Structure-Color Preserving Network for Hyperspectral Image Super Resolution

Bin Pan, Qiaoying Qu, Xia Xu and Zhenwei Shi

Abstract

Fusion based hyperspectral super resolution (HSR) algorithms usually utilize a low-resolution hyperspectral image and a high-resolution multispectral image to generate a high-resolution hyperspectral image, 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 multispectral images has always been the focus and difficulty of fusion based hyperspectral super resolution. In this paper, we propose a new structure-color preserving network (SCPNet) for hyperspectral super resolution, which is developed under the basis of the joint attention mechanism. The SCPNet mainly includes three modules: structure-preserving module, color-preserving module and cross fusion module. The structure-preserving module is constructed based on the spatial attention, which aims to capture and enhance the significant structure information from the high-resolution multispectral image. Meanwhile, the color-preserving module is constructed based on the channel attention, where the spectral characteristics in the low-resolution hyperspectral image are preserved during the reconstruction process. At last, we propose a cross attention based cross fusion strategy to integrate the features from the two branches, and reconstruct the final high-resolution hyperspectral image. The major contribution of SCPNet is that the structure and color information is respectively 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

Hyperspectral super resolution, structure-color preserving, attention mechanism.

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 classification [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

This work was supported by the National Key R&D Program of China under the Grant 2019YFC1510905, the National Natural Science Foundation of China under the Grant 62001251 and 62001252 and the China Postdoctoral Science Foundation under the Grant 2020M670631. *(Corresponding author: Xia Xu.)*

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

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

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

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/mutispectral 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, it is a more effective method to reconstruct HR-HSI by super resolving LR-HSI. According to 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 low-resolution hyperspectral image (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 [19–21]. Liu *et al.* employ group convolutions and covariance statistics based attention mechanism to explore the consecutive information. Mei *et al.* combine the cross-scale non-local prior with local and in-scale non-local 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 a HR-HSI using an LR-HSI and a registered multispectral image (MSI), which has attracted much attention in recent years [22],[23]. 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 higher spatial resolution while the LR-HSIs have finer spectral resolution. Traditional methods such as matrix factorization based algorithms [24–28] are proposed firstly. 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 [29–33]. In addition to the above two algorithms, Bayesian rule based algorithms have been widely applied in HSR. The dictionary is obtained through Bayesian dictionary learning, then the HR-HSI is reconstructed with dictionary and sparse coding matrix [34–37].

Recently, deep learning techniques, especially the convolutional neural networks, have presented promising performance in HSR tasks[38–41]. For example, in order to adapt to the characteristics of hyperspectral images as 3-D cube, some methods apply 3-D convolution to CNN [42–44]. Mei *et al.* [42] proposed a 3-D CNN based algorithm, which showed that 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],[45]. Xie *et al.* [8] proposed a network

which could capture the deep residual features of high-frequency information, and utilize the learned features as a priori in HR-HSI reconstruction. For fusion based hyperspectral super resolution task, the cross fusion of spatial information and spectral information is critical for the reconstruction performance. Therefore, researchers mostly focus on developing networks which could extract significant characteristics and integrate them [46–49]. Han *et al.* [49] utilized a multi-level network to upscale the spatial resolution of LR-HSI gradually and integrated the multi-scale loss functions during the training to avoid the gradient vanish. Moreover, some alternative super resolution methods have been proposed. Such as unsupervised HSR[50, 51] and HSR algorithms considering PSF[9]. Qu *et al.* [50] adopt the mutual information and assume that the characteristics follow a similar Dirichlet distribution. Kwan *et al.* [9] super resolve LR-HSI with method incorporates PSF into the deblurring and then fuse 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 which simultaneously considers the color and structure characteristics remains a challenge.

In this paper, 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.

- We propose a new spatial-attention-based structure preserving module to extract the structure details from MSIs.
- We propose a new channel-attention-based color preserving module to provide spectral invariance from LR-HSIs.
- 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) Structure Preserving Module; 2) Color Preserving Module; and 3) Cross Fusion Strategy.

A. Network Architecture

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

$$X = ZS + N_1 \tag{1.1}$$

$$Y = DZ + N_2 \tag{1.2}$$

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 superresolve 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 same parameters, which reduces the number of parameters and improves the training speed.

