

# Structure–Color Preserving Network for Hyperspectral Image Super-Resolution

Bin Pan<sup>ID</sup>, Member, IEEE, Qiaoying Qu<sup>ID</sup>, Xia Xu<sup>ID</sup>, Graduate Student Member, IEEE,  
and Zhenwei Shi<sup>ID</sup>, Member, IEEE

**Abstract**—Fusion-based hyperspectral super-resolution (HSR) algorithms usually utilize a low-resolution hyperspectral image (LR-HSI) and a high-resolution multispectral image (MSI) to generate a high-resolution hyperspectral image (HR-HSI), which have attracted increasing attention in recent years. However, how to deal with the abundant spectral information of hyperspectral images and complex structure characteristics of MSIs has always been the focus and difficulty of fusion-based HSR. In this article, we propose a new structure–color preserving network (SCPNet) for HSR, which is developed under the basis of the joint attention mechanism. The SCPNet mainly includes three modules: structure-preserving module (SPM), color-preserving module (CPM), and cross-fusion module. The SPM is constructed based on the spatial attention, which aims to capture and enhance the significant structure information from the high-resolution MSI. Meanwhile, the CPM is constructed based on the channel attention, where the spectral characteristics in the LR-HSI are preserved during the reconstruction process. Finally, we propose a cross attention-based cross-fusion strategy to integrate the features from the two branches and reconstruct the final HR-HSI. The major contribution of SCPNet is that the structure and color information is described and preserved via the joint attention mechanism. Experimental results indicate that the proposed SCPNet has presented advantages on three benchmark datasets when compared with some state-of-the-art HSR methods.

**Index Terms**—Attention mechanism, hyperspectral super-resolution (HSR), structure–color preserving.

## I. INTRODUCTION

HYPERSPECTRAL image sensors collect hundreds of wavelengths ranging from visible to long-wave infrared [1], [2]. Therefore, the hyperspectral images (HSIs) contain abundant spectral information, which has made contributions to quite a few applications such as image classi-

Manuscript received September 22, 2021; revised November 14, 2021 and November 30, 2021; accepted December 5, 2021. Date of publication December 13, 2021; date of current version March 1, 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2019YFC1510905, in part by the National Natural Science Foundation of China under Grant 62001251 and Grant 62001252, and in part by the China Postdoctoral Science Foundation under Grant 2020M670631.

(Corresponding author: Xia Xu.)

Bin Pan and Qiaoying Qu are with the School of Statistics and Data Science and the Key Laboratory of Pure Mathematics and Combinatorics of Ministry of Education, Nankai University, Tianjin 300071, China, and also with the Science and Technology on Special System Simulation Laboratory, Beijing Simulation Center, Beijing 100854, China (e-mail: panbin@nankai.edu.cn; quqiaoying@mail.nankai.edu.cn).

Xia Xu is with the College of Computer Science, Nankai University, Tianjin 300071, China (e-mail: xuxia@nankai.edu.cn).

Zhenwei Shi is with the Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China (e-mail: shizhenwei@buaa.edu.cn).

Digital Object Identifier 10.1109/TGRS.2021.3135028

fication [3]–[5] and target detection [6], [7]. However, HSIs usually suffer from low spatial resolution due to the limitations of hardware. Researchers adopt two methods to acquire a high-resolution hyperspectral image (HR-HSI): improving the spatial resolution of low-resolution hyperspectral images (LR-HSI) [8]–[10] and reconstructing spectral information from high-resolution RGB/multispectral image (HR-RGB/HR-MSI) [11], [12]. However, since the complex correlation among channels of hyperspectral images is difficult to be reconstructed from RGB/multispectral images (MSIs), it is a more effective method to reconstruct HR-HSI by super-resolving LR-HSI. According to the difference of the required inputs, current hyperspectral super-resolution (HSR) methods can be roughly divided into two categories: single-image-based super-resolution with just one input and fusion-based super-resolution with two input images.

Single-image-based HSR methods directly rely on one LR-HSI as the input. Since no auxiliary information is available, single-image-based methods often lead to spectral or texture distortions, especially when the upscaling factor is large. To solve the insufficiency of priors, researchers attempt to exploit the abundant spectral correlations among spectral bands. Some methods based on sparse and dictionary representation or low-rank prior have been proposed [10], [13], [14]. Moreover, deep learning-based single-image super-resolution methods have achieved excellent performance [15]–[18]. Based on single-image super-resolution, researchers proposed many hyperspectral image super-resolution methods [17], [19], [20]. Liu *et al.* [17] employed group convolutions and covariance statistics-based attention mechanism to explore the consecutive information. Mei *et al.* [20] combined the cross-scale nonlocal prior with local and in-scale nonlocal priors to improve the performance. However, single-image-based HSR methods usually require critical priori information, but the manually designed prior may not be well representations of the data.

Fusion-based HSR refers to generating an HR-HSI using an LR-HSI and a registered MSI, which has attracted much attention in recent years [21], [22]. Compared with single-image-based HSR, fusion-based HSR is more feasible since most hyperspectral platforms usually integrate synchronous multispectral sensors as well. In this case, the MSIs usually have a higher spatial resolution, while the LR-HSIs have a finer spectral resolution. Traditional methods, such as matrix factorization-based algorithms [23]–[27], are proposed first. These algorithms, respectively, decompose the LR-HSI and

the high-resolution multispectral image (HR-MSI) into a coefficient matrix and a basis matrix under some priors. Moreover, tensor decomposition-based models have also been widely utilized since the hyperspectral image is a 3-D cube. In these methods, hyperspectral images are represented as 3-D tensors and the image tensors are decomposed into the product of kernel tensors and projection matrices using tensor decomposition techniques, which could account simultaneously for all spectral–spatial information [28]–[32]. In addition to the above two algorithms, Bayesian rule-based algorithms have been widely applied in HSR. The dictionary is obtained through the Bayesian dictionary learning, and then, the HR-HSI is reconstructed with dictionary and sparse coding matrix [33]–[36].

Recently, deep learning techniques, especially the convolutional neural networks, have presented a promising performance in HSR tasks [37]–[40]. For example, in order to adapt to the characteristics of hyperspectral images as 3-D cube, some methods apply 3-D convolution to CNN [41]–[43]. Mei *et al.* [41] proposed a 3-D CNN-based algorithm, which showed that the 3-D CNN could achieve excellent performance in HSR. These methods facilitate the representation of correlations among successive bands. Besides, the correlations can also be enhanced with sufficient capture of the residual information among spectral bands [8], [44]. Xie *et al.* [8] proposed a network that could capture the deep residual features of high-frequency information and utilize the learned features as *a priori* in the HR-HSI reconstruction. For a fusion-based HSR task, the cross fusion of spatial information and spectral information is critical to the reconstruction performance. Therefore, researchers mostly focus on developing networks, which could extract significant characteristics and integrate them [45]–[48]. Han *et al.* [48] utilized a multilevel network to upscale the spatial resolution of LR-HSI gradually and integrated the multiscale loss functions during the training to avoid the gradient vanish. Moreover, some alternative super-resolution methods have been proposed. Such as unsupervised HSR [49], [50] and HSR algorithms consider point spread function (PSF) [9]. Qu *et al.* [49] adopted the mutual information and assumed that the characteristics follow a similar Dirichlet distribution. Kwan *et al.* [9] super-resolved LR-HSI with method, which incorporates PSF into the deblurring and then fuses an HR color image with enhanced HSI.

However, recent deep learning-based HSR algorithms may suffer from the color and structure distortions. They usually utilize feature extractors that are not appropriate for both two inputs. In this way, the insufficiency of feature representation for the spectral information and the structure characteristics leads to loss of color–structure information. Therefore, how to design a network that simultaneously considers the color and structure characteristics remains a challenge.

In this article, we propose a structure–color preserving network (SCPNet) for fusion-based hyperspectral image super-resolution, which aims at extracting the spatial details from MSIs while preserving the spectral information in the LR-HSIs. The kernel of SCPNet is a newly developed joint attention mechanism, which is composed of three modules: structure-preserving module (SPM), color-preserving module (CPM), and the fusion module. The SPM is designed to

capture the significant structure information from the MSIs and to introduce the structure details to the obtained HR-HSIs based on the spatial attention. Meanwhile, the CPM tries to preserve the spectral characteristics in the LR-HSIs during the reconstruction process via a channel attention approach. Finally, the SPM and CPM are integrated based on a new cross attention-based fusion strategy.

