

Learning Spatial-Spectral Prior for Super-Resolution of Hyperspectral Imagery

Junjun Jiang , Member, IEEE, He Sun, Xianming Liu , Member, IEEE, and Jiayi Ma , Member, IEEE

Abstract—Recently, single gray/RGB image super-resolution reconstruction task has been extensively studied and made significant progress by leveraging the advanced machine learning techniques based on deep convolutional neural networks (DCNNs). However, there has been limited technical development focusing on single hyperspectral image super-resolution due to the high-dimensional and complex spectral patterns in hyperspectral image. In this article, we make a step forward by investigating how to adapt state-of-the-art deep learning based single gray/RGB image super-resolution approaches for computationally efficient single hyperspectral image super-resolution, referred as SSPSR. Specifically, we introduce a spatial-spectral prior network (SSPN) to fully exploit the spatial information and the correlation between the spectra of the hyperspectral data. Considering that the hyperspectral training samples are scarce and the spectral dimension of hyperspectral image data is very high, it is nontrivial to train a stable and effective deep network. Therefore, a group convolution (with shared network parameters) and progressive upsampling framework is proposed. This will not only alleviate the difficulty in feature extraction due to high dimension of the hyperspectral data, but also make the training process more stable. To exploit the spatial and spectral prior, we design a spatial-spectral block (SSB), which consists of a spatial residual module and a spectral attention residual module. Experimental results on some hyperspectral images demonstrate that the proposed SSPSR method enhances the details of the recovered high-resolution hyperspectral images, and outperforms state-of-the-arts. The source code is available at [Online]. Available: <https://github.com/junjun-jiang/SSPSR>.

Index Terms—Hyperspectral remote sensing, image super-resolution, deep convolutional neural networks (DCNNs), spatial-spectral prior.

I. INTRODUCTION

UNLIKE human eyes, which can only be exposed to visible light, hyperspectral imaging is an imaging technique for collection and processing information across the entire range of electromagnetic spectrum [1]. The most important feature of hyperspectral imaging is the combination of imaging technology

Manuscript received November 26, 2019; revised April 1, 2020 and May 6, 2020; accepted May 6, 2020. Date of publication May 20, 2020; date of current version July 6, 2020. The work was supported in part by the National Natural Science Foundation of China under Grants 61971165, 61922027 and 61773295 and in part by the Fundamental Research Funds for the Central Universities. This article was recommended by Associate Editor Prof. Ulugbek S. Kamilov. (*Corresponding authors:* Junjun Jiang and Xianming Liu.)

Junjun Jiang and Xianming Liu are with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China, and also with the Peng Cheng Laboray, Shenzhen 518000, China (e-mail: jiangjunjun@hit.edu.cn; csxm@hit.edu.cn).

He Sun is with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China (e-mail: 19s103179@hit.edu.cn).

Jiayi Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jyyma2010@gmail.com).

Digital Object Identifier 10.1109/TCI.2020.2996075

and spectral detection technology. While imaging the spatial features of the target, each spatial pixel in a hyperspectral image is dispersed to form dozens or even hundreds of narrow spectral bands for continuous spectral coverage. Therefore, hyperspectral images have a strong spectral diagnostic capability to distinguish materials that look similar for humans.

However, the hyperspectral imaging system is often compromised due to the limitations of the amount of the incident energy. There is always a tradeoff between the spatial and spectral resolution of the real imaging process. With the increase of spectral features, if all other factors are kept constant to ensure a high signal-to-noise ratio (SNR), the spatial resolution will inevitably become a victim. Therefore, how to obtain a reliable hyperspectral image with high-resolution still remains a very challenging problem.

Super-resolution reconstruction can infer a high-resolution image from one or sequential observed low-resolution images [2]. It is a post-processing technique that does not require hardware modifications, and thus could break through the limitations of the imaging system. According to whether the auxiliary information (such as panchromatic, RGB, or multispectral image) is utilized, hyperspectral image super-resolution techniques can be divided into two categories: *fusion based hyperspectral image super-resolution* (sometimes called hyperspectral image pansharpening) and *single hyperspectral image super-resolution* [3]. The former merges the observed low-resolution hyperspectral image with the higher spatial resolution auxiliary image to improve the spatial resolution of the observed hyperspectral image. These fusion approaches based on Bayesian inference, matrix factorization, sparse representation, or recently advanced deep learning techniques have flourished in recent years and achieved considerable performance [4]–[6]. However, most of these methods all assume that the input low-resolution hyperspectral image and the high-resolution auxiliary image are well co-registered. In practical applications, obtaining such well co-registered auxiliary images would be difficult, if not impossible [7]–[9].

Compared with fusion based hyperspectral image super-resolution, single hyperspectral image super-resolution has received less attention and there has been limited advancement due to the spectral patterns in hyperspectral images and no additional auxiliary information. To exploit the abundant spectral correlations among successive spectral bands, several single hyperspectral image super-resolution approaches based on sparse and dictionary learning or low-rank approximation have been developed [10]–[13]. However, these hand-crafted priors can

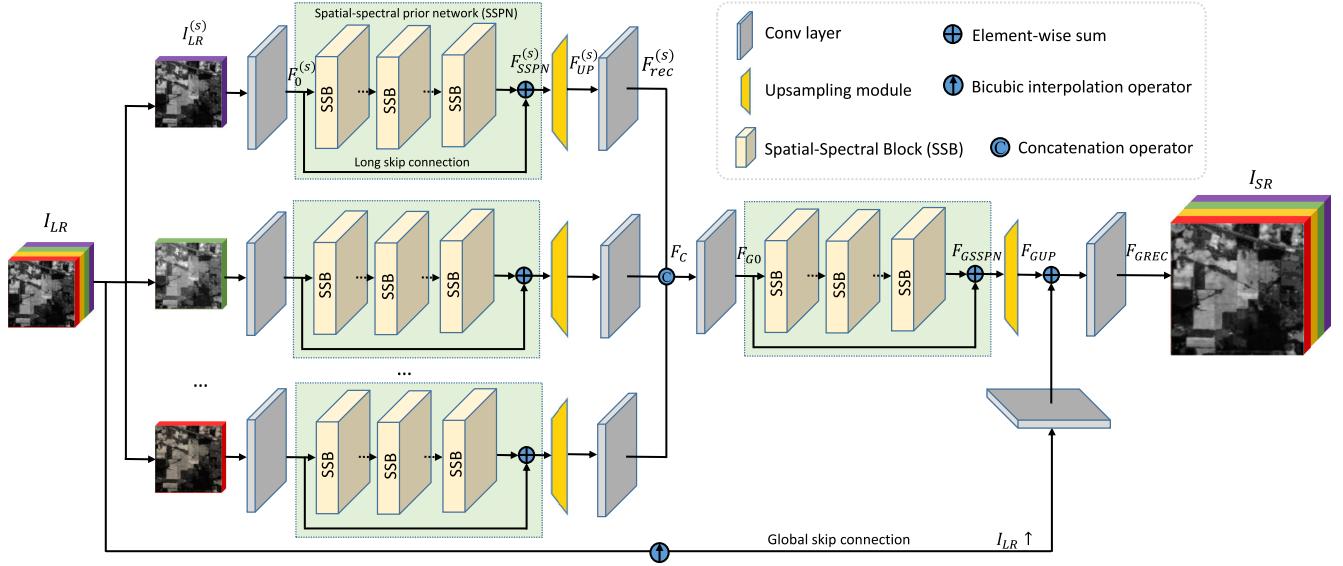


Fig. 1. The overall network architecture of the proposed SSPSR network.

only reflect the characteristics of one aspect of the hyperspectral data.

Recently, deep convolutional neural network (DCNN) has shown extraordinary capability of modelling the relationship between the low-resolution images and high-resolution ones, i.e., single gray/RGB image super-resolution task [14]–[16]. The practiced rationale in these schemes can be summarized as follows: given a very large number of example pairs of original images and their corrupted versions, a deep network can be learned to restore the degraded image to its source.

Specifically, compared with the single gray/RGB image super-resolution based on deep learning, in the single hyperspectral image super-resolution task, it is nontrivial to train a computationally efficient and effective deep network. This is mainly due to the following reasons: on the one hand, hyperspectral images are not as popular as natural images, the training sample number of available hyperspectral image dataset is extremely small. Even if we can collect a lot of images, hyperspectral images may be obtained by different hyperspectral cameras. The differences in the number of spectral bands and imaging conditions will make it more difficult to establish a unified deep network. On the other hand, the spectral dimensionality of hyperspectral image data itself is very high. Unlike traditional gray/RGB images, hyperspectral images often have hundreds of contiguous spectral bands, which calls for larger dataset to guarantee the training process. Otherwise, it is easy to cause the over-fitting problem.