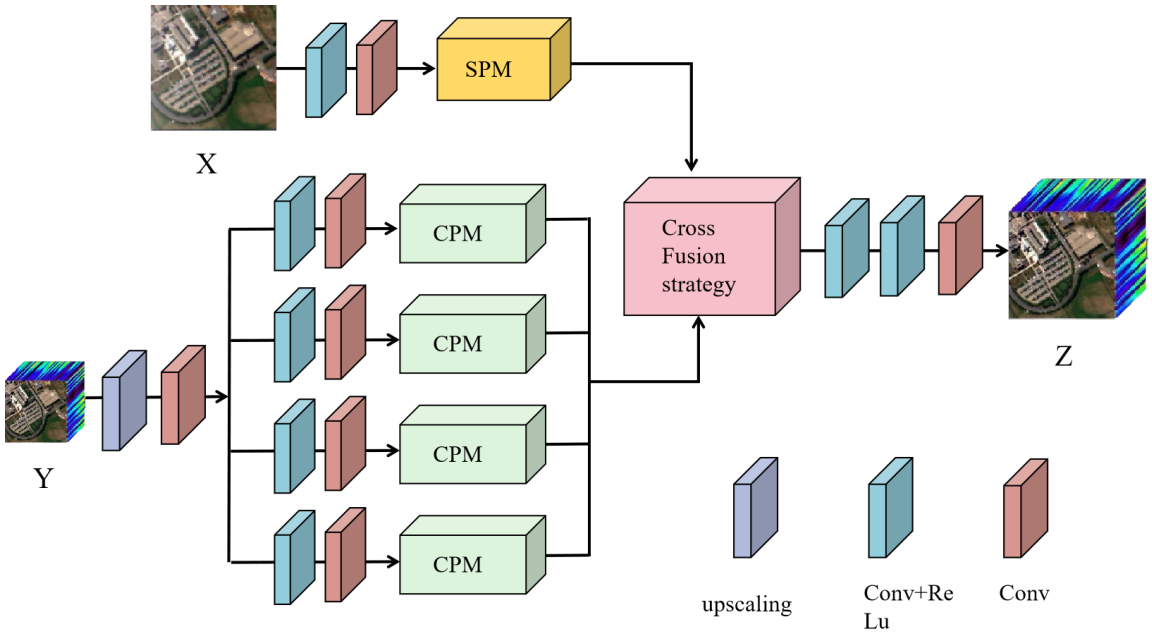


Fig. 1: Illustration of proposed SCPNet framework. This network contains three parts: Structure Preserving Module (SPM), Color Preserving Module (CPM) and cross fusion strategy. SPM and CPM are proposed for feature extraction of HR-MSI and LR-HSI while the cross fusion strategy is proposed for the fusion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

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

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

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

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

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

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

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

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

Spatial information determines the texture and details while spectral information determines color of a image. So through capturing and ehancing spatial-spectral information, SCPNet could obtain image with more accurate structure and color.

Most fusion based HSR methods realize feature fusion by concatenating feature maps. Inspired by Yao [51], we adopt a newly proposed cross fusion strategy based on 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:

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

where $CF(\cdot)$ is the cross fusion strategy.

B. Spatial Attention based Structure Preserving Module

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

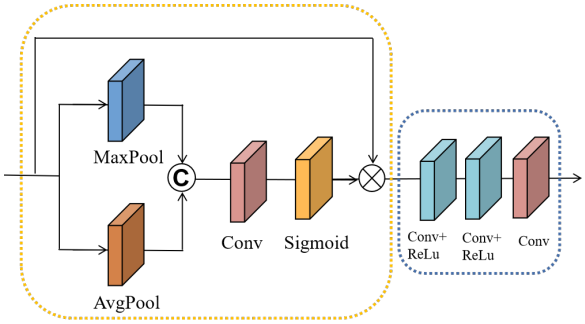


Fig. 2: Illustration of SPM, which consists of a spatial attention block and a feature extraction block. The spatial attention block is in the yellow box and the feature preserving block is in the blue box.

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 $SAB(\cdot)$ denote the spatial attention block. The SPM can be expressed as:

$$X' = Conv(ReLU(Conv(ReLU(Conv(SAB(F_2)))))) \tag{2.1}$$

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

$$Ms(F) = \sigma(f^{7\times7}([AvgPool(F), MaxPool(F)])) \tag{2.3}$$

$AvgPool(\cdot)$ represents the average pooling operation while $MaxPool(\cdot)$ represents the maximum pooling operation. $f^{7\times7}(\cdot)$ denotes the 7×7 convolution layer. And $\sigma(\cdot)$ represents the $sigmoid$ activation fuction.

As the pixel values of a 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.

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 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, 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 color correction branch to obtain the result. The structure of CPM is shown in Fig 3.

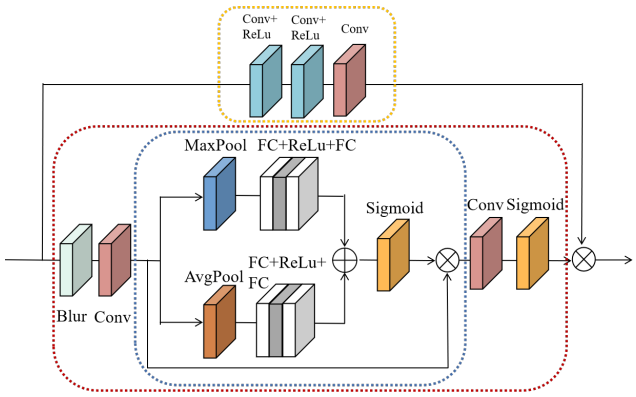


Fig. 3: Illustration of CPM, which contains a color correction branch and a feature preserving branch. The color correction module is in the red box, the channel attention block is in the blue box and the feature preserving block is in the yellow box.

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

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

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

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

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

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

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

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

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

Since the spectral coherence of hyperspectral images is critical for 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: The pseudocode for SCPNet.

Input: HR MSI $\mathbf{X}$, LR HSI $\mathbf{Y}$.
Output: reconstructed HR HSI $\mathbf{Z}$.

while the SCPNet has not converged yet **do**
For $\mathbf{X}$:
Shallow feature extraction: $F_1 = Conv(ReLU(Conv(X)))$
Input the shallow features into the SPM: $X' = SPM(F_1)$
For $\mathbf{Y}$:
Upsampling the HR-HSI: $Y_1 = Upsampling(Y)$
Divide the upscaled image into four groups: $[Y_1, Y_2, Y_3, Y_4] = Conv(Y_1)$
Shallow feature extraction: $F_{1i} = Conv(ReLU(Conv(Y_i))) \quad i = 1, 2, 3, 4$
Input the shallow features into the CPM: $F_{2i} = CPM(F_{1i})$
Feature fusion:
$Z = FC(X', Y')$
Renturn Z

D. Cross Attention based Cross Fusion Strategy

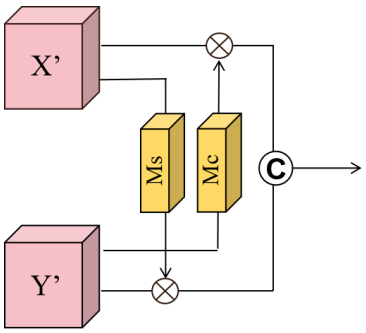


Fig. 4: Illustration of cross fusion strategy. Ms module is utilized to compute the spatial weight coefficients, and the structure is same as spatial attention block. Meanwhile, the Mc module is adopted to compute the spectral weight, and the structure is same as channel attention block.