The major contributions of SCPNet can be summarized as follows.

- 1) We propose a new spatial attention-based SPM to extract the structure details from MSIs.
- 2) We propose a new channel attention-based CPM to provide spectral invariance from LR-HSIs.
- 3) We design a new cross attention-based cross-fusion strategy to achieve joint spatial–spectral information preservation for the final obtained HR-HSIs.

## II. METHODOLOGY

This section presents the architecture of SCPNet, which consists of three parts: 1) SPM; 2) CPM; and 3) cross-fusion strategy.

### A. Network Architecture

Some notations of terms are as follows. The two inputs HR-MSI and LR-HSI and the reconstructed HR-HSI are denoted as  $X \in \mathbb{R}^{swsh \times c}$ ,  $Y \in \mathbb{R}^{wh \times C}$ , and  $Z \in \mathbb{R}^{swsh \times C}$ , respectively, where  $h$  and  $w$  represent the height and width of LR-HSI, respectively.  $C$  and  $c$  represent the number of channels of HR-MSI and LR-HSI, respectively, while  $s$  denotes the upscaling factor. The degradation model is as follows:

$$X = ZS + N_1 \quad (1.1)$$

$$Y = DZ + N_2 \quad (1.2)$$

where  $S \in \mathbb{R}^{C \times c}$  is the spectral response function and  $D \in \mathbb{R}^{wh \times swsh}$  is the downsampling operation.

The flowchart of SCPNet is shown in Fig. 1. We first use bicubic interpolation to super-resolve the LR-HSI to the specified resolution. After upsampling operation, LR-HSI is divided into four groups on average according to the number of spectral bands. Then, the four groups are input to branches with the same parameters, which reduces the number of parameters and improves the training speed.

After a few convolution and ReLU layers, SPM and CPM are utilized for deep feature extraction. Let  $\text{Conv}(\cdot)$  and  $\text{ReLU}(\cdot)$  denote convolution layer and ReLU function, respectively.  $\text{SPM}(\cdot)$  and  $\text{CPM}(\cdot)$  represent the SPM and CPM, respectively. Then, the feature extraction operations of SCPNet can be expressed as

$$F_2 = \text{Conv}(\text{ReLU}(\text{Conv}(X))) \quad (1.3)$$

$$X' = \text{SPM}(F_2) \quad (1.4)$$

$$[Y_1, Y_2, Y_3, Y_4] = \text{Conv}(Y) \quad (1.5)$$

$$F_{1i} = \text{Conv}(\text{ReLU}(\text{Conv}(Y_i))), \quad i = 1, 2, 3, 4 \quad (1.6)$$

$$F_{2i} = \text{CPM}(F_{1i}), \quad i = 1, 2, 3, 4 \quad (1.7)$$

$$Y' = [F_{21}, F_{22}, F_{23}, F_{24}] \quad (1.8)$$

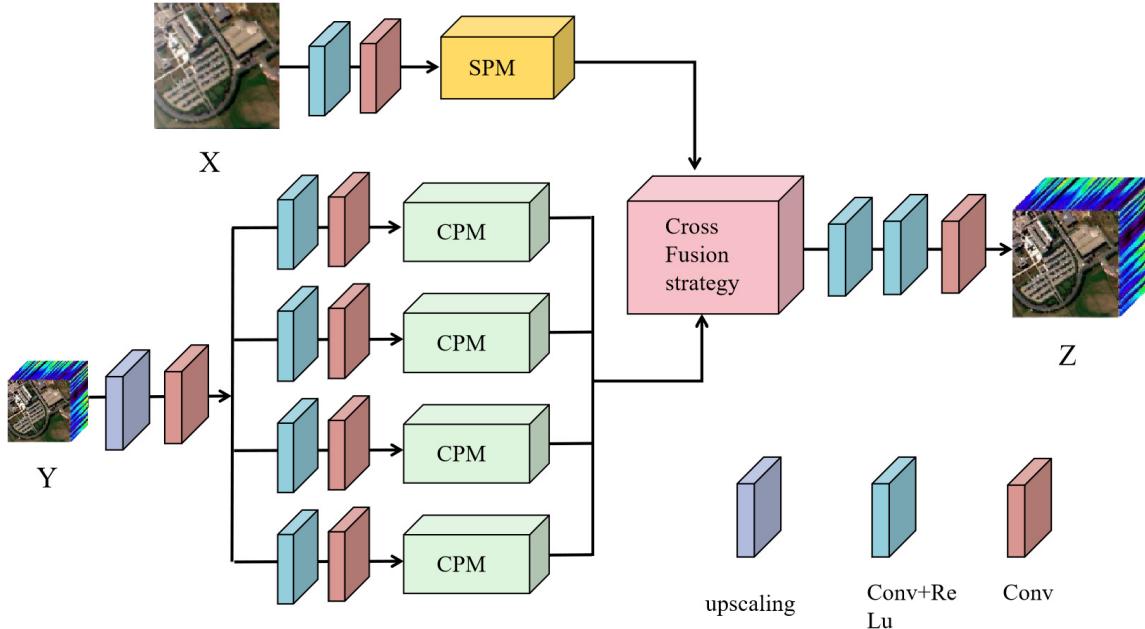


Fig. 1. Illustration of the proposed SCPNet framework. This network contains three parts: SPM, CPM, and cross-fusion strategy. SPM and CPM are proposed for feature extraction of HR-MSI and LR-his, while the cross-fusion strategy is proposed for the fusion of spatial-spectral information.

where  $X'$  and  $Y'$  are feature maps through SPM and CPM.

Spatial information determines the texture and details, whereas spectral information determines color of an image. Thus, through capturing and enhancing spatial-spectral information, SCPNet could obtain an image with more accurate structure and color.

Most fusion-based HSR methods realize feature fusion by concatenating feature maps. Inspired by [50], we adopt a newly proposed cross-fusion strategy based on a cross attention mechanism. The fusion strategy realizes cross fusion under the guidance of spatial-spectral information, which ensures the preserving of structure-color characteristics. The formula is as follows:

$$Z = \text{CF}(X', Y') \quad (1.9)$$

where  $\text{CF}(\cdot)$  is the cross-fusion strategy.

### B. Spatial Attention-Based SPM

HR-MSI possesses abundant spatial information, which is significant for the structure preserving of reconstructed HR-HSI. Thus, we propose the SPM to extract the spatial characteristics. The architecture of SPM is shown in Fig. 2.

SPM consists of a spatial attention block and a feature-preserving block. The spatial attention block captures and enhances the spatial details important for reconstruction, while the feature-preserving block facilitates the representation of details. The feature-preserving block is composed of stacked convolution layers and ReLU activation functions. Let  $\text{SAB}(\cdot)$  denote the spatial attention block. The SPM can

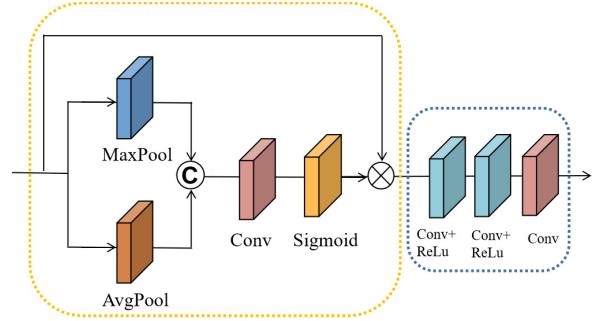


Fig. 2. Illustration of SPM, which consists of a spatial attention block and a feature extraction block. The spatial attention block is in the yellow box and the feature-preserving block is in the blue box.

be expressed as

$$X' = \text{Conv}(\text{ReLU}(\text{Conv}(\text{ReLU}(\text{Conv}(\text{SAB}(F)))))) \quad (2.1)$$

$$\text{SAB}(F) = \text{Ms}(F) \otimes F \quad (2.2)$$

$$\text{Ms}(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F), \text{MaxPool}(F)])). \quad (2.3)$$

$\text{AvgPool}(\cdot)$  represents the average pooling operation, while  $\text{MaxPool}(\cdot)$  represents the maximum pooling operation.  $f^{7 \times 7}(\cdot)$  denotes the  $7 \times 7$  convolution layer and  $\sigma(\cdot)$  represents the sigmoid activation function.