In order to deal with the above problems caused by the lack of data and the inability to fully exploit the spatial information and spatial correlation characteristics in hyperspectral data, a *group convolution* (with shared network parameters) and *progressive upsampling* framework is proposed in this paper, which can greatly reduce the size of the model and make it feasible to obtain stable training results under small data conditions. For exploiting the spatial and spectral correlation characteristics of hyperspectral data, we carefully design the spatial-spectral

prior network (SSPN), which cascades multiple spatial-spectral blocks (SSBs). For each SSB, it contains a spatial residual module and a spectral attention residual module. The former consists of a standard residual block which is used to exploit spatial information of the hyperspectral data, while the latter consists of a spectral attention residual module which is used to extract spectral correlations. Through short and long skip connections, a residual in residual architecture is formed, which makes the spatial-spectral feature extraction more efficient.

Fig. 1 shows the network architecture of our spatial-spectral prior network based super-resolution network (SSPSR). The input low-resolution hyperspectral image is firstly divided into several overlap groups. For each group, a branch network is applied to extract the spatial-spectral features of the input grouped hyperspectral images (a subset of the entire hyperspectral linages) and upscale them with a smaller unsampling factor (compared with the final target). And then, the output features of all branches are concatenated and fed to the following global spatial-spectral feature extraction and upsampling networks. Note that in order to let the SSPN in branch network and global network share the same structure, we insert a “reconstruction” layer after each branch upsampling module. Similar to many previous super-resolution networks, we also adapt a global residual structure to facilitate the prediction of the target. Therefore, in the proposed SSPSR network, the transmission of information flow is very flexible by designing these short (refer to residual spatial/spectral blocks), long (refer to the spatial-spectral prior network), global skip links. During the training phase, we share the network parameters of each branch across all groups, which avoids heavy computational cost and simplifies the complex optimization process. Comprehensive ablation studies demonstrate the effectiveness of each component and the fusion strategy used in the proposed method. Comparison results with state-of-the-art single hyperspectral image super-resolution methods on two public datasets demonstrate the effectiveness of the proposed SSPSR network.

We summarize the main contributions of this paper as follows. Considering the limited hyperspectral training samples and the high dimensionality of spectral bands, it is difficult to learn the mapping relationship from low-resolution space to high-resolution space in one-step upsampling. Inspired by the idea of some general image super-resolution methods, which conduct super-resolution progressively, we apply the progressive upsampling scheme to the single hyperspectral image super-resolution task and verify its effectiveness. In addition, we propose a spectral grouping and parameter sharing strategy to greatly reduce the parameters of the model and alleviate the difficulty in feature extraction. Inspired by the efficient residual learning and attention mechanism, we develop a spatial-spectral feature extraction network to fully exploit the spatial-spectral prior of hyperspectral images.

The rest of this paper is organized as follows: Section II presents the related work of hyperspectral image super-resolution. In Section III, we give the details of our SSPSR network architecture and the SSB. Then, the network configuration and experimental results including ablation analysis are reported in Section IV. Finally, some conclusions are drawn in Section V.

II. RELATED WORK

In this section, we briefly review some methods that are most relevant to our work, which include *fusion based hyperspectral image super-resolution*, *single hyperspectral image super-resolution*, and *single gray/RGB image super-resolution*. A list of hyperspectral image super-resolution resources collected by Jiang can be found at [17].

A. Fusion Based Hyperspectral Image Super-Resolution

Remote sensing image fusion is a very challenging problem with long history. Generally speaking, this problem can be classified to two categories, pansharpening and super-resolution. In order to improve the spatial resolution of the multispectral images, some previous works cast the fusion problem into a variational reconstruction task by blending a panchromatic image with higher resolution. This is often referred as pansharpening. A taxonomy of pansharpening based fusion methods can be found in the literature [18]–[21].

Recently, low-resolution hyperspectral image and high-resolution multispectral image fusion based spatial resolution improvement technique, which is often referred as hyperspectral image super-resolution, has received extensive attention. For example, Yokoya *et al.* [5] proposed a coupled nonnegative matrix factorization (CNMF) based approach to infer the high-resolution hyperspectral images with a pair of high-resolution multispectral image and low-resolution hyperspectral image. To exploit the redundancy and correlation in spectral domain, some approaches have been proposed by exploiting the sparsity [22], non-local similarity [23], [24], superpixel-guided self-similarity [25], clustering manifold structure [26], tensor and low-rank constraints [27], [28]. Most recently, some deep learning based methods have gradually become popular due to its superior performance and fewer assumptions regarding the image prior

[29]–[32]. Inspired by the iterative optimization based on the observation model, some deep unfolding network for fusion based hyperspectral image super-resolution methods are becoming popular in recent years [33]–[35]. The common idea of the above fusion based hyperspectral image super-resolution methods is to borrow high-frequency spatial information from high-resolution auxiliary image, and fuse these information to the target high-resolution hyperspectral image. Though these approaches have achieved very good performance, the major drawback of them is that a well co-registered auxiliary image with a higher resolution is needed. However, obtaining such a well co-registered auxiliary image would be arduous, if not impossible in practical applications [7]–[9].

B. Single Hyperspectral Image Super-Resolution

Without co-registered auxiliary image, single hyperspectral image super-resolution methods have still attracted considerable attention in reality. The pioneer work is proposed by Akgun *et al.* [36], in which a hyperspectral image acquisition model and the projection onto convex sets (POCS) algorithm [37] is applied to reconstruct the high-resolution hyperspectral image. By incorporating the low-rank and group-sparse constraints, Huang *et al.* [10] developed a novel method to tack with the unknown blurring problem. Recently, variants of sparse representations and dictionary learning based approaches are widely studied [12], [38]. However, these methods have some drawbacks. First, they usually need to solve some complex and time consuming optimization problems in the testing phase. Second, the image priors are often hand-crafted and based on the internal example without consideration of any external information from external samples. Due to the superior performance in many computer vision problems, deep learning techniques have also been introduced into the single hyperspectral image super-resolution task very recently. For example, Yuan *et al.* [39] and Xie *et al.* [40] firstly super-resolved the hyperspectral image based on the DCNNs, and then applied the nonnegative matrix factorization (NMF) to guarantee the spectral characteristic for the intermediate results. Essentially, they utilized DCNNs and matrix factorization to exploit the spatial and spectral features, separately, in a non-end-to-end manner. In [41], Mei *et al.* introduced a 3D full convolutional neural network to extract the feature of hyperspectral images. Although 3D convolution can well exploit the spectral correlation, the computational complexity is very large. Li *et al.* [42] proposed a grouped deep recursive residual network (GDRRN) by designing a group recursive module and embedding it into a global residual structure. This group-wise convolution and recursive structure can guarantee that it could yield very good performance. In our previous work [43], a feature pyramid block is designed to extract multi-scale features of the hyperspectral images. Most recently, inspired by the work of [44], which states that the image prior can be found within a CNN itself, Sidorov *et al.* [45] developed an effective single hyperspectral-image restoration algorithm. In general, these deep methods achieve better results than traditional methods. However, due to the limited hyperspectral training samples and the high dimensionality of spectral bands, it is difficult to fully

exploit the spatial information and the correlation among the spectra of the hyperspectral data.

C. Single Gray/RGB Image Super-Resolution

Recently, DCNN based approaches have achieved excellent performance over the single gray/RGB image super-resolution problem. The seminal work by Dong *et al.* [14] proposes a three layer convolutional neural network for the end-to-end image super-resolution(SRCNN) and achieved much better performance over conventional non-deep learning based methods. Benefiting from the residual learning, in VDSR [46] and DRCN [47] Kim *et al.* introduced very deep network for image super-resolution and achieved better results than the three layer SRCNN. The residual structure was then adopted in LapSRN [48], DRRN [49], and EDSR [15]. By simply attaching residual blocks, introducing the feedback, or incorporating non-local operations into a recurrent neural network, RDN [50], DBPN [51], and NLRN [52] are proposed. Inspired by the SE block [53], Zhang *et al.* developed a very deep network named RCAN by incorporating the channel attention module [16]. Most recently, Dai *et al.* introduced the non-local block and presented a second-order attention network (SAN) to capture the long-range dependencies [54]. Although fascinating results have been achieved, these methods are designed for the gray/RGB images, which have only one or three channels. When directly applying these approaches to the hyperspectral image, they will neglect the spectral correlations among spectra of the hyperspectral data, hindering the representation capacity of the network. In addition, for single gray/RGB image super-resolution, when using one- or three-channel pictures as network input, in order to extract features, a feature map of 64 (or more) channels is usually used. Similarly, if we also apply this 20-fold (or more) parameter growth network design scheme to hyperspectral images which have hundreds of channels, it will lead to a sharp increase in parameters. However, there is not enough hyperspectral data to support the model training like for the gray/RGB images.