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') respectively. 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:

$$Ms(X') = \sigma(f^{7\times7}([AvgPool(X'), MaxPool(X')])) \tag{4.1}$$

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

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

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

$$Z = Conv(Relu(Conv(Relu(Conv(X'', Y''))))) \tag{4.5}$$

E. Loss Function

To obtain the optimal network parameter set, the losses between training samples and ground truth need to be minimized. The L2 loss is generally utilized in most methods to maximize the PSNR. However, the L2 loss often fails to capture the underlying multimodal distributions of the HR patches, which results in the over-smoothness of

TABLE I: The PSNR, SSIM and SAM values by different methods on ICVL dataset.

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

TABLE II: The PSNR, SSIM and SAM values by different methods on DFC2018 Houston dataset.

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

TABLE III: The PSNR, SSIM and SAM values by different methods on TG1HRSSC dataset.

Scale		GSA	CNMF	FUMI	HiBCD	CSTF	GLORIA	TFNet	SCPNet
×2	PSNR	37.1437	42.4636	43.1266	41.4067	**46.7022**	40.9173	35.6435	43.5513
	SSIM	0.8679	0.9370	0.9733	0.9492	0.9708	0.9505	0.9192	**0.9754**
	SAM	6.2052	4.0938	**2.3896**	6.3981	2.5533	4.9034	5.0556	3.4060
×4	PSNR	36.3512	41.1210	45.4276	37.2205	**45.9850**	33.0627	36.4373	42.7133
	SSIM	0.8706	0.9250	0.9738	0.9133	0.9690	0.8227	0.9313	**0.9829**
	SAM	6.9451	4.3561	2.7147	11.9277	2.8272	15.5473	4.6637	**2.5809**
×8	PSNR	33.8293	39.0538	**42.7513**	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	**0.9800**
	SAM	11.0983	5.1371	4.2480	17.9753	4.5432	23.8257	5.1649	**2.8458**

the reconstructed images. Therefore, We select the L1 loss since it provides better convergence. The loss function is shown in Eq. (5.1).

$$loss = \sum_{i=1}^{N} \left| I(i) - \widehat{I}(i) \right| \tag{5.1}$$

More training details are provided in Section III.

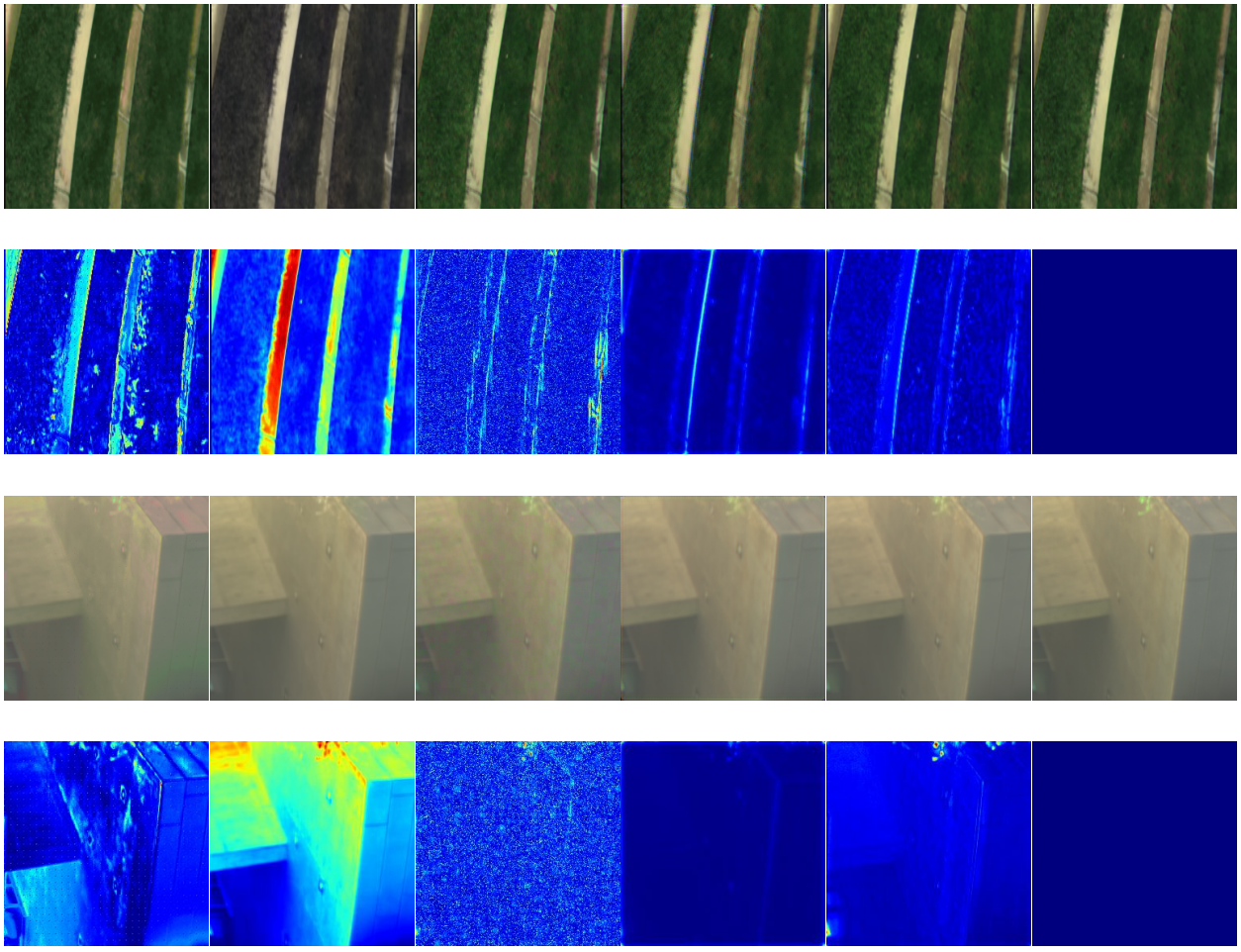


Fig. 5: Visual display of ICVL dataset when upscaling factor is 8.