As the pixel values of an image obey a certain distribution, there are strong similarities among image patches in a feature map. We extract the correlated characteristics through maximum pooling, average pooling, and sigmoid function and utilize them for subsequent feature extraction. The coherence could be explicitly modeled by learning the weight of each pixel.

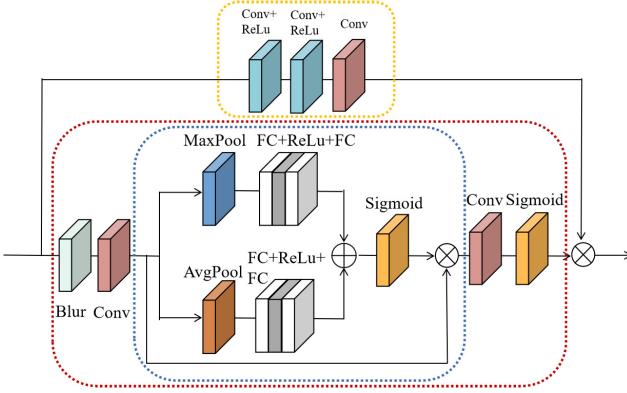


Fig. 3. Illustration of CPM, which contains a color correction branch and a feature-preserving branch. The color correction module is in the red box, the channel attention block is in the blue box, and the feature-preserving block is in the yellow box.

### C. Channel Attention-Based Color-Preserving Module

We propose CPM for the feature extraction of LR-HSI, which consists of two branches: the color correction branch and the feature-preserving branch. With the color correction branch, we obtain a set of weight coefficients, which could guide the feature representation of feature-preserving block. We first apply a Gaussian blur kernel to the input feature maps and then two convolution layers, a channel attention block and a sigmoid function. The blurring operation ensures that only the low-frequency information such as color characteristic passes through this branch, while high-frequency information such as fine texture of feature maps is blocked. Apart from the blur kernel, channel attention is another unit that facilitates the color preserving. Then, the other feature-preserving branch focuses on the representation of spectral information to promote the color preserving of reconstruction. Finally, we multiply the output of feature-preserving block and the weight coefficients from the color correction branch to obtain the result. The structure of CPM is shown in Fig. 3.

Let  $CCM(\cdot)$  represent the color correction operation. Then, the CPM can be expressed as

$$\alpha_i = CCM(F_{1i}), \quad i = 1, 2, 3, 4 \quad (3.1)$$

$$F'_{1i} = \text{Conv}(\text{ReLU}(\text{Conv}(\text{ReLU}(\text{Conv}(F_{1i}))))), \quad i = 1, 2, 3, 4 \quad (3.2)$$

$$F_{2i} = \alpha_i \otimes F'_{1i}, \quad i = 1, 2, 3, 4 \quad (3.3)$$

$$CCM(\cdot) = \sigma(\text{Conv}(\text{CAB}(\text{Conv}(\text{Blur}(\cdot))))) \quad (3.4)$$

$$\text{CAB}(F) = \text{Mc}(F) \otimes F \quad (3.5)$$

$$\text{Mc}(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) \quad (3.6)$$

$$\text{MLP}(\cdot) = \text{FC}(\text{ReLU}(\text{FC}(\cdot))) \quad (3.7)$$

where  $F_{1i}$  and  $F_{2i}$  are the input and output of CPM,  $\alpha_i$  is the weight coefficients from the color correction branch for the  $i$ th group of LR-HSI,  $\text{Blur}(\cdot)$  represents the blur kernel,  $\text{CAB}(\cdot)$  denotes the CAB operation,  $\text{MLP}(\cdot)$  represents the MLP layers, which contains two fully connected layers and a ReLU layer, and  $\sigma(\cdot)$  represents the sigmoid activation function.

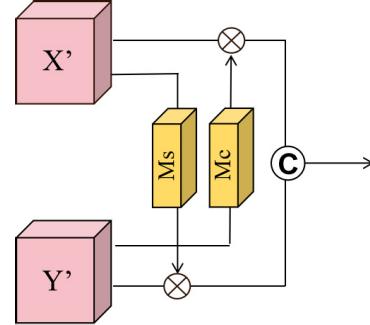


Fig. 4. Illustration of cross-fusion strategy.  $Ms$  module is utilized to compute the spatial weight coefficients, and the structure is the same as the spatial attention block. Meanwhile, the  $Mc$  module is adopted to compute the spectral weight, and the structure is the same as channel attention block.

Since the spectral coherence of hyperspectral images is critical to reconstruction, we employ the channel attention block to extract and retain this feature. Average pooling and maximum pooling integrate feature characteristics along the spatial dimensions, which forces the CPM to focus on the spectral relation. Then, the sigmoid function converts spectral information into weights, which explicitly expresses the spectral information. In this way, the spectral information and interspectral correlation significant for hyperspectral reconstruction are integrated into the feature maps.

---

### Algorithm 1 Pseudocode for SCPNet

---

```

Input: HR MSI  $X$ , LR HSI  $Y$ .
Output: reconstructed HR HSI  $Z$ .
1 while the SCPNet has not converged yet do
2 For  $X$ :
3 Shallow feature extraction:
    $F_1 = \text{Conv}(\text{ReLU}(\text{Conv}(X)))$ 
4 Input the shallow features into the SPM:  $X' = SPM(F_1)$ 
5 For  $Y$ :
6 Upsampling the HR-HSI:  $Y_1 = \text{Upsampling}(Y)$ 
7 Divide the upscaled image into four groups:
    $[Y_1, Y_2, Y_3, Y_4] = \text{Conv}(Y_1)$ 
8 Shallow feature extraction:
    $F_{1i} = \text{Conv}(\text{ReLU}(\text{Conv}(Y_i))) \quad i = 1, 2, 3, 4$ 
9 Input the shallow features into the CPM:
    $F_{2i} = \text{CPM}(F_{1i})$ 
10 Feature fusion:
11  $Z = FC(X', Y')$ 
12 Return  $Z$ 
```

---

### D. Cross Attention-Based Cross-Fusion Strategy

In order to further exploit the spatial-spectral information from the SPM and CPM, we apply a cross-fusion strategy to fuse the feature maps across branches. First, we compute spatial attention weight coefficients from the branch of HR-MSI ( $X'$ ) and spectral attention weight coefficients from the branch of LR-HSI ( $Y'$ ). Then, we multiply the input feature maps with the attention weight coefficients from the other

branch to integrate the significant information. In this way, spatial information from the HR-MSI branch could be integrated into the LR-HSI branch, and spectral information from the LR-HSI branch could be fused into the HR-MSI branch. Then, we concatenate the feature maps after information transfer and input the resulting feature map into several convolution and activation function layers.

As shown in Fig. 4, the cross-fusion strategy can be expressed as

$$\text{Ms}(X') = \sigma(f^{7 \times 7}([\text{AvgPool}(X'), \text{MaxPool}(X')])) \quad (4.1)$$

$$\text{Mc}(Y') = \sigma(\text{MLP}(\text{AvgPool}(Y')) + \text{MLP}(\text{MaxPool}(Y'))) \quad (4.2)$$

$$X'' = \text{Ms}(Y') \otimes X' \quad (4.3)$$

$$Y'' = \text{Mc}(X') \otimes Y' \quad (4.4)$$

$$Z = \text{Conv}(\text{Relu}(\text{Conv}(\text{Relu}(\text{Conv}(X'', Y''))))). \quad (4.5)$$

#### E. Loss Function

To obtain the optimal network parameter set, the losses between training samples and ground truth need to be minimized. The *L*<sub>2</sub> loss is generally utilized in most methods to maximize the peak signal-to-noise ratio (PSNR). However, the *L*<sub>2</sub> loss often fails to capture the underlying multimodal distributions of the HR patches, which results in the oversmoothness of the reconstructed images. Therefore, we select the *L*<sub>1</sub> loss since it provides better convergence. The loss function is shown in the following equation:

$$\text{loss} = \sum_{i=1}^N |I(i) - \hat{I}(i)|. \quad (5.1)$$

More training details are provided in Section III.

### III. EXPERIMENTS

This section presents the experiment results and analysis. To validate the superiority of SCPNet, we conduct comparison experiments and ablation experiments on three datasets: ICVL dataset, DFC2018 Houston dataset, and TG1HRSSC dataset.