III. THE PROPOSED SSPSR METHOD

A. Network Architecture

In Fig. 1, we show the network architecture of the proposed SSPSR method. It mainly consists of two parts: the branch networks and global network. For each branch network or the global network, it includes shallow feature extraction, spatial-spectral deep feature extraction, upsampling module, and reconstruction part. We denote $I_{LR} \in \mathbb{R}^{h \times w \times C}$ the input low-resolution hyperspectral image, $I_{SR} \in \mathbb{R}^{H \times W \times C}$ the corresponding output high-resolution hyperspectral image, and $I_{HR} \in \mathbb{R}^{H \times W \times C}$ the ground truth (original high-resolution hyperspectral image) of the input image I_{LR} . Our goal is to predict the high-resolution hyperspectral image I_{SR} from the input low-resolution hyperspectral image I_{LR} by the proposed end-to-end super-resolution reconstruction network,

$$I_{SR} = H_{Net}(I_{LR}), \quad (1)$$

where $H_{Net}(\cdot)$ denotes the function of the proposed SSPSR method.

Different from previous methods, which treat the hyperspectral images as multiple single channel images (reconstructing them separately) or as a whole, we divide the whole hyperspectral image into some groups. In this way, we can not only exploit the correlations among neighboring spectral bands of hyperspectral images, but also reduce the dimensionality of features of each group. Inspired by the success of the recently proposed residual network structure, which has achieved very good performance in the field of image restoration, we specifically design a SSB based on residual network structure. As shown in Fig. 1, the proposed SSPSR network contains several branch networks and a global network. For each branch network and the global network, they first extract the shallow features and feed them to the SSPN, then upscale the outputs of SSPN with an intermediate upsampling factor. By cascading the parallel branch networks with the global network, we can super-resolve the input low-resolution hyperspectral image in a coarse-to-fine manner. In the following, we will give details of the branch network and global network, respectively.

1) *The Branch Network:* Specifically, the input low-resolution hyperspectral image I_{LR} is firstly divided into S groups, $I_{LR} = \{I_{LR}^{(1)}, I_{LR}^{(2)}, \dots, I_{LR}^{(S)}\}$. It should be noted that, in our settings the neighboring groups may have overlaps. More details about the settings can be found at the experiment section. For each group $I_{LR}^{(s)}$, we directly apply one convolutional layer to obtain its shallow features $F_0^{(s)}$ as investigated in previous work [15], [16],

$$F_0^{(s)} = H_{FE}(I_{LR}^{(s)}), \quad (2)$$

where $H_{FE}(\cdot)$ denotes convolution operation, i.e., feature extraction layer. $F_0^{(s)}$ is then used for deep feature extraction with the proposed SSPN. Consequently, we can further have

$$F_{SSPN}^{(s)} = H_{SSPN}(F_0^{(s)}), \quad (3)$$

where $H_{SSPN}(\cdot)$ denotes the function of the proposed SSPN, which contains R SSBs and we will present its details in the following.

The output of SSPN can be treated as the deep features of one grouped hyperspectral images. In order to alleviate the burden of the final super-resolution reconstruction, we adopt a strategy of progressive super-resolution reconstruction. Particularly, we add an upsampling module in the middle of the network (before feeding the output of branch SSPN to the global SSPN), which has proven to be a very effective technique, especially when the magnification is very large. Thus, by the upsampling module we obtain the upscaled feature maps,

$$F_{UP}^{(s)} = H_{UP}(F_{SSPN}^{(s)}), \quad (4)$$

where $H_{UP}(\cdot)$ and $F_{UP}^{(s)}$ denote an upsampling module and upscaled features respectively. In this paper, we leverage the *PixelShuffle* [55] operator to conduct the upsampling procedure.

Before feeding the upscaled features to the following global SSPN, we add one *Conv* layer after each branch upsampling

module to reduce the number of feature channels to the spectral number of each input group. Therefore, the output of the branch network will have the same channels as the input grouped hyperspectral images, and we call this layer as a “reconstruction” layer,

$$F_{rec}^{(s)} = H_{rec}(F_{UP}^{(s)}), \quad (5)$$

where $H_{rec}(\cdot)$ denotes the “reconstruction” layer (Here we use a lowercase term “rec” to represent a pseudo-reconstruction operation). By this *Conv* layer, each branch can be seen as a super-resolution reconstruction subnetwork. Another purpose of designing this layer is to make the branch SSPN and global SSPN have the same network structure.

2) *The Global Network*: After extracting features from different groups with the branch networks, we concatenate them together from all branches (as shown in the “concatenation operator” of Fig. 1), i.e., $F_C = [F_{rec}^{(1)}, F_{rec}^{(2)}, \dots, F_{rec}^{(S)}]$. It should be noted that if the neighboring groups have overlaps, the integrated feature maps can be generated according to their original spectral band position and by averaging feature values in the overlapping bands. Similar to the local branch, before feeding the contacted features into the global SSPN, we apply one *Conv* layer to extract the “shallow features”,

$$F_{G0} = H_{GFE}(F_C), \quad (6)$$

where $H_{GFE}(\cdot)$ is similar to $H_{FE}(\cdot)$ and is used to extract corresponding “shallow features” of the input contacted features of all branch networks.

And then, we further feed F_{G0} into the global SSPN, whose structure is the same as the local one,

$$F_{GSSPN} = H_{GSSPN}(F_{G0}), \quad (7)$$

where $H_{GSSPN}(\cdot)$ refers to the global version of $H_{SSPN}(\cdot)$. In this way, we extract the spatial-spectral features F_{GSSPN} of the input hyperspectral images.

To upscale the obtained features to the target size, here we apply upsampling module once more (progressively reconstruction) to generate the upscaled spatial-spectral feature maps,

$$F_{GUP} = H_{GUP}(F_{GSSPN}), \quad (8)$$

where H_{GUP} refers to the global version of H_{UP} .

The final super-resolved hyperspectral images can be then obtained via one reconstruction layer by feeding the upscaled spatial-spectral features and the upscaled input hyperspectral images,

$$I_{SR} = H_{GREC}(F_{GUP} + H_{GFE2}(I_{LR} \uparrow)), \quad (9)$$

where $I_{LR} \uparrow$ refers to the Bicubic upsampling version of the input low-resolution hyperspectral images, $H_{GFE2}(\cdot)$ is similar to $H_{GFE}(\cdot)$ and is used to extract shallow features of the input Bicubic upscaled hyperspectral images for residual learning, and $H_{GREC}(\cdot)$ is the reconstruction operation that has one *Conv* layer. Here, “ $+H_{GFE}(I_{LR} \uparrow)$ ” is referred to as the residual learning.

B. Spatial-Spectral Prior Network (SSPN)

Image super-resolution is a very ill-posed problem, which calls for additional prior (regularization) to constrain the reconstruction procedure. Traditional approaches all try to design sophisticated regularization terms such as total variation (TV), sparse, low-rank, by hand [22], [23], [25]–[27]. Therefore, the performance of these algorithms is highly dependent on whether the designed prior can well characterize the observed data. As for the hyperspectral image super-resolution problem, it is crucial to effectively exploit the intrinsic properties of hyperspectral images, *i.e.*, the non-local self-similarity in spatial and the high correlation across spectra. Previous manually designed constraints are insufficient for accurate hyperspectral image restoration.

In this paper, we advocate a spatial-spectral feature extraction network (SSPN) to exploit the spatial and spectral prior. In particular, SSPN cascades R spatial-spectral blocks (SSBs) and can be formulated as,

$$F_{SSB_r}^{(s)} = H_{SSB_R}(H_{SSB_{R-1}}(\dots H_{SSB_1}(F_0^{(s)}) \dots)), \quad (10)$$

where $H_{SSB_r}(\cdot)$ refers to the function of the r -th SSB, and $F_{SSB_{r-1}}^{(s)}$ is the input of the r -th SSB and $F_{SSB_r}^{(s)}$ is the extracted features. Noted that we use the notations from the local branch network to demonstrate the detailed design of the local SSPN, and the global SSPN is the same to the local one.

To facilitate the prediction of the target, the long skip connection is further introduced in SSPN. This will lead to the direct passing of the low frequency features of the current features to the end, and let the current residual body pay more attention to the high frequency information. Therefore, the output of the SSPN can be obtained by

$$F_{SSPN}^{(s)} = H_{SSB_R}(H_{SSB_{R-1}}(\dots H_{SSB_1}(F_0^{(s)}) \dots)) + F_0^{(s)}. \quad (11)$$

Here, “ $+F_0^{(s)}$ ” is referred to as the residual learning (same as below). This residual in residual structure can enable fast as well as stable training.

