Exploring the Relationship between 2D/3D Convolution for Hyperspectral Image Super-Resolution

Qiang Li, Qi Wang, *Senior Member, IEEE,* and Xuelong Li, *Fellow, IEEE*

***Abstract*—Hyperspectral image super-resolution (SR) methods based on deep learning have achieved significant progress recently. However, previous methods lack the joint analysis between spectrum and horizontal or vertical direction. Besides, when both 2D and 3D convolution are in the network, the existing models can not effectively combine the two. To address these issues, in this paper, we propose a novel hyperspectral image SR method by exploring the relationship between 2D/3D convolution (ERCSR). Our method alternately employs 2D and 3D units to solve the problem of structural redundancy by sharing spatial information during reconstruction for existing model, which can enhance the learning ability of 2D spatial domain. Importantly, compared with the network using 3D unit, i.e., 2D unit is replaced by 3D unit, it can not only reduce the size of the model, but also improve the performance of the model. Furthermore, to exploit the spectrum fully, the split adjacent spatial and spectral convolution (SAEC) is designed to parallelly explore information between spectrum and horizontal or vertical direction in space. Experiments on widely used benchmark datasets demonstrate that the proposed approach outperforms state-of-the-art SR algorithms across different scales in terms of quantitative and qualitative analysis.**

***Index Terms*—Hyperspectral image, super-resolution (SR), convolutional neural networks (CNNs), hybrid convolution, split adjacent spatial and spectral convolution (SAEC).**

I. Introduction

HYPERSPECTAL imaging system gathers tens to hundreds of spectral bands from the object area to obtain hyperspectral image. While collecting the spatial information, the spectrum is also obtained. This has greatly improved the degree of information richness, so it is widely applied in mineral exploration [1], medical diagnosis [2], etc. Nevertheless, the physical limitations of spectral sensors often hinder the acquisition of high-resolution hyperspectral image in practical applications. It affects the subsequent analysis for high-level tasks, such as image classification [3], [4], change detection [5], and anomaly detection [6].

To solve this challenge, the hyperspectral image super-resolution (SR) is proposed [7]–[12]. It aims to restore LR hyperspectral image to high-resolution (HR) hyperspectral image, so as to better and accurately describe objects. Since substances behave distinction in different bands of spectral signals, usually, some bands in hyperspectral image are select to examine in practical applications [13]–[16]. Thus, unlike natural image (RGB image) used for SR task, the spectral distortion needs to be considered for hyperspectral image SR. It means that the spectral distortion should be reduced as much as possible during reconstruction, which is also an important index to evaluate the restored hyperspectral image.

Hyperspectral image usually divides more bands within the limited spectrum to improve spectral resolution. As a result, its spatial resolution is lower than that of natural images or multispectral images. Inspired by this discovery, the researches propose many SR methods by fusing LR hyperspectral image with its corresponding HR RGB image [17], [18]. These methods generate the corresponding RGB image by integrating the HR hyperspectral image and its spectral dimension using same camera spectral response (CSR) [19]. Although the approaches have obtained good results, the differences of CSR in datasets or scenes are ignored, obtaining poor robustness. Later, Fu *et al.* [20] design an automatic CSR selection mechanism to address the above trouble. Nevertheless, the fusion strategy claims that the image pair is well matched in different datasets or scenes, which makes it extremely difficult in practical applications. Thus, the hyperspectral image SR is executed without using fusion strategy in our paper.

Due to the strong representation ability of convolutional neural networks (CNNs), the performance of natural image SR has been greatly advanced in recent years [21], [22]. It aims to learn the mapping function between LR and HR RGB image by means of supervision. Aiming at the inherent properties of hyperspectral image, various methods using 2D convolution are designed by referring to natural image SR methods [23]–[26]. For example, inspired by deep recursive residual network [27], Li *et al.* [24] present a new grouped recursive module and embed it into the global residual structure (GDRRN). To avoid the spectral distortion, the network joints Spectral Angle Mapper (SAM) with Mean Squared Error (MSE) to optimize the network parameters during reconstruction. Nevertheless, the designed loss function influences the performance of spatial resolution. Li *et al.* [25] propose deep spectral difference network. After achieving the spatial reconstruction of hyperspectral image, similarly, the post-processing is carried out to avoid the spectral distortion. As spectral information is not utilized, the type of above algorithms generally has poor performance.

Since hyperspectral image contains abundant spectral information, the methods employing 3D convolution have become a

This work was supported by the National Natural Science Foundation of China under Grant U1864204, 61773316, U1801262, and 61871470.