#### A. Compared Methods and Performance Evaluation Measures

Seven methods are selected as comparative methods: GSA [51], CNMF [23], FUMI [25], CSTF [52], GLORIA [53], TFNet [54], and HiBCD [55]. Among them, GSA, CNMF, and FUMI are both representative matrix factorization-based HSR methods; GLORIA, HiBCD, TFNet, and CSTF are the state-of-the-art HSR methods. GLORIA and HiBCD are based on matrix factorization, while CSTF and TFNet are factorization- and deep learning-based HSR methods, respectively.

We utilize PSNR, structural similarity index (SSIM), and spectral angle mapper (SAM) to evaluate the quality of reconstructed HSIs. The better the reconstruction effect, the higher the PSNR and SSIM value, the lower the SAM value.

#### B. Datasets and Settings

The datasets we utilize are listed as follows.

1) *ICVL*: The ICVL dataset contains 201 images. These hyperspectral images cover a 400-700-nm spectral range with 31 bands. Subimages with size of  $31 \times 256 \times 256$  are utilized as ground truth. Then, the corresponding LR-HSIs are constructed by applying a Gaussian spatial filter on each band of the HR-HSI and downsampling every 2/4/8 pixels in both height and width directions. A three-band HR-MSI with size of  $256 \times 256$  is constructed by filtering the HR-HSI with a spectral responses function [27].

2) *DFC2018 Houston*: The data were obtained over the University of Houston campus and its neighborhood. These HSIs were collected with 50 spectral channels from 380 to 1050 nm. Subimages with size of  $50 \times 200 \times 200$  are used as the ground truth. Then, the corresponding LR-HSI and HR-MSI are constructed via the same operations as ICVL.

3) *TG1HRSSC*: TG1HRSSC dataset is a space hyperspectral remote sensing scene classification dataset acquired by Tiangong-1 hyperspectral imager. We select visible near-infrared spectral data that cover 400–900 nm with 54 effective bands for the experiment. Subimages with size of  $54 \times 256 \times 256$  are used as the ground truth. Then, the corresponding LR-HSI and HR-MSI are constructed via the same operations as ICVL.

#### C. Comparison Experiments on Synthetic Datasets

In order to verify the performance of the proposed method SCPNet for HSR, we conduct comparison experiments on three hyperspectral datasets. Also, we choose three kinds of upscaling factors for experiments: 2, 4, and 8. The details and analysis of experiments can be seen as follows.

1) *Performance on ICVL Dataset*: We randomly select 60 subimages from the ICVL dataset for training and 20 for testing. Then, we construct the corresponding LR-HSIs and HR-MSIs with Gaussian filter and SRF, respectively. Table I summarizes the PSNR, SSIM, and SAM values of SCPNet and comparative algorithms for different upscaling factors. SSIM reflects the structural difference between reconstructed images and real images, while SAM measures the spectral difference between two images. Thus, the performance of SPM and CPM can be demonstrated by the values of SSIM and SAM. Moreover, PSNR represents the overall difference between the reconstructed image and the real one, and thus, PSNR could also reflect the reconstruction effect. For the representative traditional optimization-based super-resolution algorithms, CNMF performs better than GSA, which indicates that appropriate prior information benefits super-resolution performance. FUMI provides with better results than CNMF, which means that sum-to-one and nonnegativity constraints help improve the performance. For the state-of-the-art super-resolution comparative algorithms, the performance of TFNet is the worst in the  $\times 2$  case. HiBCD performs poor in the  $\times 4$  case. Also, in  $\times 8$  case, GLORIA is the worst. CSTF surpasses GLORIA, TFNet, and HiBCD a lot in both  $\times 4$  and  $\times 8$  cases, while it gets worse performance than GLORIA in the  $\times 2$  case. Compared with both deep learning- and optimization-based comparison methods, SCPNet has the best performance in all three cases. Since the SSIM and SAM of

TABLE I  
PSNR, SSIM, AND SAM VALUES BY DIFFERENT METHODS ON THE ICVL DATASET

Scale		GSA	CNMF	FUMI	HiBCD	CSTF	GLORIA	TFNet	SCPNet
$\times 2$	PSNR	29.3731	35.3415	33.5582	33.8501	38.0348	40.2977	38.5241	<b>40.8147</b>
	SSIM	0.6453	0.8454	0.7345	0.9630	0.8653	0.9810	0.9134	<b>0.9906</b>
	SAM	6.0583	3.3435	4.6143	5.1462	2.4734	2.3234	3.2747	<b>1.4722</b>
$\times 4$	PSNR	29.0430	34.3442	36.9734	30.4961	38.3786	34.9232	37.8434	<b>42.2033</b>
	SSIM	0.7307	0.8106	0.8918	0.9316	0.8869	0.9740	0.9359	<b>0.9889</b>
	SAM	9.7251	3.8963	2.8703	6.6594	2.3228	4.4032	2.8148	<b>1.4497</b>
$\times 8$	PSNR	25.8619	38.3451	37.2233	29.3375	39.6238	28.7139	37.0159	<b>40.1055</b>
	SSIM	0.7716	0.6987	0.9471	0.9252	0.9164	0.9409	0.9388	<b>0.9889</b>
	SAM	13.2301	5.0453	2.6378	7.5202	1.9213	11.1151	3.0403	<b>1.6079</b>

TABLE II  
PSNR, SSIM, AND SAM VALUES BY DIFFERENT METHODS ON THE DFC2018 HOUSTON DATASET

Scale		GSA	CNMF	FUMI	HiBCD	CSTF	GLORIA	TFNet	SCPNet
$\times 2$	PSNR	32.2142	38.6136	39.3360	28.8065	42.3147	38.9604	35.3147	<b>43.1563</b>
	SSIM	0.8152	0.9314	0.9406	0.8894	0.9455	0.9731	0.9569	<b>0.9907</b>
	SAM	5.8020	3.1126	2.7528	17.0610	2.6018	3.4616	2.9131	<b>1.3103</b>
$\times 4$	PSNR	31.5929	36.7528	39.8612	25.9030	42.0802	30.1886	34.7288	<b>42.2311</b>
	SSIM	0.8658	0.9040	0.9562	0.8541	0.9462	0.9195	0.9519	<b>0.9889</b>
	SAM	6.2262	3.6216	2.5605	20.5555	2.6545	11.9335	3.2387	<b>1.4447</b>
$\times 8$	PSNR	28.3718	33.9525	37.4007	24.6918	<b>41.5519</b>	25.9161	35.4862	41.1030
	SSIM	0.8786	0.8441	0.9457	0.8318	0.9453	0.8651	0.9574	<b>0.9897</b>
	SAM	9.7769	4.6247	3.2550	23.0552	2.7339	22.3022	2.7990	<b>1.6928</b>

TABLE III  
PSNR, SSIM, AND SAM VALUES BY DIFFERENT METHODS ON THE TG1HRSSC DATASET

Scale		GSA	CNMF	FUMI	HiBCD	CSTF	GLORIA	TFNet	SCPNet
$\times 2$	PSNR	37.1437	42.4636	43.1266	41.4067	<b>46.7022</b>	40.9173	35.6435	43.5513
	SSIM	0.8679	0.9370	0.9733	0.9492	0.9708	0.9505	0.9192	<b>0.9754</b>
	SAM	6.2052	4.0938	<b>2.3896</b>	6.3981	2.5533	4.9034	5.0556	3.4060
$\times 4$	PSNR	36.3512	41.1210	45.4276	37.2205	<b>45.9850</b>	33.0627	36.4373	42.7133
	SSIM	0.8706	0.9250	0.9738	0.9133	0.9690	0.8227	0.9313	<b>0.9829</b>
	SAM	6.9451	4.3561	2.7147	11.9277	2.8272	15.5473	4.6637	<b>2.5809</b>
$\times 8$	PSNR	33.8293	39.0538	<b>42.7513</b>	34.8726	41.8198	30.6919	35.7888	41.6064
	SSIM	0.8298	0.8975	0.9498	0.8717	0.9212	0.7583	0.9221	<b>0.9800</b>
	SAM	11.0983	5.1371	4.2480	17.9753	4.5432	23.8257	5.1649	<b>2.8458</b>

SCPNet are both better than others, we assume that SPM and CPM successfully capture and preserve the structure characteristics and color information respectively. Especially for SAM, our SCPNet is far superior to the others. For example, when the upscaling factor is 4, SCPNet is the only one with an SAM value below 2. Even the second lowest SAM of method CSTF is almost 1.5 times bigger than that of ours. In addition, when the LR-HSI is upsampled eight times, the value of SSIM is significantly higher than other algorithms, which means that the SPM could preserve the structure information effectively.