In this paper, we specifically design the SSB to exploit the spatial-spectral information from the hyperspectral images. In particular, each SSB has two parts, *i.e.* a spatial residual module and a spectral attention residual module. The architecture of SSB is illustrated in Fig. 2. For the spatial residual module, we leverage the standard residual block with 3×3 convolutions to extract the spatial features,

$$F_{Spa_r}^{(s)} = F_{SSB_{r-1}}^{(s)} + H_{SSB_r-Spa}(F_{SSB_{r-1}}^{(s)}), \quad (12)$$

where $H_{SSB_r-Spa}(\cdot)$ refers to the function of the spatial residual module for the r -th SSB, and F_{Spa_r} is the spatial feature for the r -th SSB. The standard residual block can well extract the spatial information of a hyperspectral image.

However, due to the strong correlation between the spectra of a hyperspectral image, standard residual convolutional networks cannot effectively extract the spectral dependencies. The spectral correlation, which is characterized by that there exists strong correlation among neighboring spectral bands of hyperspectral

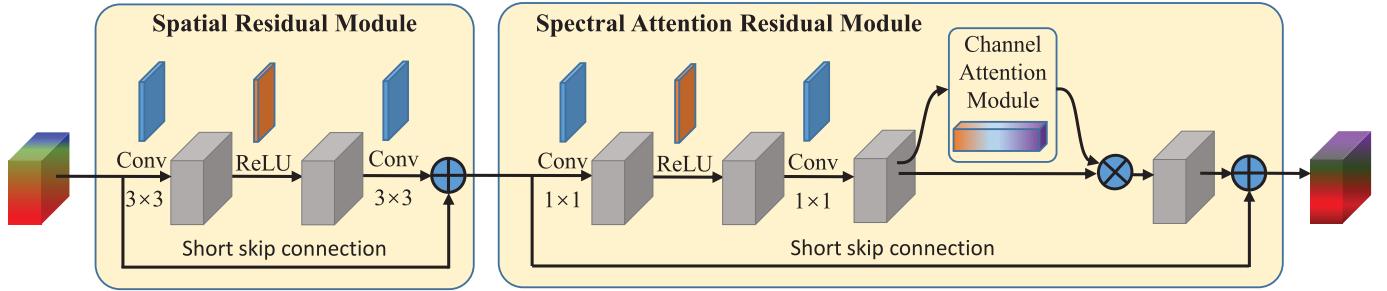


Fig. 2. The network architecture of the spatial-spectral block (SSB), which consists of a spatial residual module and a spectral attention residual module. “+” and “ \times ” denote element-wise addition and element-wise multiplication, respectively.

image, has been widely used for hyperspectral image reconstruction and analysis [5], [56]. To exploit this correlation, we can use all the spectral bands $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_C$ to obtain the newly reconstructed spectral band \mathbf{x}'_i , i.e., $\mathbf{x}'_i = w_{i,1}\mathbf{x}_1 + w_{i,2}\mathbf{x}_2 + \dots + w_{i,C}\mathbf{x}_C$. $\mathbf{w}_i = [w_{i,1}, w_{i,2}, \dots, w_{i,C}]$ are the linear combination (reconstruction) weights. If similar spectral bands share similar weights, the correlation information will be embedded in the reconstructed spectral band, thus exploiting the correlation among neighboring spectral bands of hyperspectral image. If we relax the weights to any learnable parameters, this will be equal to learning a set of weight vectors $\{\mathbf{w}_i\}_i$, and thus obtaining a new representation of the hyperspectral image. Mathematically, this can be achieved by some 1×1 filters (bottleneck layer), whose weights are $\{\mathbf{w}_i\}_i$. By designing a spectral network with 1×1 filters, we can expect to fully exploit the correlations between different spectral bands. It is worth noting that we further apply the ReLU layer to enhance its representation ability. Therefore, the structure of the SSB is designed as the combination of a spatial residual module and a spectral attention residual module as shown in Fig. 2. Thus, we have

$$F_{SSB_r}^{(s)} = F_{Spa_r}^{(s)} + H_{SSB_r-Spc}(F_{Spa_r}^{(s)}), \quad (13)$$

where $H_{SSB_r-Spc}(\cdot)$ denotes the spectral network of the r -th SSB.

To further improve the representation ability of spectral information as well as the entire network, we are inspired by Zhang *et al.* [16] and introduce the channel attention mechanism to adaptively rescale each channel-wise feature by modeling the interdependencies across feature spectra. Specifically, a global average pooling layer is applied to the extracted feature maps of previous spectral network to obtain a global context embedding vector. And then, two thin fully connected layers with a simple gating mechanism (by sigmoid function) is applied to learn nonlinear interactions between spectra. Then we obtain the final channel scaling coefficient vector $T \in \mathbb{R}^{1 \times 1 \times C}$, which is used to reweight the extracted feature maps. The output of the spectral attention residual module is simply computed by

$$F_{SSB_r}^{(s)} = F_{Spa_r}^{(s)} + T H_{SSB_r-Spc}(F_{Spa_r}^{(s)}). \quad (14)$$

C. Loss Function

In order to measure the super-resolution performance, several cost functions have been investigated to make the super-resolution results approximate to ground truth high-resolution images. In the current literature, l_2 , l_1 , perceptual, and adversarial losses are the most commonly used loss functions. When compared with perceptual and adversarial losses, which may restore details that do not exist in the original images and is undesirable in remote sensing field, l_2 and l_1 losses are more credible. As for l_2 loss, it encourages finding pixel-wise averages of plausible solutions which are typically overly-smooth. Due to that l_1 loss can effectively penalize small errors and maintain better convergence throughout the training phase, we adopt l_1 loss to measure the reconstruction accuracy of the network. Specifically, the l_1 loss is defined by mean absolute error (MAE) between all the reconstructed images and the ground truth:

$$\mathcal{L}_1(\Theta) = \frac{1}{N} \sum_{n=1}^N \|I_{HR}^n - H_{Net}(I_{LR}^n)\|_1, \quad (15)$$

where $H_{Net}(I_{LR}^n)$ and I_{HR}^n are the n -th reconstructed high-resolution hyperspectral image and ground truth hyperspectral image, respectively. N denotes the number of images in one training batch, and Θ refers the parameter set of our network.

However, above-mentioned loss is primarily designed for general image restoration tasks. Although they can well preserve the spatial information of the super-resolution results, the reconstructed spectral information may be distorted due to the ignorance of the correlations among spectral features. In order to simultaneously ensure the spatial and spectral credibility of the reconstruction results, we introduce the spatial-spectral total variation (SSTV) [57]. It extends the conventional total variation model and accounts for both the spatial and the spectral correlation. In this paper, we add the SSTV to the l_1 loss to impose spatial and spectral smoothness simultaneously,

$$\mathcal{L}_{SSTV}(\Theta) = \frac{1}{N} \sum_{n=1}^N (\|\nabla_h I_{SR}^n\|_1 + \|\nabla_w I_{SR}^n\|_1 + \|\nabla_c I_{SR}^n\|_1), \quad (16)$$

where ∇_h , ∇_w , and ∇_c are functions to compute the horizontal, vertical, and spectral gradient of I_{SR}^n .

TABLE I
AVERAGE QUANTITATIVE PERFORMANCE BY DIFFERENT LOSS FUNCTIONS
OVER FOUR TESTING IMAGES OF CHIKUSEI DATASET WITH RESPECT TO SIX
PQIS WHEN THE UPSAMPLING FACTOR IS 4

Losses	l_2	l_1	$l_1 + \text{SSTV}$
CC \uparrow	0.9535	0.9560	0.9565
SAM \downarrow	2.5152	2.3581	2.3527
RMSE \downarrow	0.0117	0.0115	0.0114
ERGAS \downarrow	5.1304	4.9903	4.9313
PSNR \uparrow	40.0703	40.3515	40.3612
SSIM \uparrow	0.9401	0.9437	0.9441

In summary, the final objective loss for the proposed model is a weighted sum of the two losses:

$$\mathcal{L}_{total}(\Theta) = \mathcal{L}_1 + \alpha \mathcal{L}_{SSTV}, \quad (17)$$

where α is used to balance the contributions of different losses. In our experiments, we set it as a constant, $\alpha = 1e - 3$.

In Table I, we report the reconstruction results (in terms of objective measurements) when using different losses (more details regarding the experimental settings can be found at the experiment section). Clearly, l_1 loss is much more suitable for our task, because it can effectively penalize small errors and maintain better convergence throughout the training phase. By introducing the SSTV constraint, slightly better results can be achieved.

D. Implementation Details

We use Pytorch libraries¹ to implement and train the proposed SSPSR network. We train different models to super-resolve the hyperspectral images for scale factors 4 and 8 with random initialization. We use the ADAM optimizer [58] with an initial learning rate of 1e-4 which decays by a factor of 10 when it reaches 30 epochs. In our experiments, we find it will take 40 epochs to achieve a stable performance. The models are trained with a batch size of 32. As in many previous work, we also apply the Bicubic interpolation to downsample the high-resolution hyperspectral images to obtain the corresponding low-resolution hyperspectral images.