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 [52], CNMF [24], FUMI [26], CSTF [53], GLORIA [54], TFNet [55] and HiBCD [56]. 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 based and deep learning based HSR methods respectively.

We utilize peak signal to noise ratio (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 data set contains 201 images. These hyperspectral images cover a 400-700 nm spectral range with 31 bands. Subimages with size of $31\times256\times256$ 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-bands HR-MSI with size of 256×256 is constructed by filtering the HR-HSI with a spectral responses function [27].

2) DFC2018 Houston: The data was 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 $50\times200\times200$ 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 data set acquired by Tiangong-1 hyperspectral imager. We select visible near-infrared spectral data which covers a 400-900 nm with 54 effective bands for the experiment. Subimages with size of $54\times256\times256$ 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 comparision experiments on three hyperspectral datasets. And we choose three kinds of upscaling factors for experiments: 2, 4 and 8. The details and analysis of experiments can be seen as follows.

Performance on ICVL Dataset. We randomly select 60 subimages from ICVL dataset for training and 20 for testing. Then we construct corresponding LR-HSIs and HR-MSIs with Guassian 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 image and real image 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, 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 sum-to-one and nonnegativity constraints help improve the performance. For the state-of-the-art super resolution comparative

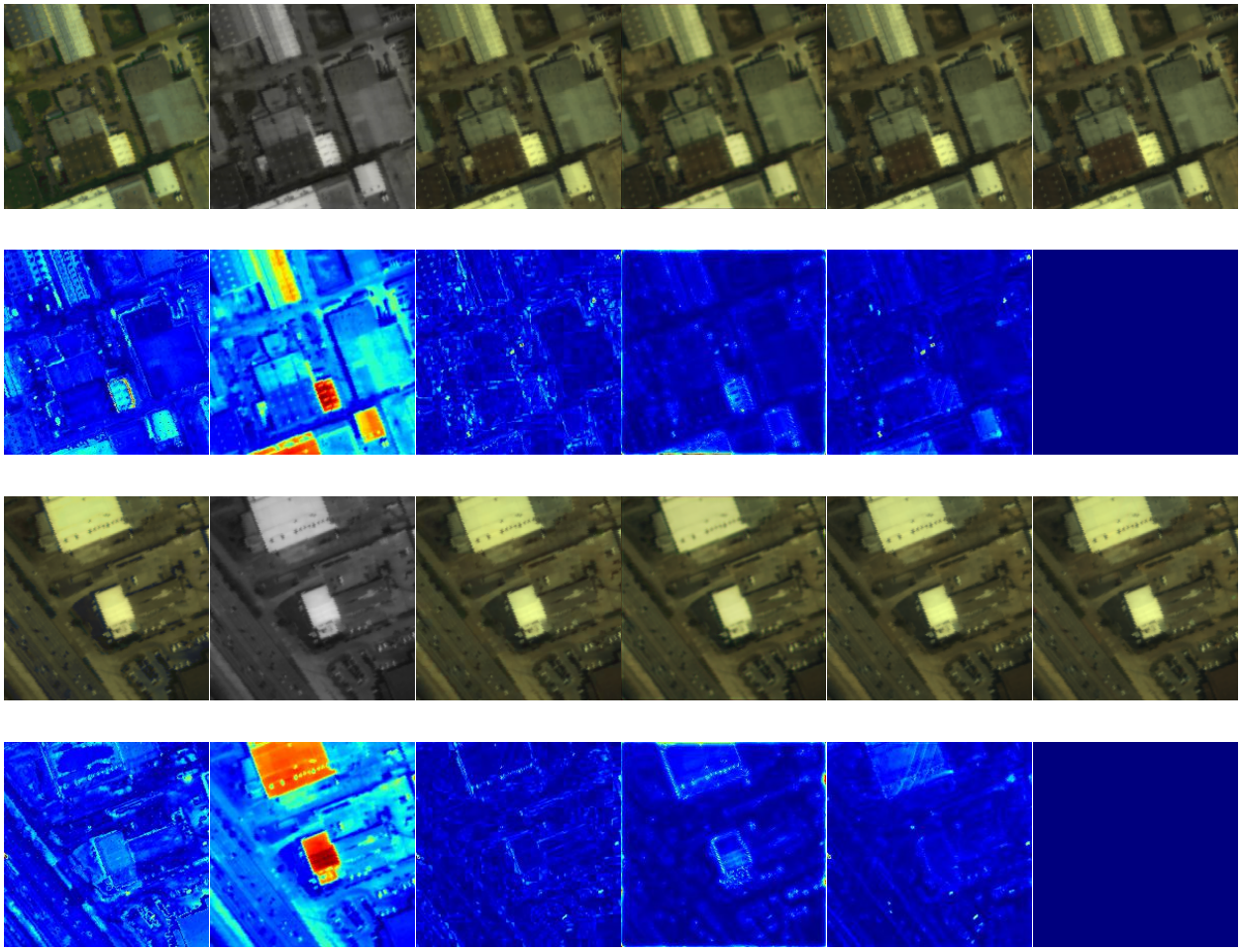


Fig. 6: Visual display of DFC2018 Houston dataset when upscaling factor is 8.