We choose four comparison methods for visualization: TFNet, FUMI, HiBCD, and CSTF. Fig. 5 shows the visualization results of the proposed method and the comparison method, where representative subimages of “BGU-0522-1113”

and “BGU-0522-1127” are chosen as examples. For reconstructed HR-HSIs, we select three bands of red (22nd band), green (14th band), and blue (7th band) and then concatenate them to generate the synthetic RGB images. In Fig. 5, the first and third rows list synthetic RGB images, and the second and fourth rows represent error maps between reconstructed HR-HSIs and ground truth. As can be seen from the reconstructed images, there is a color difference between the ground truth and the reconstructed images obtained by HiBCD, which indicates that HiBCD does not sufficiently capture the spectral information, thus leading to color distortion. What is more, from the error maps, we find that other comparative methods fail to adequately address the deviations in texture detail. The SCPNet method performs better both in structure

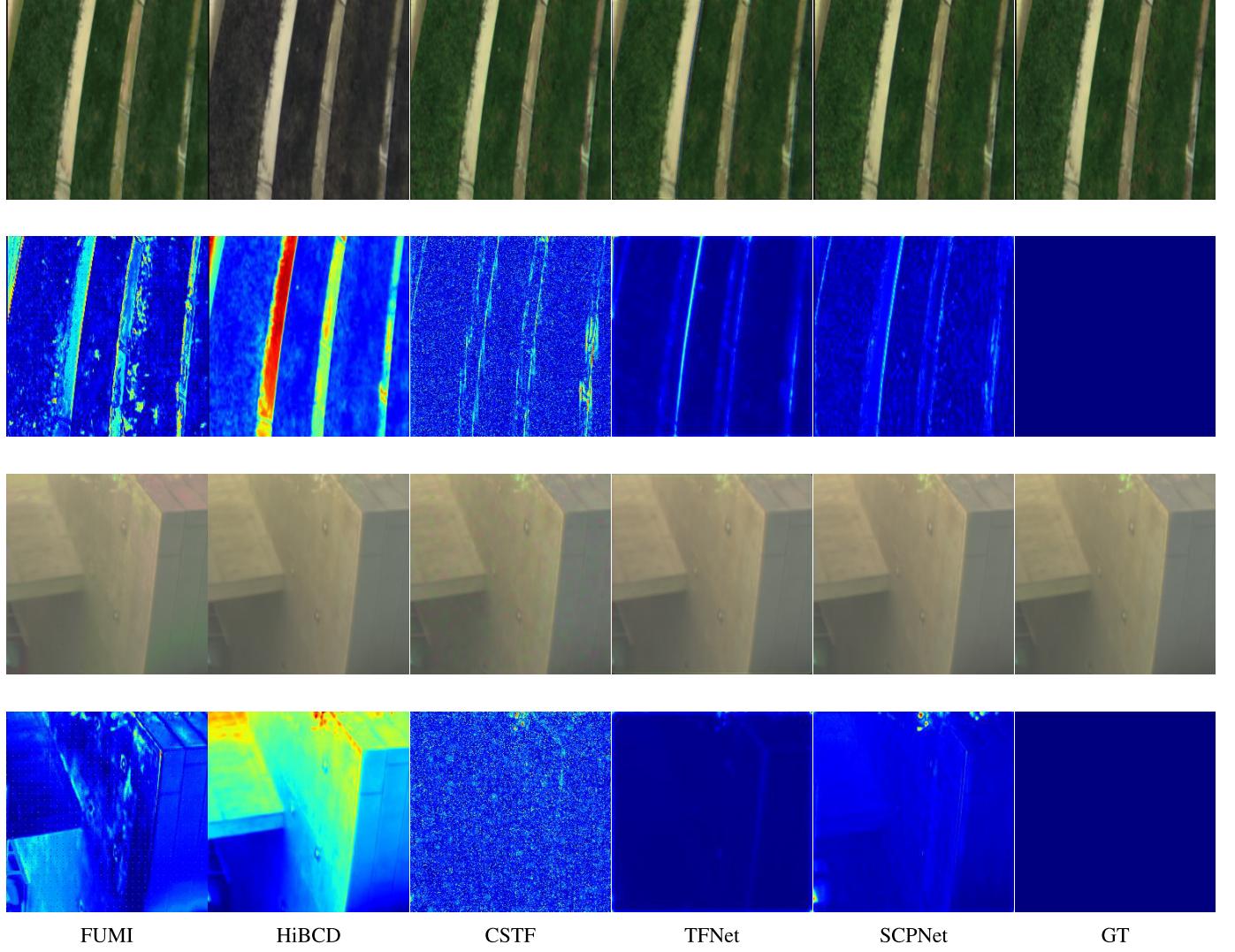


Fig. 5. Visual display of ICVL dataset when the upscaling factor is 8.

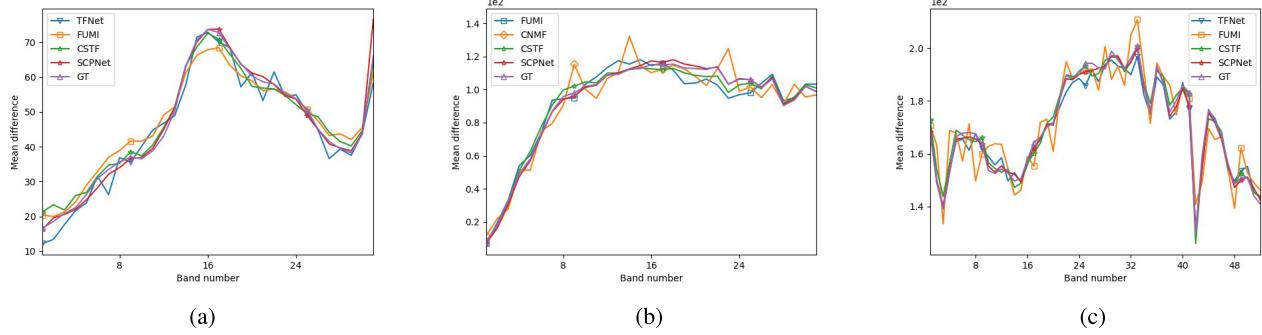


Fig. 6. Spectral curves of the selected pixel in the reconstructed HSI from three dataset. (a) ICVL. (b) DFC2018 Houston. (c) TG1HRSSC.

and color. To further validate the reconstruction ability of SCPNet, we randomly select one pixel from each reconstructed HR-HSIs and plot their spectral curves, which are shown in Fig. 6(a). From the 1st channel to the 31st channel, the pixel values of image reconstructed by SCPNet are always the closest to the real image. Moreover, an image obtained by the FUMI method is significantly different from the real one in the first 20 channels, while the overall difference of the

image obtained by the TFNet method fluctuates greatly from the first channel to the last channel. According to the curves shown in Fig. 6(a), our SCPNet gets the best reconstruction performance.

