDER DIFFERENT IMAGE SIZES

Image Size	Methods								
	BM3D	PARAFAC	LRMR	BM4D	TDL	ITSReg	LLRT	HSI-DeNet(CPU)	HSI-DeNet(GPU)
180*180*10	1.863	3.191	4.422	6.939	4.037	106.256	51.527	0.624	**0.013**
380*380*10	14.233	3.962	65.294	30.28	8.938	510.181	300.254	2.794	**0.023**
512*512*10	24.821	6.667	113.234	59.095	18.221	939.81	369.392	5.870	**0.026**

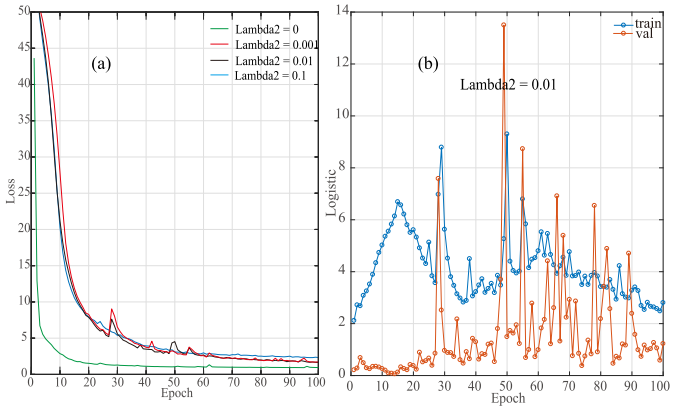


Fig. 18. Analysis of the regularization parameter λ_2. (a) Loss curve of the generator. (b) Loss curve of the discriminator.

(7) degenerates to the conventional reconstruction loss as in (5). However, as the goal of the second term controlled by λ_2 is not PSNR value-oriented, it is for the perceptual appearance, as shown in Fig. 13. As a tradeoff, we empirically set $\lambda_2 = 0.01$.

7) Training Convergence: We show the training loss of both the generator and the discriminator to judge the training convergence. For the generator, we can observe from Figs. 16 and 18(a), the training loss decreases and converges to a very low value at different depths and different parameters. For the discriminator, we used the discriminator similar to that of the DCGAN [74]. However, the loss of the DCGAN is meaningless [74]. We plot the adversarial loss under different values of λ_2 in Fig. 18(b). As we expected, the loss curve does not exhibit any regularity. As far as we know, only the WGAN [75] could provide the reasonable curve/guidance for the discriminator convergence judgment. We would like to try more sophisticated GAN, such as WGAN as our future work.

8) Limitation: Right now, the trained HSI-DeNet is only suitable for the specific noise level. For example, as shown in Fig. 19, if we want to restore the degraded HSI with noise level $\sigma = 50$, we have to use the trained HSI-DeNet with the HSI data degraded with $\sigma = 50$. The denoising result in Fig. 19(b) by HSI-DeNet_50 is with satisfactory appearance. However, if we import the noise image with either $\sigma = 25$ or $\sigma = 75$ to the $\sigma = 50$ based trained HSI-DeNet, we could observe that the results [Fig. 19(c) and (d)] are either oversmoothed or with residual noise. This heavily restricts the application of our HSI-DeNet for real HSIs with the unknown noise level. Fortunately, the recent studies are working toward

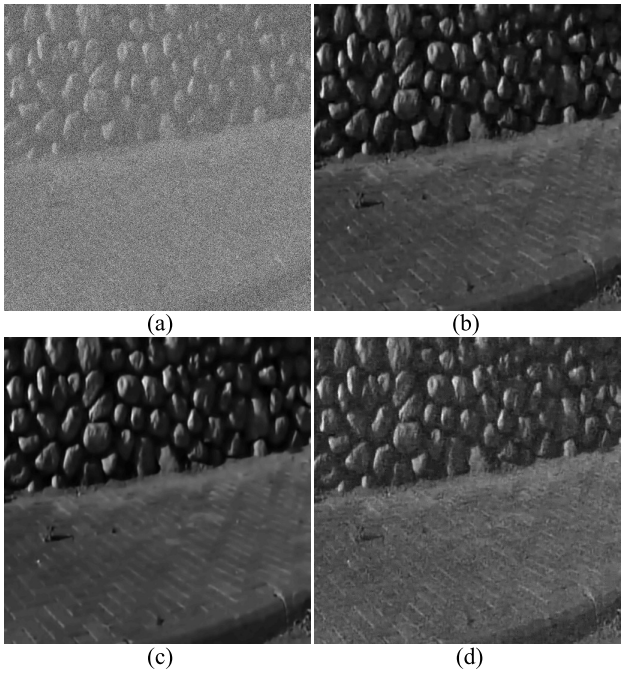


Fig. 19. Limitation of our HSI-DeNet. (a) Degraded image in CVIL with random noise $\sigma = 50$. The denoising result by the HSI-DeNet trained with HSI data degraded with (b) $\sigma = 50$, (c) $\sigma = 25$, and (d) $\sigma = 75$.

this direction by training one single network for the general restoration task [76], [77]. We will take this point as our future work.