algorithms, performance of TFNet is the worst in $\times 2$ case. HiBCD performs poor in $\times 4$ case. And 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 $\times 2$ case. Compared with both deep learning based and optimization based comparison methods, SCPNet has the best performance in all 3 cases. Since the SSIM and SAM of 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 8 times, the value of SSIM is significantly higher than other algorithms, which means 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

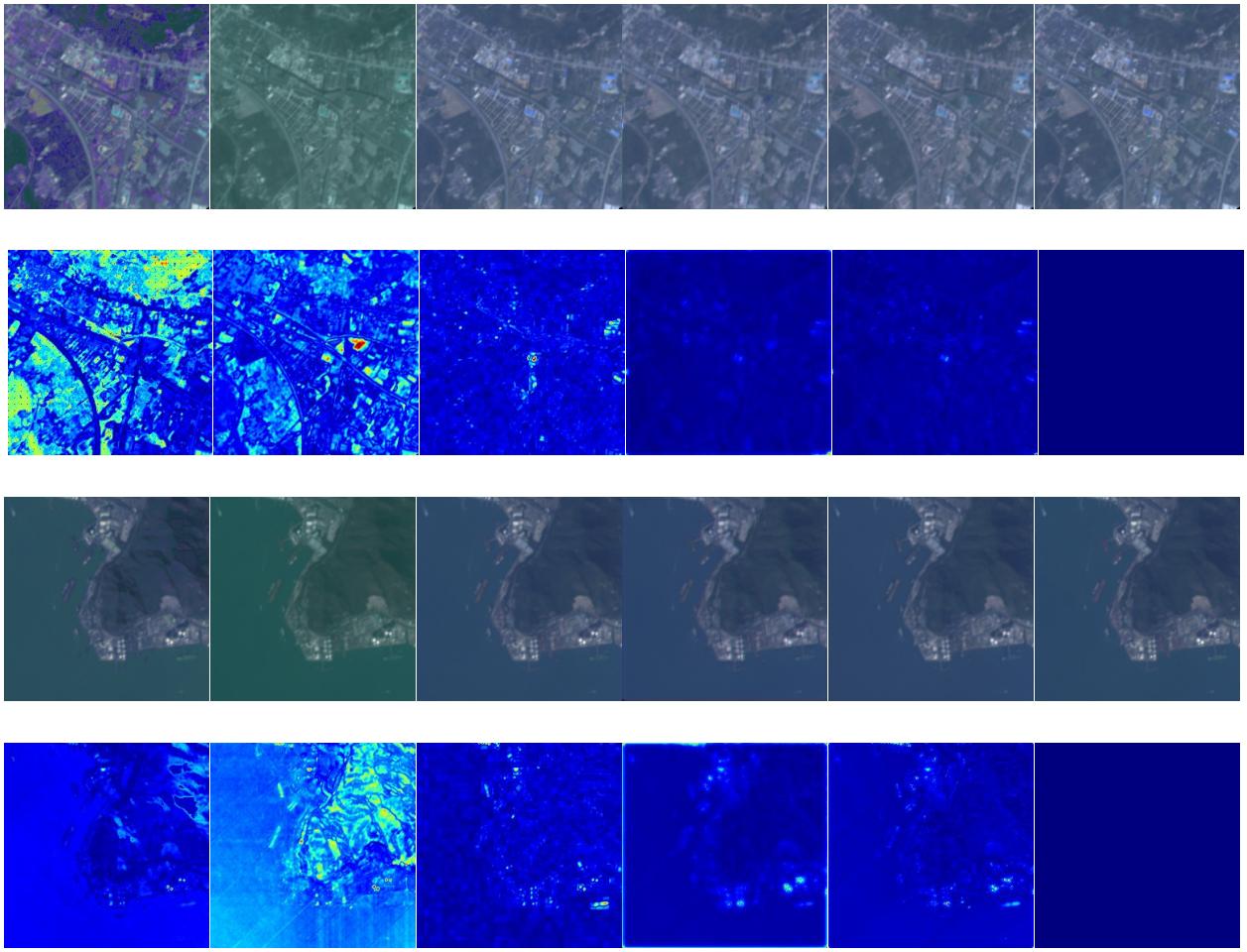
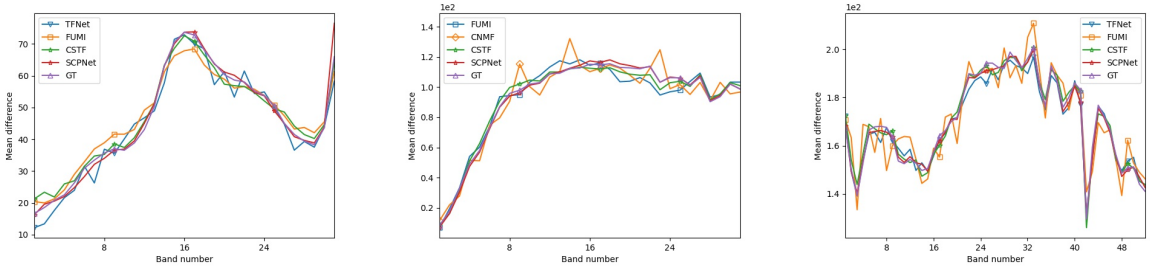


Fig. 7: Visual display of TG1HRSSC dataset when upscaling factor is 8.