*2) Performance on the DFC2018 Houston Dataset:* For the DFC2018 Houston dataset, we randomly select 60 subimages for training and 20 for testing. More details and analysis on the DFC2018 Houston dataset are provided in Table II. For

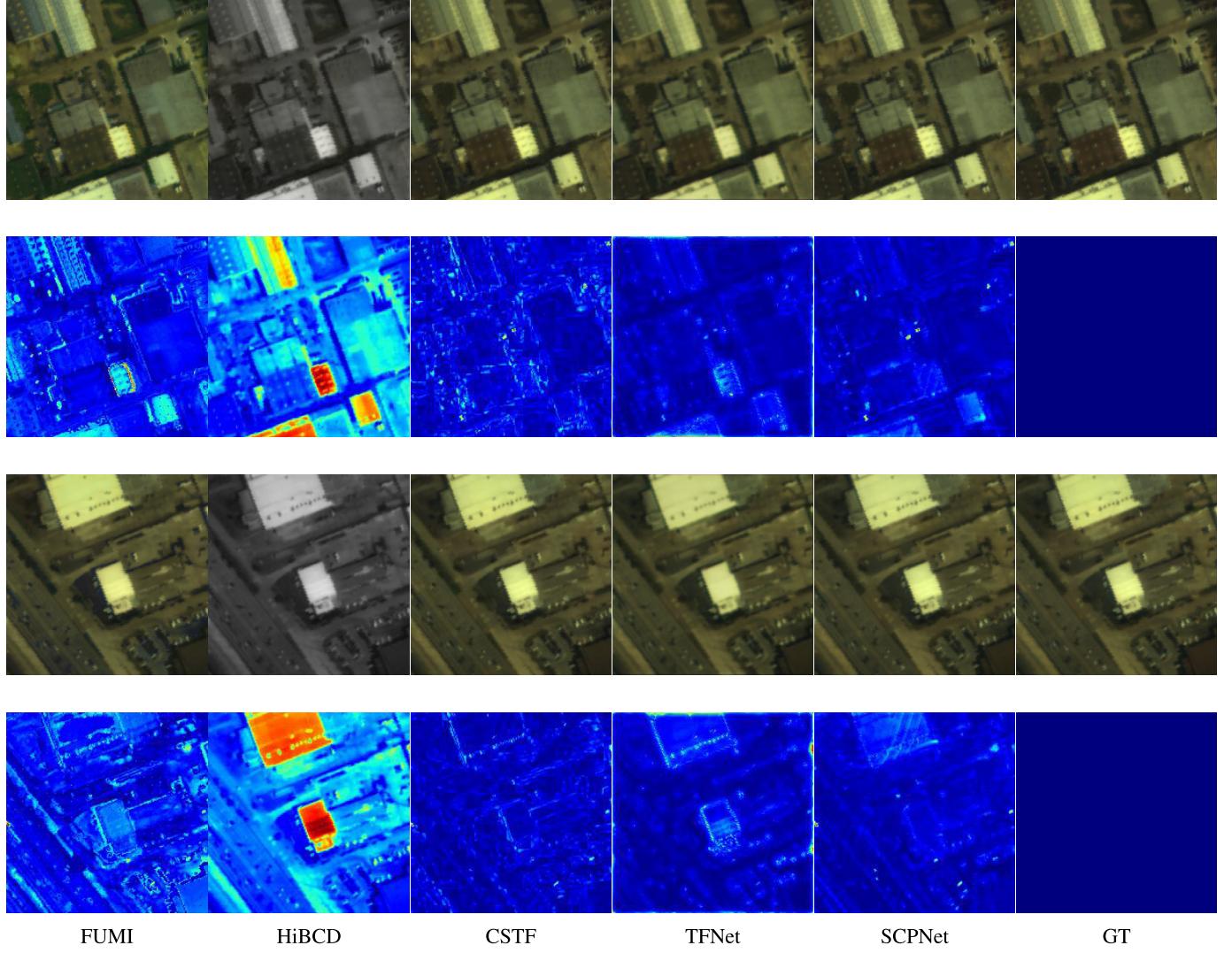


Fig. 7. Visual display of DFC2018 Houston dataset when the upscaling factor is 8.

TABLE IV  
ABLATION STUDY ON THE ICVL DATASET WHEN THE UPSCALING FACTOR IS 8

Options	Baseline	1st	2nd	3rd	4th	5th	6th	SCPNet
SPM		✓			✓		✓	✓
CPM			✓		✓	✓		✓
FS				✓		✓	✓	✓
PSNR	35.5075	38.9906	38.8906	39.1059	39.0339	39.0727	39.7619	<b>40.1055</b>
SSIM	0.9639	0.9853	0.9842	0.9860	0.9858	0.9865	0.9865	<b>0.9889</b>
SAM	3.6051	1.7869	1.8560	1.7303	1.7875	1.7257	1.7227	<b>1.6079</b>

the representative super-resolution algorithms, FUMI performs better than GSA and CNMF. CNMF provides better results than GSA. For the state-of-the-art super-resolution comparison algorithms, the performance of CSTF is the best in all cases. HiBCD and GLORIA perform poor in both  $\times 4$  and  $\times 8$  cases. In  $\times 2$  and  $\times 4$  cases, our SCPNet performs best for all the three assessments. Especially under the evaluations of SSIM and SAM, SCPNet remains a significant advantage

in reconstruction. Though the PSNR value of SCPNet in  $\times 8$  case is slightly smaller than CSTF, SSIM and SAM values still maintain our competitive superiority. Thus, we conclude that SCPNet could reconstruct more accurate HR-HSI than other comparison methods under quantitative assessment criteria.

Fig. 7 shows the visualization results of the proposed method and comparison methods. For reconstructed HR-HSIs,

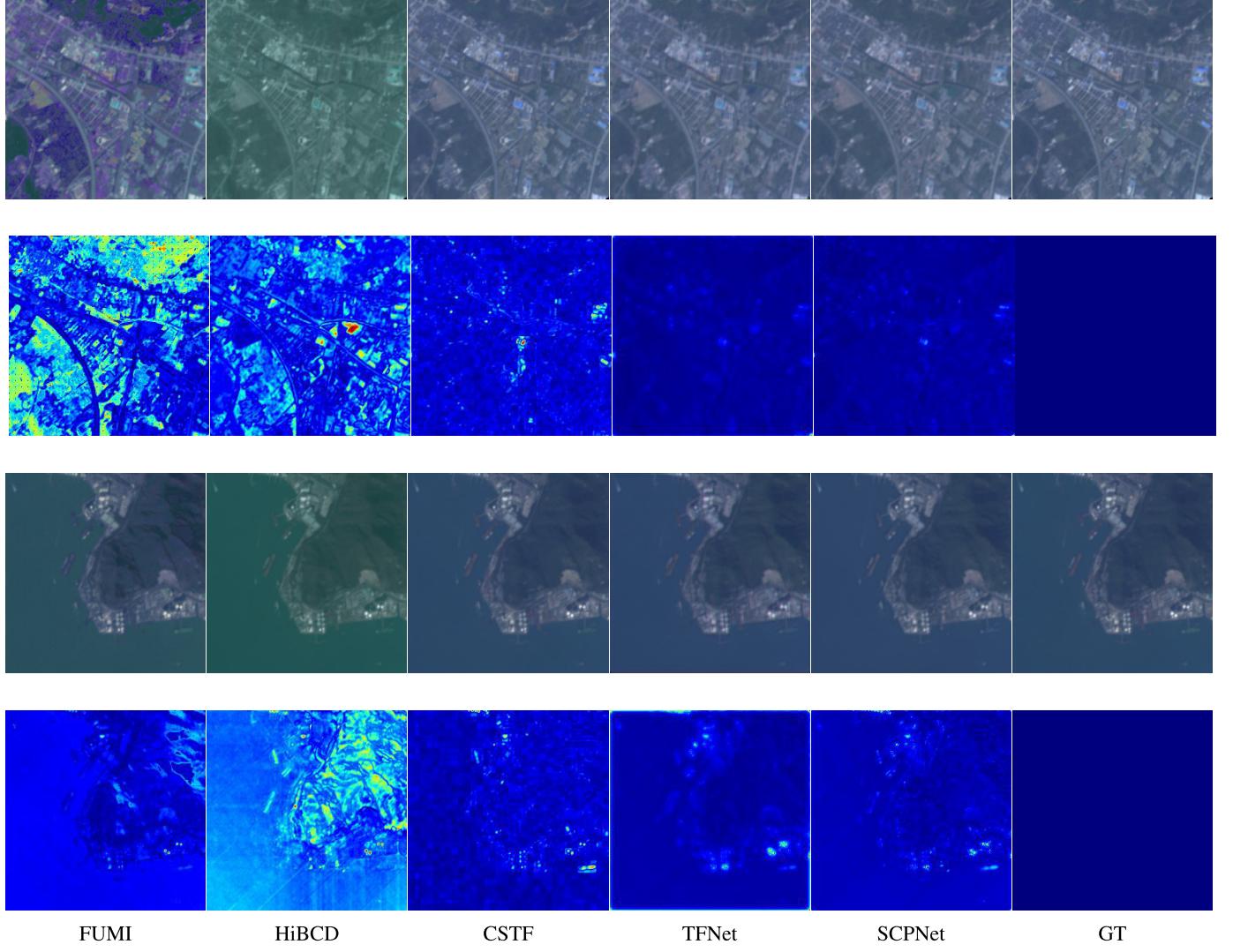


Fig. 8. Visual display of TG1HRSSC dataset when the upscaling factor is 8.