The authors are with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an 710072, China (e-mail: liqmges@gmail.com, crabwq@gmail.com, xuelong_li@nwpu.edu.cn) (*Corresponding author: Qi Wang.*)

research topic, which is based on the fact that spectral features can improve the performance for spatial resolution [28]–[30]. Mei *et al.* [28] first propose 3D full CNN (3D-FCNN) to explore both spatial context and spectral correlation. Because the spectral information is considered, the network obtains better performance. Yang *et al.* [29] develop multi-scale wavelet 3D convolutional neural network with embedding and predicting subnet. The network requires pre-processing and post-processing in terms of wavelet transformation. All the methods described above use regular 3D convolution to process hyperspectral image, and there are many similar methods, such as [31] and [32]. Different from 2D convolution, a regular 3D convolution is performed by convoluting 3D kernel and feature map. It results in a significant increase in network parameters. Considering this shortcoming, the researchers modify the filter $k \times k \times k$ as $k \times 1 \times 1$ and $1 \times k \times k$ [30], [33]–[35]. Typical algorithms have SSRNet [33] and MCNet [30]. By doing so, the network parameters are reduced dramatically, making it possible to design the network more deeply. As for SSRNet algorithm [33], all layers are conducted by above operation. However, it generates redundant information in feature maps along the spectral dimension due to the existence of high similarity among bands. Moreover, when the model can explore spectral dimension, it lacks of more learning ability in space. Later, Li *et al.* [30] propose mixed convolution module (MCNet) by sharing spatial information to design several 2D and 3D units. The model effectively addresses the existing drawbacks in SSRNet. However, it adopts parallel structure to extract the features, resulting in module redundancy.

With respect to the above descriptions, it can be concluded that how to effectively combine the 2D and 3D unit still needs more research efforts. Additionally, all the above methods only consider the relationship of space and spectrum using such convolution operation (i.e., the filter is $k\times1\times1$ and $1\times k\times k$), ignoring the exploration between spectrum and horizontal or vertical direction in space (see Fig. 1). Motivated by these discoveries, in this paper, hyperspectral image SR is achieved via exploring the relationship between 2D/3D convolution (ERCSR). In summary, our main contributions are as follows:

- A new structure that appears alternately through 2D and 3D units is proposed. Under sharing spatial information between 2D and 3D unit, it overcomes the problem of redundancy caused by parallel structure in MCNet [30]. Besides, it also improves the learning ability of spatial domain via designing more 2D units.
- The split adjacent spatial and spectral convolution (SAEC) is proposed. By separating the filter, it fully explores the potential features between spectrum and horizontal or vertical direction in space, which alleviates the spectral distortion of the reconstructed image.
- Extensive experiments on three public datasets demonstrate that the proposed model outperforms the state-of-the-art methods across different scales in both quantitatively and qualitatively.

The remainder of this paper is organized as follows: Section II describes several existing typical networks. Section III introduces the proposed ERCSR, including network structure, enhanced hybrid convolution module, etc. Then, experiments on public datasets are performed to verify our method in Section IV. Finally, the conclusion is given in Section V.

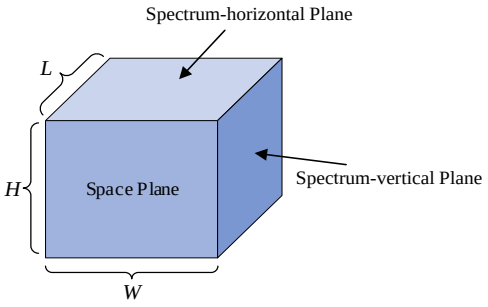


Fig. 1. Illustration of hyperspectral image cube with spatial convolution in space plane, spectrum-horizontal convolution in spectrum-horizontal plane, and spectrum-vertical convolution in spectrum-vertical plane, where W and H are the width and height of each band in the spatial domain, L represents the total number of the bands.

II. Related Work

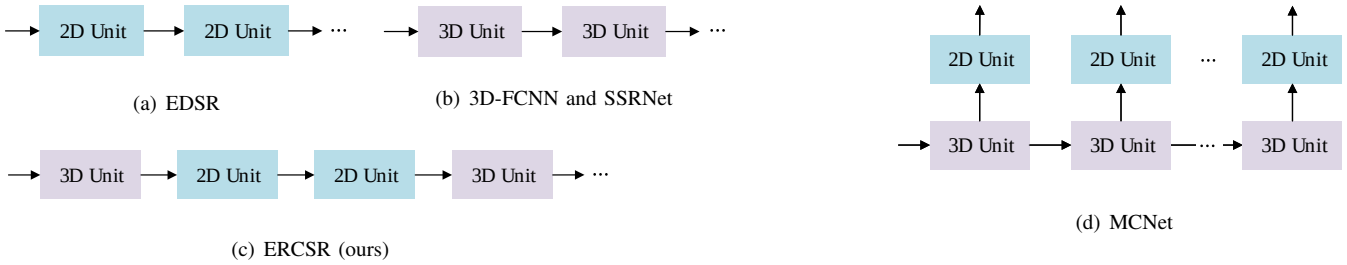
In this section, we describe in detail the existing typical networks applying 2D and 3D convolution, including EDSR [36], 3D-FCNN [28], SSRNet [33], and MCNet [30]. Fig. 2 shows simplified structure of these methods. Here, 2D and 3D unit refer to the use of 2D and 3D convolution in corresponding unit, respectively.