(a) ICVL (b) DFC2018 Houston (c) TG1HRSSC

Fig. 8: Spectral curves of the selected pixel in the reconstructed HSI from three dataset.

red (22th 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 color difference between 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's more, from the error maps we find that other comparative methods fail to adequately address the deviations in texture detail. SCPNet method performs better both in structure 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 demonstrated in Fig. 8 (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, image obtained by FUMI method is significantly different from the real one in the first 20 channels, while the overall difference of the image obtained by TFNet method fluctuates greatly from the first channel to the last channal. According to the curves shown in Fig. 8 (a), our SCPNet gets the best reconstruction performance.

Performance on DFC2018 Houston Dataset. For the DFC2018 Houston dataset, we randomly select 60 subimages for training and 20 for testing. More details and analysis on DFC2018 Houston dataset are provided in Table II. For the representative super resolution algorithms, FUMI performs better than GSA and CNMF. CNMF provides with better results than GSA. For the state-of-the-art super resolution comparison algorithms, 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 superioity. So we conclude that SCPNet could reconstruct more accurate HR-HSI than other comparison methods under quantitative assessment criteria.

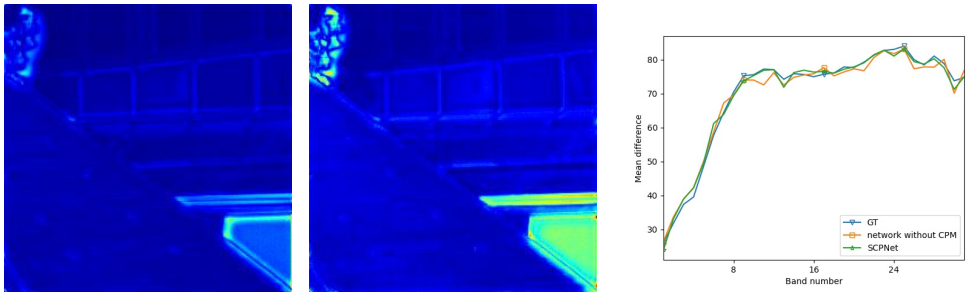
Fig. 6 shows the visualization results of the proposed method and comparison methods. For reconstructed HR-HSIs, we select three bands of red (17th band), green (11th band) and blue (5th band), and then concatenate them to generate the synthetic RGB images. From Fig. 6, we can see that in terms of edge 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 with the reconstructed spectra of DFC2018 Houston dataset in Fig. 8 (b) same as ICVL dataset. Image obtained by our SCPNet is closest to the real one from the first channel to the last. And image obtained by CNMF fluctuates the most.

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

Fig. 7 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 depicted in Fig. 8 (c), which also shows the performance of SCPNet.

D. Ablation Study



(a) reconstruction error map of SCPNet (b) reconstruction error map of network without SPM (c) spectral curves of SCPNet and network without CPM

Fig. 9: visual display of ablation experiment and SCPNet

To verify the contribution of the SPM, CPM and fusion strategy, we conduct ablation experiments on 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 can not 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 performace.

Fig. 9 represents the visualization results of ablation experiments. Fig. 9 (a) and Fig.9 (b) are reconstruction error maps of SCPNet and network without SPM. Obviously, SCPNet reconstructs HR-HSI with more accurate texture, which means 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, thus it proves the function of CPM.

TABLE IV: Ablation Study on 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	**40.1055**
SSIM	0.9639	0.9853	0.9842	0.9860	0.9858	0.9865	0.9865	**0.9889**
SAM	3.6051	1.7869	1.8560	1.7303	1.7875	1.7257	1.7227	**1.6079**

IV. Conclusion

We propose a fusion based HSR framework SCPNet with three modules based on joint attention mechanism: a structure preserving module 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 both 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. So we conclude that the SCPNet is a promising algorithm not only in numerical assessments but also in visual effects. Although fusion based hyperspectral super resolution methods have achieved excellent performance, these methods do not consider the point spread function (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] Pedram Ghamisi, Naoto Yokoya, Jun Li, Wenzhi Liao, Sicong Liu, Javier Plaza, Behnood Rasti, and Antonio Plaza. Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 5(4):37–78, 2017.

[2] Jose M. Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser Nasrabadi, and Jocelyn Chanussot. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and Remote Sensing Magazine*, 1(2):6–36, 2013.

[3] Shaohui Mei, Jingyu Ji, Yunhao Geng, Zhi Zhang, Xu Li, and Qian Du. Unsupervised spatialspectral feature learning by 3d convolutional autoencoder for hyperspectral classification. *IEEE Transactions on Geoscience and Remote Sensing*, 57(9):6808–6820, 2019.

[4] Shaohui Mei, Xingang Li, Xiao Liu, Huimin Cai, and Qian Du. Hyperspectral image classification using attention-based bidirectional long short-term memory network. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–12, 2021.

[5] Shaohui Mei, Xiaofeng Chen, Yifan Zhang, Jun Li, and Antonio Plaza. Accelerating convolutional neural network-based hyperspectral image classification by step activation quantization. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–12, 2021.

[6] Zhengxia Zou and Zhenwei Shi. Hierarchical suppression method for hyperspectral target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 54(1):330–342, 2016.

[7] Yuxiang Zhang, Bo Du, Liangpei Zhang, and Tongliang Liu. Joint sparse representation and multitask learning for hyperspectral target detection. *IEEE Transactions on Geoscience and Remote Sensing*, 55(2):894–906, 2017.

[8] Weiying Xie, Jie Lei, Yuhang Cui, Yunsong Li, and Qian Du. Hyperspectral pansharpening with deep priors. *IEEE Transactions on Neural Networks and Learning Systems*, PP:1–15, 06 2019.

[9] Chiman Kwan, Joon Hee Choi, Stanley H. Chan, Jin Zhou, and Bence Budavari. A super-resolution and fusion approach to enhancing hyperspectral images. *Remote Sensing*, 10(9), 2018.