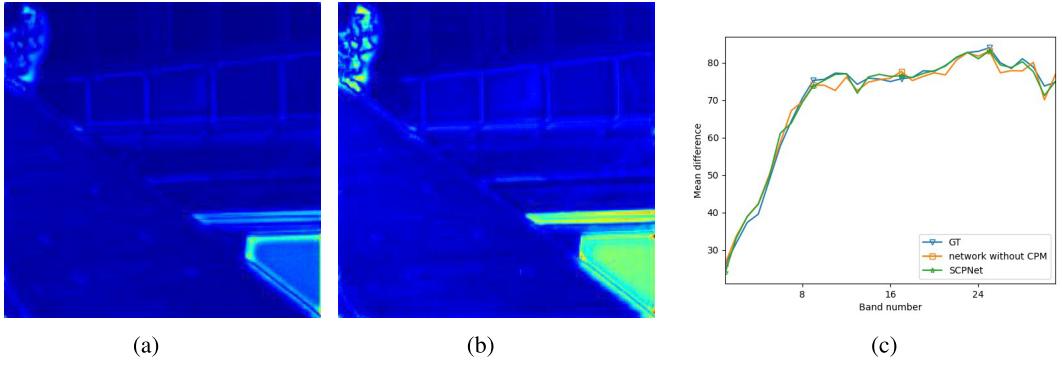


Fig. 9. Visual display of ablation experiment and SCPNet. (a) Reconstruction error map of SCPNet. (b) Reconstruction error map of network without SPM. (c) Spectral curves of SCPNet and network without CPM.

we select three bands of red (17th band), green (11th band), and blue (5th band) and then concatenate them to generate the synthetic RGB images. From Fig. 7, we can see that in terms of edge texture, two error maps from the proposed method are both darker than other algorithms, which demonstrates that

our network preserves structure characteristics best. Compared with HiBCD, our network captures spectral information and preserves color features, which can be represented in error maps too. Moreover, we also provide the reconstructed spectra of the DFC2018 Houston dataset in Fig. 6 (b) same as the

ICVL dataset. The image obtained by our SCPNet is closest to the real one from the first channel to the last. Also, the image obtained by CNMF fluctuates the most.

*3) Performance on the TG1HRSSC Dataset:* For experiments on the TG1HRSSC dataset, we select 40 images for training and 7 for testing. Table III shows the numerical results on TG1HRSSC dataset. Our network gets the best SSIM values in  $\times 2$ ,  $\times 4$ , and  $\times 8$  cases, which means that the SPM contributes to preserve color characteristics effectively. Moreover, when the upscaling factor is 4 or 8, spectral information of HR-HSI from SCPNet is maintained since the SAM values are the smallest. However, the performance of CSTF and FUMI sometimes outperforms ours on the PSNR measure. Thus, we need to validate the reconstruction performance through visualization results.

Fig. 8 shows the visualization results, where representative images “city-015-VNI-2013041514” and “port-001-VNI-2013010214” are chosen as examples. For reconstructed HR-HSIs, we select three bands of red (27th band), green (17th band), and blue (7th band) and then concatenate them to generate the synthetic RGB images. From synthetic RGB images and error maps, it is obvious that results from FUMI and HiBCD still remain distortions in structure and color. On the contrary, images from our network perform well both in terms of structure and color. Besides, spectral curves are shown in Fig. 6(c), which also shows the performance of SCPNet.

#### D. Ablation Study

To verify the contribution of the SPM, CPM, and fusion strategy, we conduct ablation experiments on the ICVL dataset. Table IV summarizes the numerical evaluation of comparative methods by three assessments. Let the model with convolution layers instead of SPM, CPM, and fusion strategy be baseline.

For network without SPM, the spatial information cannot be efficiently represented, while the network without CPM fails to capture the spectral details completely. Besides, since the spatial-spectral information cannot be well integrated, the networks without cross-fusion strategy perform worse than methods with it. Therefore, we conclude that the SPM, CPM, and fusion strategy all contribute to improve the performance.

Fig. 9 represents the visualization results of ablation experiments. Fig. 9(a) and (b) shows the reconstruction error maps of SCPNet and network without SPM. Obviously, SCPNet reconstructs HR-HSI with more accurate texture, which means that SPM contributes to the preservation of structure information. Fig. 9(c) shows the reconstructed spectral curves of SCPNet and network without CPM. The reconstructed image spectrum obtained by the full network is closer to the ground truth, and thus, it proves the function of CPM.

## IV. CONCLUSION

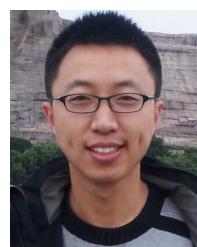
We propose a fusion-based HSR framework SCPNet with three modules based on a joint attention mechanism: an SPM based on spatial attention mechanism, a color-preserving module based on channel attention mechanism, and a cross-fusion strategy based on cross attention mechanism. SPM and CPM

capture spatial and spectral information, respectively, while the fusion strategy integrates information to reconstruct structure and color preserved HR-HSI. Comparison experiments demonstrate that our SCPNet outperforms all the other methods and the ablation study shows the contributions of all three modules. Thus, we conclude that the SCPNet is a promising algorithm not only in numerical assessments but also in visual effects. Although fusion-based HSR methods have achieved excellent performance, these methods do not consider the PSF, which may lead to poor reconstruction results when the upsampling factor is large. Therefore, in the future, we will incorporate PSF to improve the reconstruction performance.

## REFERENCES

- [1] P. Ghamisi *et al.*, “Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art,” *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.
- [2] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, “Hyperspectral remote sensing data analysis and future challenges,” *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [3] S. Mei, J. Ji, Y. Geng, Z. Zhang, X. Li, and Q. Du, “Unsupervised spatial-spectral feature learning by 3D convolutional autoencoder for hyperspectral classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6808–6820, Dec. 2019.
- [4] S. Mei, X. Li, X. Liu, H. Cai, and Q. Du, “Hyperspectral image classification using attention-based bidirectional long short-term memory network,” *IEEE Trans. Geosci. Remote Sens.*, early access, Aug. 11, 2021, doi: [10.1109/TGRS.2021.3102034](https://doi.org/10.1109/TGRS.2021.3102034).
- [5] S. Mei, X. Chen, Y. Zhang, J. Li, and A. Plaza, “Accelerating convolutional neural network-based hyperspectral image classification by step activation quantization,” *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5502012.
- [6] Z. Zou and Z. Shi, “Hierarchical suppression method for hyperspectral target detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 330–342, Jan. 2016.
- [7] Y. Zhang, B. Du, L. Zhang, and T. Liu, “Joint sparse representation and multitask learning for hyperspectral target detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 894–906, Feb. 2017.
- [8] W. Xie, J. Lei, Y. Cui, Y. Li, and Q. Du, “Hyperspectral pansharpening with deep priors,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1529–1543, May 2020.
- [9] C. Kwan, J. Choi, S. Chan, J. Zhou, and B. Budavari, “A super-resolution and fusion approach to enhancing hyperspectral images,” *Remote Sens.*, vol. 10, no. 9, p. 1416, Sep. 2018.
- [10] S. Gou, S. Liu, S. Yang, and L. Jiao, “Remote sensing image super-resolution reconstruction based on nonlocal pairwise dictionaries and double regularization,” *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 12, pp. 4784–4792, Dec. 2014.
- [11] L. Gao, D. Hong, J. Yao, B. Zhang, P. Gamba, and J. Chanussot, “Spectral superresolution of multispectral imagery with joint sparse and low-rank learning,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2269–2280, Mar. 2021.
- [12] S. Mei, Y. Geng, J. Hou, and Q. Du, “Learning hyperspectral images from RGB images via a coarse-to-fine CNN,” *Sci. China Inf. Sci.*, vol. 65, no. 5, May 2022, Art. no. 152102.
- [13] Y. Wang, X. Chen, Z. Han, and S. He, “Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization,” *Remote Sens.*, vol. 9, no. 12, p. 1286, 2017.
- [14] S. Tang, Y. Xu, L. Huang, and L. Sun, “Hyperspectral image super-resolution via adaptive dictionary learning and double  $\ell_1$  constraint,” *Remote Sens.*, vol. 11, no. 23, p. 2809, Nov. 2019.
- [15] K. Jiang, Z. Wang, P. Yi, and J. Jiang, “Hierarchical dense recursive network for image super-resolution,” *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107475.
- [16] E. Liu, Z. Tang, B. Pan, and Z. Shi, “Spatial-spectral feedback network for super-resolution of hyperspectral imagery,” 2021, *arXiv:2103.04354*.
- [17] D. Liu, J. Li, and Q. Yuan, “A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution,” *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7711–7725, Sep. 2021.