Unless otherwise specified, in the following experiments we set the spectral band number (p) of each group to 8 and the overlap (o) between neighboring groups to 2. To efficiently process the “edge” spectral bands, we adopt a so called “fallback” dividing strategy. When the last group has less than p spectral bands, we select the last p bands as the last group. Therefore, the number of groups can be obtained by the following equation,

$$S = \text{ceil} \left(\frac{C - o}{p - o} \right), \quad (18)$$

where $\text{ceil}(\cdot)$ is the function that rounds the elements of to the nearest integers towards infinity. In the SSPN, the number of spatial-spectral blocks (R) is set to 3. We set the size of all Conv layers to 3×3 except for that in the spectral residual modules, where the kernel size is set to 1×1 . To ensure that the size of the feature map is not changed, the zero-padding strategy is applied for these Conv layers with kernel size 3×3 . The

¹[Online]. Available: <https://pytorch.org>

Conv layers in shallow feature extraction and SSPN have $C = 256$ filters, except for that in the channel-downscaling, i.e., the reconstruction network after the upscaled features at the branch networks (please refer to Eq. (5)).

IV. EXPERIMENTS AND RESULTS

In this section, we present a detailed analysis and evaluation of our approach on three public hyperspectral image datasets, which include two remote sensing hyperspectral image datasets, i.e., Chikusei dataset [59]² and Pavia Center dataset,³ and one nature hyperspectral image dataset, i.e., CAVE dataset [60].⁴ We compare the proposed method with eight comparison methods, including four state-of-the-art deep single gray/RGB image super-resolution methods, VDSR [46], EDSR [15], RCAN [16], and SAN [54], and four representative and most relevant deep single hyperspectral image super-resolution methods, TLCNN [39], 3DCNN [41], GDRRN [42], and DeepPrior [45]. We carefully adjust hyperparameters of these comparison methods to achieve their best performance. Bicubic interpolation is introduced as the baseline. The source code is available at <https://github.com/junjun-jiang/SSPSR>.

Evaluation measures: Six widely used quantitative picture quality indices (PQIs) are employed to evaluate the performance of our method, including cross correlation (CC) [61], spectral angle mapper (SAM) [62], root mean squared error (RMSE), erreur relative globale adimensionnelle de synthese (ERGAS) [63], peak signal-to-noise ratio (PSNR), and structure similarity (SSIM) [64]. For PSNR and SSIM of the reconstructed hyperspectral images, we report their mean values of all spectral bands. CC, SAM, and ERGAS are three widely adopted quality indices in HS fusion task, while the remaining three indices are commonly used quantitative image restoration quality indices. The best values for these indices are 1, 0, 0, 0, $+\infty$, and 1, respectively.

A. Ablation Studies

The proposed SSPSR method contains four main components including Grouping Strategy (GS), Progressive Upsampling (PU), Parameter Sharing (PS), and Spectral Attention (SA). In order to validate the effectiveness of these components, we modify our model and compare their variants. We use the training images from Chikusei dataset as a training set, and evaluate the super-resolution performance (in terms of average objective results) on the four testing images from Chikusei dataset (more details regarding the experimental settings on Chikusei dataset can be found in the following subsection). Table II tabulates the four variants of the proposed method, in which d denotes the upsampling scale. In the following, we will give the detailed analysis about them.

Grouping Strategy (GS): To effectively exploit the correlation among neighboring spectral bands of hyperspectral image and reduce the parameters of the model, we design a grouping

²[Online]. Available: <https://www.sali.t.u-tokyo.ac.jp/hyperdata/>

³[Online]. Available: http://www.ehu.eus/ccwintco/index.php?title=Hyper_spectral_Remote_Sensing_Scenes

⁴[Online]. Available: https://www.cs.columbia.edu/CAVE/databases/multi_spectral/

TABLE II
ABLATION STUDY. QUANTITATIVE COMPARISONS AMONG SOME OTHER VARIANTS OF THE PROPOSED SSPSR METHOD OVER FOUR TESTING IMAGES OF CHIKUSEI DATASET WITH RESPECT TO SIX PQIs

Models	d	CC↑	SAM↓	RMSE↓	ERGAS↓	PSNR↑	SSIM↑
Our	4	0.9565	2.3527	0.0114	4.9313	40.3612	0.9441
Our - w/o GS	4	0.9548	2.4048	0.0116	5.0399	40.1901	0.9424
Our - w/o PU	4	0.9520	2.5239	0.0119	5.2329	39.9185	0.9388
Our - w/o PS	4	0.9537	2.4152	0.0118	5.0991	40.0712	0.9410
Our - w/o SA	4	0.9563	2.3597	0.0115	4.9443	40.3408	0.9438
Our	8	0.8766	4.0127	0.0191	8.3355	35.8368	0.8538
Our - w/o GS	8	0.8622	4.5121	0.0199	8.8459	35.3857	0.8427
Our - w/o PU	8	0.8585	4.5542	0.0202	9.0285	35.2489	0.8358
Our - w/o PS	8	0.8732	4.0587	0.0194	8.4621	35.7074	0.8522
Our - w/o SA	8	0.8760	4.0198	0.0192	8.3650	35.8144	0.8538

SG: Grouping Strategy, PU: Progressive Upsampling.

PS: Parameter Sharing, SA: Spectral Attention.

strategy to divide the input hyperspectral image into some overlap groups. In order to verify the effectiveness of this strategy, we remove the grouping strategy and treat them as one group. As shown in Table II, “Our - w/o GS”, where the grouping strategy is discarded, is getting worse. The grouping strategy leads to a considerable performance improvement, *e.g.*, +0.17 dB for $\times 4$ and +0.45 dB for $\times 8$. As for other objective indicators, the gains are also considerable.

In addition to above with/without GS comparisons, we also report the number of parameters and FLOPs as well as the six PQIs of our method under some typical setting for the spectral band numbers (p) of each group and overlaps (o) between neighboring groups. The group number S is calculated by Eq. (18). As shown in Table II, when $p = 128$ and $o = 0$, our method considers all the spectral bands as a whole group ($S = 1$) and there is no grouping strategy, *i.e.*, the case of “Our - w/o GS”. When $p = 1$, $o = 0$, and $S = 128$ our method will treat each spectral band as a group and this can be seen as a special case, *i.e.*, the band-wise grouping. From the results, we can see that regardless of whether we treat all spectra as a whole or treat them separately, their performance cannot be compared with our proposed grouping strategy. When comparing the two schemes, the band-wise one obtained better performance due to the combination of grouping and parameter sharing. However, it will also greatly increase the computational overhead (please refer to the FLOPs). Because the more branches of the model, the more calculations are required.

We also report the performance of the proposed methods with different settings for the overlaps between neighboring groups, *i.e.*, $p = 8$ and $o = 0, 2, 4, 6$. With the increase of overlap (from $o = 0$ to $o = 6$), the performance of our method will be gradually improved, but the calculation amount of the model is also constantly expanding. It is worth noting that because we adopt a strategy of parameter sharing, when we fix the spectral band number p and change the overlap o , the parameters of the model are the same. In order to achieve a balance among the number of parameters and FLOPs and the objective results, in this paper, we set the p and o to 8 and 2, respectively.

Progressive Upsampling (PU): To learn the end-to-end relationship between low-resolution input and high-resolution output, there are two commonly used upsampling frameworks, pre-upsampling super-resolution and post-upsampling

super-resolution. They either increase the parameters of the network or increase the difficulty of training. Inspired by Laplacian pyramid super-resolution network [48], we leverage a progressive upsampling super-resolution framework. In this way, it decomposes a difficult task into some easy tasks, thus not only greatly reducing the learning difficulty but also obtaining better performance. In Table II, we report the performance of the proposed SSPSR method without the PU strategy, *i.e.*, “Our - w/o PU”. We remove the upsampling module in the branch networks and obtain the variant of our method. We can see that our method with PU achieves better performance on all the six indices, including the spatial reconstruction fidelity (*e.g.*, RMSE, PSNR and SSIM) and the spectral consistency (CC, SAM, and ERGAS). Especially when the upsampling factor is large, this strategy appears to be paramount. For example, the improvement of CC and PSNR of $\times 8$ is greater than that of $\times 4$, *e.g.*, +0.045 and +0.45 dB for $\times 4$, and +0.181 and +0.58 dB for $\times 8$.