The training of our HSI-DeNet is better to be performed on the GPU platform. Although our model could also be trained on the CPU, it requires several days to be completed. Larger training data sets are needed to improve the generalization of our model to accommodate all kinds of complex noise category and HSI data. Right now, we do not train one single model for all situations. Instead, we train each model for one specific task. The trained model obtains its best performance for the specific noise category.

V. CONCLUSION

In this paper, we introduce the deep CNN to remove the mixed noise in the HSIs. The proposed method significantly advances the HSIs' restoration field in three aspects: complex noise modeling, spectral-spatial structure preserving, and running time. The implicit yet powerful representative ability of the CNN enables us to better model the mixed noise in the HSIs. The learned 2-D filters with multiple channels

inherently match the multidimensional property of HSIs, better preserving the spatial-spectral structure correlation. Last but not least, the simple operations in the HSI-DeNet make the algorithm extremely fast for testing. Our method has been tested on various simulated and real HSIs, and achieved better restoration performance than the compared methods in terms of both quantitative and qualitative assessments. In the future, we would like to introduce the real 3-D CNN [78] into the HSIs' restoration, and also extend the CNN method to other interesting HSI tasks, such as super-resolution, unmixing, and so on.

REFERENCES

[1] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. CVPR*, Jul. 2017, pp. 5344–5353.

[2] Y. Chang, L. Yan, H. Fang, and H. Liu, "Simultaneous destriping and denoising for remote sensing images with unidirectional total variation and sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 6, pp. 1051–1055, Jun. 2014.

[3] Y. Chang, L. Yan, H. Fang, and C. Luo, "Anisotropic spectral-spatial total variation model for multispectral remote sensing image destriping," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1852–1866, Jun. 2015.

[4] 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, Sep. 2016.

[5] A. Garzelli and F. Nencini, "Hypercomplex quality assessment of multi/hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 4, pp. 662–665, Oct. 2009.

[6] L. Zhang, W. Wei, Y. Zhang, C. Shen, A. van den Hengel, and Q. Shi, "Cluster sparsity field for hyperspectral imagery denoising," in *Proc. ECCV*, 2016, pp. 631–647.

[7] Y. Fu, A. Lam, I. Sato, and Y. Sato, "Adaptive spatial-spectral dictionary learning for hyperspectral image restoration," *Int. J. Comput. Vis*, vol. 122, no. 2, pp. 228–245, 2017.

[8] H. Shen and L. Zhang, "A MAP-based algorithm for destriping and inpainting of remotely sensed images," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 5, pp. 1492–1502, May 2009.

[9] X. Cao *et al.*, "Low-rank matrix factorization under general mixture noise distributions," in *Proc. IEEE Conf. ICCV*, Dec. 2015, pp. 1493–1501.

[10] D. Meng and F. De La Torre, "Robust matrix factorization with unknown noise," in *Proc. IEEE Conf. ICCV*, Dec. 2013, pp. 1337–1344.

[11] Y. Chang, L. Yan, and S. Zhong, "Hyper-laplacian regularized unidirectional low-rank tensor recovery for multispectral image denoising," in *Proc. IEEE Conf. CVPR*, Jul. 2017, pp. 4260–4268.

[12] W. Dong, G. Li, G. Shi, X. Li, and Y. Ma, "Low-rank tensor approximation with laplacian scale mixture modeling for multiframe image denoising," in *Proc. IEEE Conf. ICCV*, Dec. 2015, pp. 442–449.

[13] 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, Aug. 2017.

[14] 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. CVPR*, Jun. 2014, pp. 2949–2956.

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

[16] Y. Chang, L. Yan, H. Fang, S. Zhong, and Z. Zhang. (2017). "Weighted low-rank tensor recovery for hyperspectral image restoration." [Online]. Available: https://arxiv.org/abs/1709.00192

[17] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. NIPS*, 2009, pp. 2080–2088.

[18] 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, Jan. 2013.

[19] 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, Jul. 2017.

[20] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.