[10] Shuiping Gou, Shuzhen Liu, Shuyuan Yang, and Licheng Jiao. Remote sensing image super-resolution reconstruction based on nonlocal pairwise dictionaries and double regularization. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 7(12):4784–4792, 2014.

[11] Lianru Gao, Danfeng Hong, Jing Yao, Bing Zhang, Paolo Gamba, and Jocelyn Chanussot. Spectral superresolution of multispectral imagery with joint sparse and low-rank learning. *IEEE Transactions on Geoscience and Remote Sensing*, 59(3):2269–2280, 2021.

[12] Shaohui Mei, Yunhao Geng, Junhui Hou, and Qian Du. Learning hyperspectral images from rgb images via a coarse-to-fine cnn. *Science China Information Sciences*, 65:152102, 05 2022.

[13] Yao Wang, Xi'ai Chen, Zhi Han, and Shiying He. Hyperspectral Image Super-Resolution via Nonlocal Low-Rank Tensor Approximation and Total Variation Regularization. *Remote Sensing*, 9(12):1286, December 2017.

[14] Songze Tang, Yang Xu, Lili Huang, and Le Sun. Hyperspectral Image Super-Resolution via Adaptive Dictionary Learning and Double $\ell 1$ Constraint. *Remote Sensing*, 11(23):2809, November 2019.

[15] Kui Jiang, Zhongyuan Wang, Peng Yi, and Junjun Jiang. Hierarchical dense recursive network for image super-resolution. *Pattern Recognition*, 107:107475, 2020.

[16] Enhai Liu, Zhenjie Tang, Bin Pan, and Zhenwei Shi. Spatial-Spectral Feedback Network for Super-Resolution of Hyperspectral Imagery. *arXiv e-prints*, page arXiv:2103.04354, March 2021.

[17] Denghong Liu, Jie Li, and Qiangqiang Yuan. A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7711–7725, 2021.

[18] Kui Jiang, Zhongyuan Wang, Peng Yi, Guangcheng Wang, Tao Lu, and Junjun Jiang. Edge-enhanced gan for remote sensing image superresolution. *IEEE Transactions on Geoscience and Remote Sensing*, 57(8):5799–5812, 2019.

[19] Xinya Wang, Jiayi Ma, and Junjun Jiang. Hyperspectral image super-resolution via recurrent feedback embedding and spatial-spectral consistency regularization. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–13, 2021.

[20] Yiqun Mei, Yuchen Fan, Yuqian Zhou, Lichao Huang, Thomas S. Huang, and Honghui Shi. Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5689–5698, 2020.

[21] Denghong Liu, Jie Li, and Qiangqiang Yuan. A spectral grouping and attention-driven residual dense network for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 59(9):7711–7725, 2021.

[22] Naoto Yokoya, Claas Grohnfeldt, and Jocelyn Chanussot. Hyperspectral and multispectral data fusion: A comparative review of the recent literature. *IEEE Geoscience and Remote Sensing Magazine*, 5(2):29–56, 2017.

[23] Laetitia Loncan, Luis B. de Almeida, Jose M. Bioucas-Dias, Xavier Briottet, Jocelyn Chanussot, Nicolas Dobigeon, Sophie Fabre, Wenzhi Liao, Giorgio A. Licciardi, Miguel Simões, Jean-Yves Tourneret, Miguel Angel Veganzones, Gemine Vivone, Qi Wei, and Naoto Yokoya. Hyperspectral pansharpening: A review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):27–46, 2015.

[24] Naoto Yokoya, Takehisa Yairi, and Akira Iwasaki. Coupled non-negative matrix factorization (cnmf) for hyperspectral and multispectral data fusion: Application to pasture classification. In *2011 IEEE International Geoscience and Remote Sensing Symposium*, pages 1779–1782, 2011.

[25] Eliot Wycoff, Tsung-Han Chan, Kui Jia, Wing-Kin Ma, and Yi Ma. A non-negative sparse promoting algorithm for high resolution hyperspectral imaging. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1409–1413, 2013.

[26] Qi Wei, Jose Bioucas-Dias, Nicolas Dobigeon, Jean-Yves Tourneret, Marcus Chen, and Simon Godsill. Multiband Image Fusion Based on Spectral Unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 54(12):7236–7249, December 2016.

[27] Renwei Dian, Shutao Li, Leyuan Fang, and Qi Wei. Multispectral and hyperspectral image fusion with spatial-spectral sparse representation. *Information Fusion*, 49:262–270, 2019.

[28] Jize Xue, Yong-Qiang Zhao, Yuanyang Bu, Wenzhi Liao, Jonathan Cheung-Wai Chan, and Wilfried Philips. Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 30:3084–3097, 2021.

[29] Renwei Dian, Leyuan Fang, and Shutao Li. Hyperspectral image super-resolution via non-local sparse tensor factorization. pages 3862–3871, 07 2017.

[30] Charilaos I. Kanatsoulis, Xiao Fu, Nicholas D. Sidiropoulos, and Wing-Kin Ma. Hyperspectral super-resolution: A coupled

tensor factorization approach. *IEEE Transactions on Signal Processing*, 66(24):6503–6517, 2018.

[31] Renwei Dian, Shutao Li, and Leyuan Fang. Learning a low tensor-train rank representation for hyperspectral image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, 30(9):2672–2683, 2019.

[32] Yang Xu, Zebin Wu, Jocelyn Chanussot, and Zhihui Wei. Nonlocal patch tensor sparse representation for hyperspectral image super-resolution. *IEEE Transactions on Image Processing*, 28(6):3034–3047, 2019.