Parameter Sharing (PS): In the proposed SSPSR method, in order to make the training process more efficient, we share the network parameters of each branch across all groups. In Table II, we tabulates the comparison results of the proposed SSPSR method with and without parameter sharing strategy. Obviously, by parameter sharing, we have greatly reduced the computational complexity of the model. Although parameter sharing strategy reduces the parameters of the model, it does not weaken the representation ability of the model. Through the parameter sharing strategy,⁵ we can make full use of the training samples provided by different branches (training “more” data with only one branch network parameters), so that we get a more stable model. From the results, we can see that the overall performance of the parameter sharing strategy is even better than the parametric unsharing method on all six PQIs under $d = 4$ and $d = 8$.

Spectral Attention (SA): To exploit the spatial-spectral prior, we apply the bottleneck network (with 1×1 filters) to extract the correlations among neighboring spectral bands of hyperspectral image. In addition, the attention module is also introduced to model the interdependencies between the spectra of the hyperspectral data. To verify the effectiveness of the SA module, we compare the performance of with and without SA module. As shown in Table II, with the SA mechanism, our method has achieved a slight performance gain compared to “Our - w/o SA” that without SA mechanism. By adding the SA module, although the improvement of each objective index is relatively small, the improvement of spectral confidence (*i.e.*, SAM) is more obvious than that of spatial reconstruction confidence (*i.e.*, PSNR), 2.2% vs. 0.43% for $d = 4$ and 11% vs. 1.3% for $d = 4$. This proves that the introduction of SA will be more conducive to the representation of spectral features.

B. Results on Chikusei Dataset

The Chikusei dataset is taken by Headwall Hyperspec-VNIR-C imaging sensor, and it is an urban area in Chikusei, Ibaraki,

⁵Since the network parameters are mainly dominated by module of SSPN, we can deduce that the parameter ratio between the models with and without parameter sharing is $\frac{2}{S+1}$.

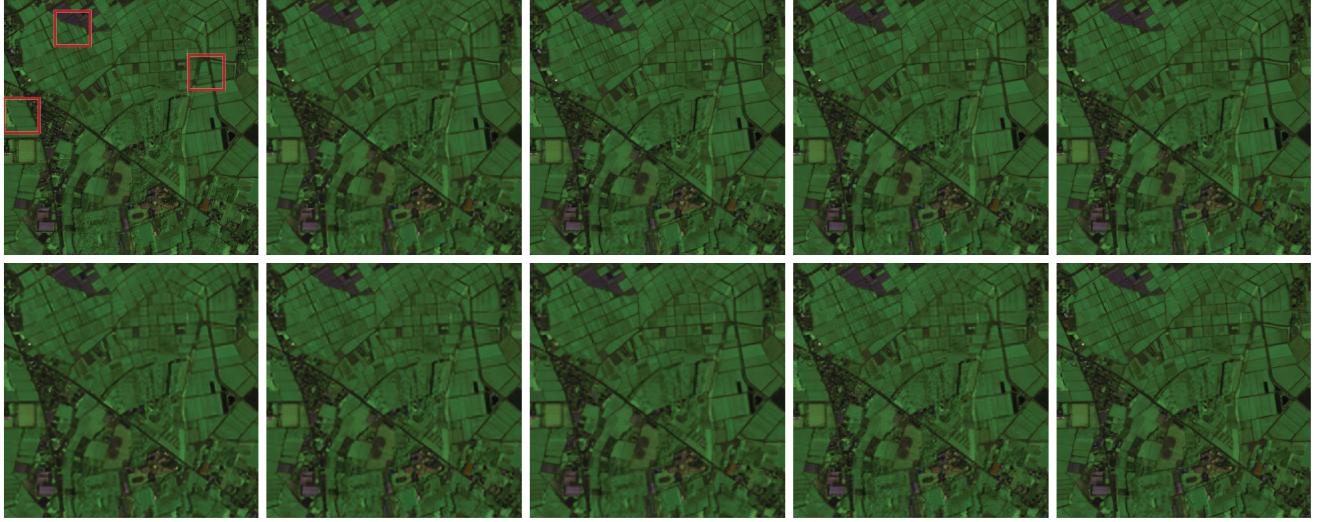


Fig. 3. Reconstructed composite images of one test hyperspectral image in Chikusei dataset with spectral bands 70-100-36 as R-G-B when the upsampling factor is $d = 4$. From left to right, top to down, they are the ground truth, results of VDSR [46], EDSR [15], RCAN [16], SAN [54], TLCNN [39], 3DCNN [41], GDRRN [42], DeepPrior [45], and the proposed SSPSR method. The bottom table shows the PSNR (dB) and SSIM results of the reconstructed RGB composite image of different methods.

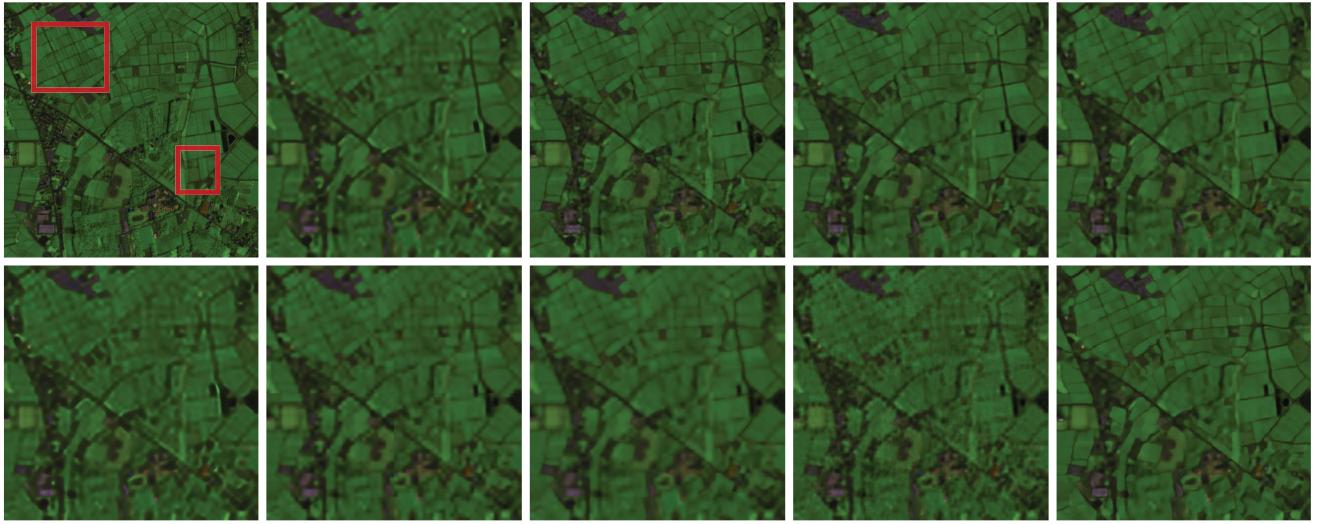


Fig. 4. Reconstructed composite images of one test hyperspectral image in Chikusei dataset with spectral bands 70-100-36 as R-G-B when the upsampling factor is $d = 8$. From left to right, top to down, they are the ground truth, results of VDSR [46], EDSR [15], RCAN [16], SAN [54], TLCNN [39], 3DCNN [41], GDRRN [42], DeepPrior [45], and the proposed SSPSR method. The bottom table shows the PSNR (dB) and SSIM results of the reconstructed RGB composite image of different methods.

Table V tabulates the average performance in terms of six PQIs over four testing images of all competing approaches. We can easily observe that the proposed SSPSR method significantly outperforms other algorithms with respect to almost all objective evaluation indexes. The average PSNR value of our method

is 0.3 dB for $\times 4$ and 0.2 dB for $\times 8$ higher than the second best method. As the most competitive general gray/RGB image super-resolution methods, EDSR, RCAN, and SAN can achieve quite pleasurable results. However, their SAM indices are relatively poor when compared with these single hyperspectral

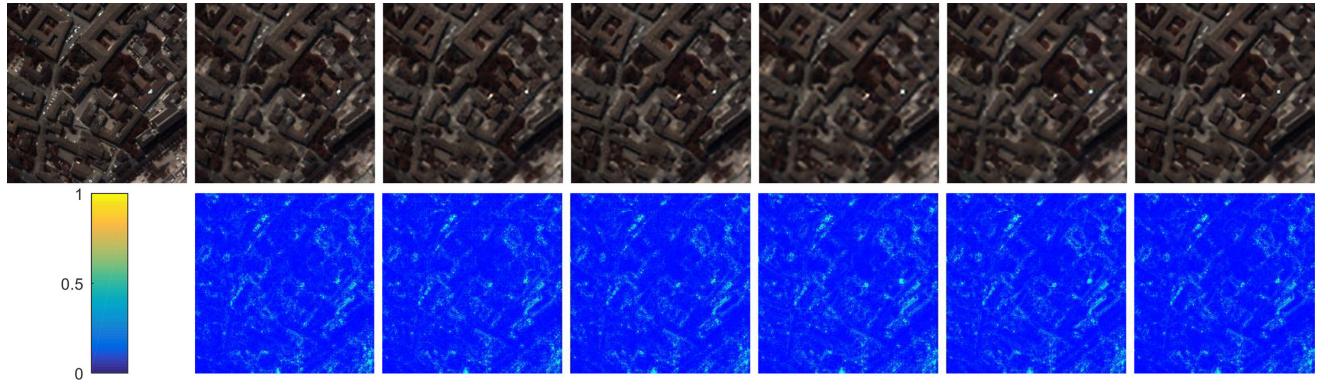


Fig. 5. Reconstructed composite images (the first row) and the error maps (the second row) of one test hyperspectral image in Pavia Center dataset with spectral bands 32-21-11 as R-G-B with upsampling factor $d = 4$. From left to right, they are the ground truth, results of EDSR [15], RCAN [16], SAN [54], 3DCNN [41], GDRRN [42], and the proposed SSPSR method. The bottom images are the reconstruction error maps of the corresponding methods.