- [18] K. Jiang, Z. Wang, P. Yi, G. Wang, T. Lu, and J. Jiang, "Edge-enhanced GAN for remote sensing image superresolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5799–5812, Jun. 2019.
- [19] X. Wang, J. Ma, and J. Jiang, "Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5503113.
- [20] Y. Mei, Y. Fan, Y. Zhou, L. Huang, T. S. Huang, and H. Shi, "Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5689–5698.
- [21] N. Yokoya, C. Grohnfeldt, 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.
- [22] L. Loncan *et al.*, "Hyperspectral pansharpening: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [23] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled non-negative matrix factorization (CNMF) for hyperspectral and multispectral data fusion: Application to pasture classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2011, pp. 1779–1782.
- [24] 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.*, May 2013, pp. 1409–1413.
- [25] Q. Wei, J. Bioucas-Dias, N. Dobigeon, J.-Y. Tourneret, M. Chen, and S. Godsill, "Multiband image fusion based on spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7236–7249, Dec. 2016.
- [26] R. Dian, S. Li, L. Fang, and Q. Wei, "Multispectral and hyperspectral image fusion with spatial-spectral sparse representation," *Inf. Fusion*, vol. 49, pp. 262–270, Sep. 2019.
- [27] J. Xue, Y.-Q. Zhao, Y. Bu, W. Liao, J. C.-W. Chan, and W. Philips, "Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 3084–3097, 2021.
- [28] R. Dian, L. Fang, and S. Li, "Hyperspectral image super-resolution via non-local sparse tensor factorization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3862–3871.
- [29] C. I. Kanatsoulis, X. Fu, N. D. Sidiropoulos, and W.-K. Ma, "Hyperspectral super-resolution: A coupled tensor factorization approach," *IEEE Trans. Signal Process.*, vol. 66, no. 24, pp. 6503–6517, Dec. 2018.
- [30] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.
- [31] 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.
- [32] K. Wang, Y. Wang, X.-L. Zhao, J. C.-W. Chan, Z. Xu, and D. Meng, "Hyperspectral and multispectral image fusion via nonlocal low-rank tensor decomposition and spectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 11, pp. 7654–7671, Nov. 2020.
- [33] Y. Zhang, A. Duijster, and P. Scheunders, "A Bayesian restoration approach for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3453–3462, Sep. 2012.
- [34] N. Akhtar, F. Shafait, and A. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3631–3640.
- [35] Y. Chang, L. Yan, X. Zhao, H. Fang, Z. Zhang, and S. Zhong, "Weighted low-rank tensor recovery for hyperspectral image restoration," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4558–4572, Nov. 2020.
- [36] 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.
- [37] W. Wei, J. Nie, Y. Li, L. Zhang, and Y. Zhang, "Deep recursive network for hyperspectral image super-resolution," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1233–1244, 2020.
- [38] J. Li *et al.*, "Hyperspectral image super-resolution by band attention through adversarial learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4304–4318, Jun. 2020.
- [39] J. Hu, X. Jia, Y. Li, G. He, and M. Zhao, "Hyperspectral image super-resolution via intrafusion network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7459–7471, Oct. 2020.
- [40] W. Xie, Y. Cui, Y. Li, J. Lei, Q. Du, and J. Li, "HPGAN: Hyperspectral pansharpening using 3-D generative adversarial networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 463–477, Jan. 2021.
- [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, p. 1139, 2017.
- [42] Q. Li, Q. Wang, and X. Li, "Exploring the relationship between 2D/3D convolution for hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 10, pp. 8693–8703, Oct. 2021.
- [43] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-D-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, May 2017.
- [44] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, Nov. 2017.
- [45] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sens.*, vol. 10, no. 5, p. 800, 2018.
- [46] S. Mei, R. Jiang, X. Li, and Q. Du, "Spatial and spectral joint super-resolution using convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 4590–4603, Jul. 2020.
- [47] J.-F. Hu, T.-Z. Huang, L.-J. Deng, T.-X. Jiang, G. Vivone, and J. Chanussot, "Hyperspectral image super-resolution via deep spatirospectral attention convolutional neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jun. 9, 2021, doi: [10.1109/TNNLS.2021.3084682](https://doi.org/10.1109/TNNLS.2021.3084682).
- [48] X.-H. Han, Y. Zheng, and Y.-W. Chen, "Multi-level and multi-scale spatial and spectral fusion CNN for hyperspectral image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 4330–4339.
- [49] Y. Qu, H. Qi, C. Kwan, N. Yokoya, and J. Chanussot, "Unsupervised and unregistered hyperspectral image super-resolution with mutual Dirichlet-net," *IEEE Trans. Geosci. Remote Sens.*, early access, May 27, 2021, doi: [10.1109/TGRS.2021.3079518](https://doi.org/10.1109/TGRS.2021.3079518).
- [50] J. Yao, D. Hong, J. Chanussot, D. Meng, X. Zhu, and Z. Xu, "Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution," in *Proc. ECCV*, 2020, pp. 208–224.
- [51] B. Aiazzi, S. Baronti, and M. Selva, "Improving component substitution pansharpening through multivariate regression of MS + pan data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3230–3239, Sep. 2007.
- [52] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.
- [53] R. Wu, W.-K. Ma, X. Fu, and Q. Li, "Hyperspectral super-resolution via global-local low-rank matrix estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 7125–7140, Oct. 2020.
- [54] Y. Zheng, J. Li, Y. Li, K. Cao, and K. Wang, "Deep residual learning for boosting the accuracy of hyperspectral pansharpening," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1435–1439, Aug. 2020.
- [55] R. Wu, C.-H. Chan, H.-T. Wai, W.-K. Ma, and X. Fu, "Hi, BCD! hybrid inexact block coordinate descent for hyperspectral super-resolution," in *Proc. ICASSP*, Apr. 2018, pp. 2426–2430.



**Bin Pan** (Member, IEEE) received the B.S. and Ph.D. degrees from the School of Astronautics, Beihang University, Beijing, China, in 2013 and 2019, respectively.

Since 2019, he has been an Associate Professor with the School of Statistics and Data Science, Nankai University, Tianjin, China. His research interests include machine learning, remote sensing image processing, and multiobjective optimization.



**Qiaoying Qu** received the B.S. degree from the College of Mathematics, Jilin University, Changchun, China, in 2021. She is currently pursuing the M.S. degree with the School of Statistics and Data Science, Nankai University, Tianjin, China.

Her research interests include machine learning, hyperspectral super-resolution, and remote sensing image processing.



**Xia Xu** (Graduate Student Member, IEEE) received the B.S. and M.S. degrees from the School of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2012 and 2015, respectively, and the Ph.D. degree from the School of Astronautics, Beihang University, Beijing, China, in 2019.

She is currently an Assistant Professor with the College of Computer Science, Nankai University, Tianjin, China. Her research interests include hyperspectral unmixing, multiobjective optimization, and remote sensing image processing.



**Zhenwei Shi** (Member, IEEE) received the Ph.D. degree in mathematics from the Dalian University of Technology, Dalian, China, in 2005.

From 2005 to 2007, he was a Post-Doctoral Researcher with the Department of Automation, Tsinghua University, Beijing, China. From 2013 to 2014, he was a Visiting Scholar with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL, USA. He is currently a Professor and the Dean of the Image Processing Center, School of Astronautics, Beihang University, Beijing. He has authored or coauthored more than 100 scientific papers in related journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON NEURAL NETWORKS, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, and IEEE Conference on Computer Vision and Pattern Recognition. His research interests include remote sensing image processing and analysis, computer vision, pattern recognition, and machine learning.

Dr. Shi received the Best Reviewer Awards for his service to IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING in 2017. He has been an Associate Editor of *Infrared Physics and Technology* since 2016.