[33] Kaidong Wang, Yao Wang, Xi-Le Zhao, Jonathan Cheung-Wai Chan, Zongben Xu, and Deyu Meng. Hyperspectral and multispectral image fusion via nonlocal low-rank tensor decomposition and spectral unmixing. *IEEE Transactions on Geoscience and Remote Sensing*, 58(11):7654–7671, 2020.

[34] Yifan Zhang, Arno Duijster, and Paul Scheunders. A bayesian restoration approach for hyperspectral images. *IEEE Transactions on Geoscience and Remote Sensing*, 50(9):3453 – 3462, 2012.

[35] Naveed Akhtar, Faisal Shafait, and Ajmal Mian. Bayesian sparse representation for hyperspectral image super resolution. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3631–3640, 2015.

[36] Yi Chang, Luxin Yan, Xi-Le Zhao, Houzhang Fang, Zhijun Zhang, and Sheng Zhong. Weighted low-rank tensor recovery for hyperspectral image restoration. *IEEE Transactions on Cybernetics*, 50(11):4558–4572, 2020.

[37] Qi Wei, Nicolas Dobigeon, and Jean-Yves Tourneret. Bayesian fusion of multi-band images. *IEEE Journal of Selected Topics in Signal Processing*, 9(6):1117–1127, 2015.

[38] Wei Wei, Jiangtao Nie, Yong Li, Lei Zhang, and Yanning Zhang. Deep recursive network for hyperspectral image super-resolution. *IEEE Transactions on Computational Imaging*, 6:1233–1244, 2020.

[39] Jiaojiao Li, Ruxing Cui, Bo Li, Rui Song, Yunsong Li, Yuchao Dai, and Qian Du. Hyperspectral image super-resolution by band attention through adversarial learning. *IEEE Transactions on Geoscience and Remote Sensing*, 58(6):4304–4318, 2020.

[40] Jing Hu, Xiuping Jia, Yunsong Li, Gang He, and Minghua Zhao. Hyperspectral image super-resolution via intrafusion network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10):7459–7471, 2020.

[41] Weiying Xie, Yuhang Cui, Yunsong Li, Jie Lei, Qian Du, and Jiaojiao Li. Hpgan: Hyperspectral pansharpening using 3-d generative adversarial networks. *IEEE Transactions on Geoscience and Remote Sensing*, 59(1):463–477, 2021.

[42] Shaohui Mei, Xin Yuan, Jingyu Ji, Yifan Zhang, and Qian Du. Hyperspectral image spatial super-resolution via 3d full convolutional neural network. *Remote Sensing*, 9:1139, 11 2017.

[43] Qiang Li, Qi Wang, and Xuelong Li. Exploring the relationship between 2d/3d convolution for hyperspectral image super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 59(10):8693–8703, 2021.

[44] Frosti Palsson, Johannes R. Sveinsson, and Magnus O. Ulfarsson. Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network. *IEEE Geoscience and Remote Sensing Letters*, 14(5):639–643, 2017.

[45] Yunsong Li, Jing Hu, Xi Zhao, Weiying Xie, and JiaoJiao Li. Hyperspectral image super-resolution using deep convolutional neural network. *Neurocomputing*, 266:29–41, 2017.

[46] Jingxiang Yang, Yongqiang Zhao, and Jonathan Chan. Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network. *Remote Sensing*, 10:800, 05 2018.

[47] Shaohui Mei, Ruituo Jiang, Xu Li, and Qian Du. Spatial and spectral joint super-resolution using convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 58(7):4590–4603, 2020.

[48] Jinfan Hu, Tingzhu Huang, Liangjian Deng, Taixiang Jiang, Gemine Vivone, and Jocelyn Chanussot. Hyperspectral image super-resolution via deep spatiospectral attention convolutional neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–15, 2021.

[49] Xian-Hua Han, YinQiang Zheng, and Yen-Wei Chen. Multi-level and multi-scale spatial and spectral fusion cnn for hyperspectral image super-resolution. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pages 4330–4339, 2019.

[50] Ying Qu, Hairong Qi, Chiman Kwan, Naoto Yokoya, and Jocelyn Chanussot. Unsupervised and unregistered hyperspectral

image super-resolution with mutual dirichlet-net. *IEEE Transactions on Geoscience and Remote Sensing*, pages 1–18, 2021.

[51] Jing Yao, Danfeng Hong, Jocelyn Chanussot, Deyu Meng, Xiaoxiang Zhu, and Zongben Xu. Cross-attention in coupled unmixing nets for unsupervised hyperspectral super-resolution. pages 208–224, 2020.

[52] Bruno Aiazzi, Stefano Baronti, and Massimo Selva. Improving component substitution pansharpening through multivariate regression of ms +pan data. *IEEE Transactions on Geoscience and Remote Sensing*, 45(10):3230–3239, 2007.

[53] Shutao Li, Renwei Dian, Leyuan Fang, and José M. Bioucas-Dias. Fusing hyperspectral and multispectral images via coupled sparse tensor factorization. *IEEE Transactions on Image Processing*, 27(8):4118–4130, 2018.

[54] Ruiyuan Wu, Wing-Kin Ma, Xiao Fu, and Qiang Li. Hyperspectral super-resolution via globallocal low-rank matrix estimation. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10):7125–7140, 2020.

[55] Yuxuan Zheng, Jiaojiao Li, Yunsong Li, Kailang Cao, and Keyan Wang. Deep residual learning for boosting the accuracy of hyperspectral pansharpening. *IEEE Geoscience and Remote Sensing Letters*, 17(8):1435–1439, 2020.

[56] Ruiyuan Wu, Chun-Hei Chan, Hoi-To Wai, Wing-Kin Ma, and Xiao Fu. Hi bcd! hybrid inexact block coordinate descent for hyperspectral super-resolution. pages 2426–2430, 2018.