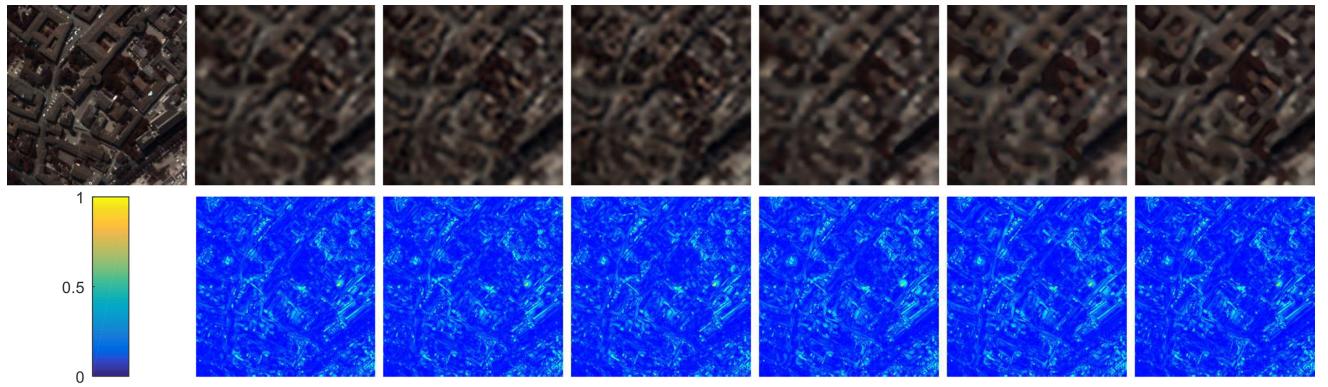


Fig. 6. Reconstructed composite images (the first row) and the error maps (the second row) of one test hyperspectral image in Pavia Center dataset with spectral bands 32-21-11 as R-G-B with upsampling factor $d = 8$. From left to right, they are the ground truth, results of EDSR [15], RCAN [16], SAN [54], 3DCNN [41], GDRRN [42], and the proposed SSPSR method. The bottom images are the reconstruction error maps of the corresponding methods.

image super-resolution methods, i.e., 3DCNN [41] and GDRRN [42].

Fig. 5 and Fig. 6 show the reconstructed composite images and error maps of one test hyperspectral image in Pavia Center dataset of the six most competitive approaches with upsampling factors $d = 4$ and $d = 8$, respectively. The results of EDSR [15], 3DCNN [41], and GDRRN [42] are very blur, while RCAN [16] and SAN [54] seem to introduce some noise. The proposed SSPSR method can maintain the main structural information. From the error maps of these methods, we can notice that the proposed method does not include obvious contour information of the image, which indicates that our method can well recover these information. It should be noted that when compared with the situation $d = 4$, the visual results with upsampling factor $d = 8$ are worse. In addition, when we compare the visual results of Fig. 4 and Fig. 6, we also notice that reconstructed results on Pavia Center dataset are worse than these on Chikusei dataset. We think this is mainly due to the limited number of the training samples of the Pavia Center database. This is also a major drawback of these deep learning based methods. That is, they require a large number of training samples, otherwise they are difficult to train a model with promising generalization ability.

D. Results on CAVE Dataset

The previous experiments are conducted on the Chikusei and Pavia Centre datasets, which are all remotely sensed hyperspectral images. To further verify the effectiveness of the proposed SSPSR method, we also conduct comparison experiments on hyperspectral images of natural scenes. Specifically, we use the CAVE multispectral image database because it is widely used in many multispectral image recovery tasks. The database consists of 32 scenes of everyday objects with spatial size of 512×512 , including 31 spectral bands ranging from 400nm to 700nm at 10 nm steps. To prepare samples for training, we randomly select 20 hyperspectral images from the database (10% samples are randomly selected for evaluations). When the upsampling factor d is 4, we extract patches with 64×64 pixels (32 pixels overlap) for training; when the upsampling factor d is 8, we let the extracted patches as 128×128 pixels (with 64 pixels overlap). The corresponding low-resolution hyperspectral image are generated by Bicubic downsampling with a factor of 4 or 8. The remaining 12 hyperspectral images of the database are used for testing, where the original images are treated as ground truth high-resolution hyperspectral images, and the low-resolution hyperspectral inputs are generated similarly as the training samples. For this dataset, we set the spectral band number (p) of

680nm by the competing methods for test images *stuffed_toys* and *real_and_fake_apples* with upsampling factors $d = 4$ and $d = 8$, respectively. From the visual reconstruction results, we can see that all the comparison methods can well reconstruct the high-resolution spatial structures of the hyperspectral images. In these error maps, we learn that the proposed method and RCAN method achieve the best reconstruction fidelity in recovering the details of the original hyperspectral images. For example, the edges of the checkerboards and the contours of dog's ears and apples (please refer to the regions marked with red boxes). In the subfigure (g), we also report the RMSE, PSNR, and SSIM results of each spectral band for the competing methods. Obviously, the proposed SSPSR method performs best in most cases. 3DCNN [41] and GDRRN [42], which are designed for the hyperspectral images, can achieve favorable results in some cases, but their performance seems to be unstable when reconstructing different spectral bands.

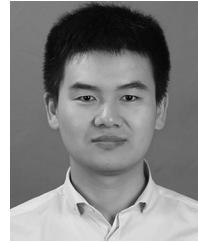
V. CONCLUSION

In this paper, a novel deep neural network based on spatial-spectral prior network (SSPN) is introduced to address the single hyperspectral image super-resolution problem. In particular, in order to discover the spatial and spatial correlation characteristics of hyperspectral data, we carefully designed a spatial-spectral prior network (SSPN) to fully exploit the spatial information and correlation among the different spectral features. In addition, to cope with the problems that the training samples of hyperspectral image are limited and the dimensionality is high, a group convolution (with shared network parameters) and progressive upsampling framework is proposed. In this way, we can expect to greatly reduce the parameters of the model and make it possible to obtain stable training results under small data and large spectral band number conditions. In our introduced network, the transmission of information flow is very flexible by the short, long, global skip links via residual learning. To regularize the network outputs, we adopt a spatial-spectral total variation (SSTV) based constraint to preserve the edge sharpness spectral correlations of the super-resolved high-resolution hyperspectral image. Evaluations on three public hyperspectral datasets demonstrate that our model not only achieves the best performance in terms of some commonly used objective indicators, but also generates clear high-resolution images which are perceptually closer to the ground truth when compared with state-of-the-arts.