[21] X. Fu, J. Huang, X. Ding, Y. Liao, and J. Paisley, "Clearing the skies: A deep network architecture for single-image rain removal," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2944–2956, Jun. 2017.

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

[23] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten, "Densely connected convolutional networks," in *Proc. IEEE Conf. CVPR*, Jul. 2017, pp. 4700–4708.

[24] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. ICLR*, 2016.

[25] J.-L. Starck, E. J. Candes, and D. L. Donoho, "The curvelet transform for image denoising," *IEEE Trans. Electron. Packag. Manuf.*, vol. 11, no. 6, pp. 670–684, Jun. 2002.

[26] E. J. Candes and D. L. Donoho, "Ridgelets: Theory and applications," Ph.D dissertation, Stanford Univ., Stanford, CA, USA, 1998.

[27] S. Mallat, *A Wavelet Tour of Signal Processing*. New York, NY, USA: Academic, 1998.

[28] H. Othman and S.-E. Qian, "Noise reduction of hyperspectral imagery using hybrid spatial-spectral derivative-domain wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 2, pp. 397–408, Feb. 2006.

[29] D. Letexier and S. Bourennane, "Noise removal from hyperspectral images by multidimensional filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 7, pp. 2061–2069, Jul. 2008.

[30] H. Peng and R. Rao, "Hyperspectral image enhancement with vector bilateral filtering," in *Proc. IEEE Conf. ICIP*, Nov. 2009, pp. 3713–3716.

[31] G. Chen and S.-E. Qian, "Denoising of hyperspectral imagery using principal component analysis and wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 973–980, Mar. 2011.

[32] L. Gómez-Chova, L. Alonso, L. Guanter, G. Camps-Valls, J. Calpe, and J. Moreno, "Correction of systematic spatial noise in push-broom hyperspectral sensors: Application to chris/proba images," *Appl. Opt.*, vol. 47, no. 28, pp. F46–F60, 2008.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. NIPS*, 2012, pp. 1097–1105.

[34] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.

[35] 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, Dec. 2006.

[36] Q. Yuan, L. Zhang, and H. Shen, "Hyperspectral image denoising employing a spectral–spatial adaptive total variation model," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3660–3677, Oct. 2012.

[37] Y.-Q. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.

[38] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. NIPS*, 2011, pp. 612–620.

[39] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.

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

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

[42] Y. Chang, L. Yan, T. Wu, and S. Zhong, "Remote sensing image stripe noise removal: From image decomposition perspective," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7018–7031, Dec. 2016.

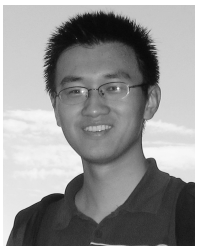
[43] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.

[44] 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, Oct. 2012.

[45] X. Guo, X. Huang, L. Zhang, and L. Zhang, "Hyperspectral image noise reduction based on rank-1 tensor decomposition," *ISPRS J. Photogramm. Remote Sens.*, vol. 83, no. 9, pp. 50–63, Sep. 2013.

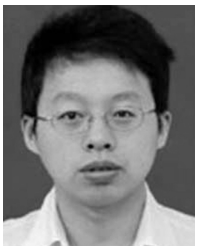
[46] A. Romero, C. Gatta, and G. Camps-Valls, "Unsupervised deep feature extraction for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1349–1362, Mar. 2016.