REFERENCES

- [1] L. J. Rickard, R. W. Basedow, E. F. Zalewski, P. R. Silvergate, and M. Landers, "Hydice: An airborne system for hyperspectral imaging," in *Proc. Opt. Eng. Photon. Aerosp. Sens.*, 1993, pp. 173–179.
- [2] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: a technical overview," *IEEE Signal Process. Mag.*, vol. 20, no. 3, pp. 21–36, May 2003.
- [3] N. Yokoya, C. Grohfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.
- [4] Q. Wei, N. Dobigeon, and J.-Y. Tourneret, "Bayesian fusion of multi-band images," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 6, pp. 1117–1127, Sep. 2015.
- [5] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2011.
- [6] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vision*, 2014, pp. 63–78.
- [7] C. Chen, Y. Li, W. Liu, and J. Huang, "Sirf: Simultaneous satellite image registration and fusion in a unified framework," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4213–4224, Nov. 2015.
- [8] Z.-W. Pan and H.-L. Shen, "Multispectral image super-resolution via rgb image fusion and radiometric calibration," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1783–1797, Apr. 2018.
- [9] Y. Zhou, A. Rangarajan, and P. D. Gader, "An integrated approach to registration and fusion of hyperspectral and multispectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3020–3033, May 2020.
- [10] H. Huang, J. Yu, and W. Sun, "Super-resolution mapping via multi-dictionary based sparse representation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2014, pp. 3523–3527.
- [11] S. He, H. Zhou, Y. Wang, W. Cao, and Z. Han, "Super-resolution reconstruction of hyperspectral images via low rank tensor modeling and total variation regularization," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2016, pp. 6962–6965.
- [12] Y. Wang *et al.*, "Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization," *Remote Sens.*, vol. 9, no. 12, 2017, Art. no. 1286.
- [13] H. Irmak, G. B. Akar, and S. E. Yuksel, "A map-based approach for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2942–2951, Jun. 2018.
- [14] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2015.
- [15] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit. Workshops*, 2017, pp. 136–144.
- [16] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vision*, 2018, pp. 286–301.
- [17] J. Jiang, "Hyperspectral image super-resolution benchmark," 2019. [Online]. Available: <https://github.com/junjun-jiang/Hyperspectral-Image-Super-Resolution-Benchmark>
- [18] B. Aiazzi, L. Alparone, S. Baronti, A. Garzelli, and M. Selva, "Twenty-five years of pansharpening: A critical review and new developments," in *Signal Image Proc. for Remote Sensing*. Boca Raton, FL, USA: CRC Press, 2012, pp. 552–599.
- [19] X. He, L. Condat, J. M. Bioucas-Dias, J. Chanussot, and J. Xia, "A new pansharpening method based on spatial and spectral sparsity priors," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 4160–4174, Sep. 2014.
- [20] K. Li, W. Xie, Q. Du, and Y. Li, "DDLPS: Detail-based deep laplacian pansharpening for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 8011–8025, Oct. 2019.
- [21] J. Ma, W. Yu, C. Chen, P. Liang, X. Guo, and J. Jiang, "Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion," *Inf. Fusion*, vol. 62, pp. 110–120, Oct. 2020.
- [22] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2015, pp. 3631–3640.
- [23] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [24] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3034–3047, Jun. 2019.
- [25] X.-H. Han, B. Shi, and Y. Zheng, "Self-similarity constrained sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5625–5637, Nov. 2018.
- [26] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5969–5982, Dec. 2018.
- [27] M. A. Veganzones, M. Simoes, G. Licciardi, N. Yokoya, J. M. Bioucas-Dias, and J. Chanussot, "Hyperspectral super-resolution of locally low rank images from complementary multisource data," *IEEE Trans. Image Process.*, vol. 25, no. 1, pp. 274–288, Jan. 2015.
- [28] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, Oct. 2019.

- [29] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "Pannet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vision*, 2017, pp. 5449–5457.
- [30] R. Dian, S. Li, A. Guo, and L. Fang, "Deep hyperspectral image sharpening," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5345–5355, Nov. 2018.
- [31] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 2511–2520.
- [32] R. A. Borsoi, T. Imbiriba, and J. C. M. Bermudez, "Super-resolution for hyperspectral and multispectral image fusion accounting for seasonal spectral variability," *IEEE Trans. Image Process.*, vol. 29, pp. 116–127, 2020.
- [33] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by ms/hs fusion net," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 1585–1594.
- [34] B. Wen, U. S. Kamilov, D. Liu, H. Mansour, and P. T. Boufounos, "Deep-casd: An end-to-end approach for multi-spectral image super-resolution," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, 2018, pp. 6503–6507.
- [35] X. Deng and P. L. Dragotti, "Deep coupled ista network for multimodal image super-resolution," *IEEE Trans. Image Process.*, vol. 29, pp. 1683–1698, 2020.
- [36] T. Akgun, Y. Altunbasak, and R. M. Mersereau, "Super-resolution reconstruction of hyperspectral images," *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1860–1875, Nov. 2005.
- [37] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Rev.*, vol. 38, no. 3, pp. 367–426, 1996.
- [38] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral image super-resolution by spectral mixture analysis and spatial-spectral group sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1250–1254, Sep. 2016.
- [39] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.
- [40] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral image super-resolution using deep feature matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6055–6067, Aug. 2019.
- [41] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3D full convolutional neural network," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1139.
- [42] Y. Li, L. Zhang, C. Dingl, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *Proc. IEEE Int. Conf. Multimedia Big Data*, 2018, pp. 1–4.
- [43] H. Sun, Z. Zhong, D. Zhai, X. Liu, and J. Jiang, "Hyperspectral image super-resolution using multi-scale feature pyramid network," in *International Forum on Digital TV and Wireless Multimedia Communications*. Berlin, Germany: Springer, 2019, pp. 49–61.
- [44] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 9446–9454.
- [45] O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vision Workshop*, 2019, pp. 3844–3851.
- [46] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 1646–1654.
- [47] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 1637–1645.
- [48] 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. Comput. Vision Pattern Recognit.*, 2017, pp. 624–632.
- [49] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 3147–3155.
- [50] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 2472–2481.
- [51] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 1664–1673.
- [52] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang, "Non-local recurrent network for image restoration," in *Proc. Ad. Neural Inf. Process. Syst.*, 2018, pp. 1673–1682.
- [53] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2018, pp. 7132–7141.
- [54] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2019, pp. 11 065–11 074.
- [55] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 1874–1883.
- [56] E. Wycoff, T.-H. Chan, K. Jia, W.-K. Ma, and Y. Ma, "A non-negative sparse promoting algorithm for high resolution hyperspectral imaging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 1409–1413.
- [57] H. K. Aggarwal and A. Majumdar, "Hyperspectral image denoising using spatio-spectral total variation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 442–446, Mar. 2016.
- [58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015.
- [59] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over Chikusei," Space Application Laboratory, University of Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27, May 2016.
- [60] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, Sep. 2010.
- [61] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [62] R. H. Yuhas, A. F. Goetz, and J. W. Boardman, "Discrimination among semi-arid landscape endmembers using the spectral angle mapper (sam) algorithm," in *Proc. JPL, Summaries 3rd Annu. JPL Airborne Geosci. Workshop*, 1992, vol. 1, pp. 147–149.
- [63] L. Wald, *Data Fusion: Definitions and Architectures: Fusion of Images of Different Spatial Resolutions*. Paris, France: Presses des MINES, 2002.
- [64] Z. Wang, A. C. Bovik, H. R. Sheikh, 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.



Junjun Jiang received the B.S. degree in Mathematics from Huaqiao University, Quanzhou, China, in 2009, and the Ph.D. degree in Computer Science from Wuhan University, Wuhan, China, in 2014.

From 2015 to 2018, he was an Associate Professor with the School of Computer Science, China University of Geosciences, Wuhan. From 2016 to 2018, he was a Project Researcher with the National Institute of Informatics (NII), Tokyo, Japan. He is currently a Professor with the School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China. His research interests include image processing and computer vision. He is the Associate Editor of Remote Sensing. He won the Best Student Paper Runner-up Award at MMM 2017, the Finalist of the World's FIRST 10K Best Paper Award at ICME 2017, and the Best Paper Award at IFTC 2018. He received the 2016 China Computer Federation (CCF) Outstanding Doctoral Dissertation Award and 2015 ACM Wuhan Doctoral Dissertation Award.

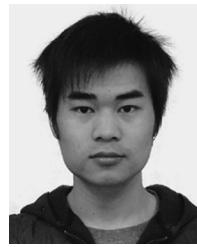


He Sun received the B.E. degree in mechatronics from the Harbin Institute of Technology, Harbin, China, in 2019. He is currently pursuing the M.S. degree in computer science at the Harbin Institute of Technology. His research interests include image processing, computer vision and its application in robotics.



Xianming Liu received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 2006, 2008, and 2012, respectively. In 2011, he spent half a year at the Department of Electrical and Computer Engineering, McMaster University, Canada, as a Visiting Student, where he was a Post-Doctoral Fellow from 2012 to 2013. He was a Project Researcher with the National Institute of Informatics (NII), Tokyo, Japan, from 2014 to 2017. He is currently a Professor with the School of Computer Science and Technology, HIT.

He was a receipt of the IEEE ICME 2016 Best Student Paper Award.



Jiayi Ma received the B.S. degree in information and computing science and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2008 and 2014, respectively.

From 2012 to 2013, he was an Exchange Student with the Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA. He was a Post-Doctoral with the Electronic Information School, Wuhan University from August 2014 to November 2015, and received an accelerated promotion to Associate Professor and Full Professor in December 2015 and December 2018, respectively. He has authored or co-authored more than 110 refereed journal and conference papers. His research interests include computer vision, machine learning, and pattern recognition.

Dr. Ma was a recipient of the Natural Science Award of Hubei Province (first class), the CAAI (Chinese Association for Artificial Intelligence) Excellent Doctoral Dissertation Award, and the CAA (Chinese Association of Automation) Excellent Doctoral Dissertation Award. He has been an Associate Editor of Neurocomputing since 2018 and a Guest Editor of Remote Sensing.