[47] Y. Wei, Q. Yuan, H. Shen, and L. Zhang, "Boosting the accuracy of multispectral image pansharpening by learning a deep residual network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1795–1799, Oct. 2017.
[48] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published, doi: 10.1109/TNNLS.2018.2798162.
[49] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[50] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. CVPR*, Jun. 2016, pp. 1646–1654.
[51] J.-F. Cai, H. Ji, C. Liu, and Z. Shen, "Framelet-based blind motion deblurring from a single image," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 562–572, Feb. 2012.
[52] T. Liu, H. Liu, Z. Chen, and A. M. Lesgold, "Fast blind instrument function estimation method for industrial infrared spectrometers," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2018.2794449.
[53] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. ICLR*, 2015.
[54] Y. Yang, J. Sun, H. Li, and Z. Xu, "Deep ADMM-net for compressive sensing MRI," in *Proc. NIPS*, 2016, pp. 10–18.
[55] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3D?" in *Proc. IEEE Conf. CVPR*, Jun. 2012, pp. 2392–2399.
[56] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. ECCV*, 2014, pp. 184–199.
[57] F. Agostinelli, M. R. Anderson, and H. Lee, "Adaptive multi-column deep neural networks with application to robust image denoising," in *Proc. NIPS*, 2014, pp. 1493–1501.
[58] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[59] K. Simonyan and A. Zisserman, "Adam: A method for stochastic optimization," in *Proc. ICLR*, 2015.
[60] B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural RGB images," in *Proc. ECCV*, 2016, pp. 19–34.
[61] A. Vedaldi and K. Lenc, "MatConvNet: Convolutional neural networks for MATLAB," in *Proc. ACM Int. Conf. Multimedia*, 2014, pp. 689–692.
[62] M. Bouali and S. Ladjal, "Toward optimal destriping of MODIS data using a unidirectional variational model," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 8, pp. 2924–2935, Aug. 2011.
[63] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. CVPR*, Jul. 2016, pp. 105–114.
[64] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. ECCV*, 2016, pp. 694–711.
[65] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 2672–2680.
[66] 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, Aug. 2007.
[67] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[68] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Berkeley, CA, USA: Presses des MINES, 2002.
[69] R. H. Yuhas, J. W. Boardman, and A. F. Goetz, "Determination of semi-arid landscape endmembers and seasonal trends using convex geometry spectral unmixing techniques," in *Proc. Summaries 4th Annu. JPL Airborne Geosci. Workshop*, 1993, pp. 620–636.
[70] R. Pande-Chhetri and A. Abd-Elrahman, "De-striping hyperspectral imagery using wavelet transform and adaptive frequency domain filtering," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 5, pp. 620–636, 2011.
[71] H. Carfantan and J. Idier, "Statistical linear destriping of satellite-based pushbroom-type images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 4, pp. 1860–1871, Apr. 2010.
[72] Y. Chang, L. Yan, and S. Zhong, "Transformed low-rank model for line pattern noise removal," in *Proc. IEEE Conf. ICCV*, Oct. 2017, pp. 1726–1734.
[73] C. Chen, W. Li, E. W. Tramel, M. Cui, S. Prasad, and J. E. Fowler, "Spectral–spatial preprocessing using multihypothesis prediction for noise-robust hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1047–1059, Apr. 2014.
[74] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. ICLR*, 2015, pp. 214–223.
[75] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018.
[76] K. Zhang, M. Zuo, and L. Zhang. (2017). "FFDNet: Toward a fast and flexible solution for CNN based image denoising." [Online]. Available: https://arxiv.org/abs/1710.04026
[77] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. ICCV*, 2017, pp. 624–632.
[78] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.



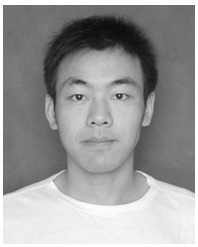
Yi Chang (S'15) received the B.S. degree in automation from the University of Electronic Science and Technology of China, Chengdu, China, in 2011, and the M.S. degree in pattern recognition and intelligent systems from the Huazhong University of Science and Technology, Wuhan, China, in 2014, where he is currently pursuing the Ph.D. degree with the School of Automation.

From 2014 to 2015, he was a Research Assistant with Peking University, Beijing, China. He was a Research Intern with the Tencent Youtu Lab, Shenzhen, China. His research interests include deep learning, image processing, and computer vision.



Luxin Yan (M'12) received the B.S. degree in electronic communication engineering and the Ph.D. degree in pattern recognition and intelligence system from Huazhong University of Science and Technology (HUST), Wuhan, China, in 2001 and 2007, respectively.

He is currently a Professor with the School of Automation, HUST. His research interests include multispectral image processing, pattern recognition, and real-time embedded systems.

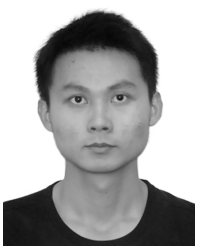


Houzhang Fang received the M.S. and Ph.D. degrees in applied mathematics and control science from the Huazhong University of Science and Technology, Wuhan, China, in 2010 and 2014, respectively.

He is currently a Lecturer with Xidian University, Xi'an, China. His research interests include image restoration, blind image deconvolution, object detection, and recognition.

Sheng Zhong received the Ph.D. degree in pattern recognition and intelligent system from the Huazhong University of Science and Technology, Wuhan, China, in 2005.

He is currently a Professor with the Huazhong University of Science and Technology. His research interests include pattern recognition, image processing, and real-time embedded systems.



Wenshan Liao received the B.S. degree from the School of Automation, Huazhong University of Science and Technology, Wuhan, China, in 2015, where he is currently pursuing the M.S. degree.

His research interests include image processing and machine learning.