A. EDSR

As for EDSR algorithm [36], the whole model uses 2D convolution to explore the natural image. The network is stacked by 16 residual units. Its simplified structure is shown Fig. 2(a). With respect to the unit, it contains two 2D convolutions, ReLU activation function, and local residual connection, whose mathematical formulation is presented in Table I. For hyperspectral image SR, the model adopting 2D convolution cannot effectively exploit spectral information to enhance the learning ability of 2D spatial domain. Therefore, the algorithm to process hyperspectral image SR has poor performance.

B. 3D-FCNN

Since hyperspectral image has rich spectral information [37], [38], Mei [28] *et al.* first introduce regular 3D convolution to implement hyperspectral image SR (3D-FCNN). The model structure contains four convolution layers. Similar to SRCNN [39], the difference is that 3D convolution is adopted for each layer instead of 2D convolution (see Fig. 2(b)). As for main module in this network, it is composed of 3D convolution and ReLU, as shown in Table I. As all convolution operations are not padded, the size of the reconstructed hyperspectral image is changed. Moreover, this method lacks of residual connection. As a result of the above problems, the performance of the algorithm is not so ideal. However, this novel approach has inspired many scholars to design networks in this way, such as [30] and SSRNet [33].



(a) EDSR (b) 3D-FCNN and SSRNet (c) ERCSR (ours) (d) MCNet

Fig. 2. Simplified structure of several methods.

TABLE I
MATHEMATICAL FORMULATIONS OF 2D AND 3D UNIT IN EDSR [36], 3D-FCNN [28], SSRNET [33], AND MCNET [30]. $conv_{2D}(\cdot)$ AND $conv_{3D}(\cdot)$ DENOTE FUNCTIONS OF 2D AND 3D CONVOLUTION, RESPECTIVELY. $\sigma(\cdot)$ IS RELU ACTIVATION FUNCTION. $R(\cdot)$ REPRESENTS RESHAPE OPERATION.

Method	Key Strategy		Mathematical Formulation
	2D Unit	3D Unit	
EDSR [36]	two 2D convolutions + ReLU + local residual learning	—	$y = \sigma(conv_{2D}(\sigma(conv_{2D}(x)))) + x$
3D-FCNN [28]	—	3D convolution + ReLU	$y = \sigma(conv_{3D}(x))$
SSRNet [33]	—	four 3D convolutions + four ReLUs + local residual learning	$y = \sigma(conv_{3D}((\sigma(conv_{3D}(x)))))$ $z = \sigma(conv_{3D}(\sigma(conv_{3D}(y)))) + x$
MCNet [30]	two 2D convolutions + ReLU	two 3D convolutions + two ReLUs + local residual learning	$y = \sigma(conv_{3D}(\sigma(conv_{3D}(x)))) + x$ $z = R(conv_{2D}(\sigma(conv_{2D}(R(y)))))$

C. SSRNet

While many deep learning methods applying 3D convolution are proposed, there is main issue that using regular 3D convolution obviously leads to a significant increase in network parameters. This prevents the network from being designed deeper. Considering this limitation, Wang *et al.* develop spatial-spectral residual network (SSRNet) [33]. The network splits regular 3D kernel $3 \times 3 \times 3$ into $1 \times 3 \times 3$ and $3 \times 1 \times 1$, namely separable 3D convolution [40], to extract spatial and spectral features, respectively. It dramatically reduces unaffordable memory usage and training time. Note that separable 3D convolution is a special form of regular 3D convolution. The whole network is conducted by three spatial-spectral residual modules, and each module is mainly composed of three 3D units. As can be seen from Table I and Fig. 2(b), when the model can explore spectral dimension, it lacks of more learning ability in space, i.e., it treats both space and spectrum with same number of layers. This problem also exists in some literature [28], [29], [32].

D. MCNet

To tackle the issue like SSRNet, Li *et al.* propose mixed 2D/3D convolution network (MCNet) [30]. This method adopts four mixed convolution modules to distinguish the mining of spatial and spectral information. By sharing spatial information, it attempts to increase the spatial exploration under the condition that the spectral content can be extracted. The simplified structure of its network is displayed in Fig. 2(d). We can observe that the output of each 3D unit is fed to the corresponding 2D unit. This parallel structure results in module redundancy. Moreover, it can be seen from the mathematical formula in Table I that there is no residual connection between 2D and 3D unit. It hinders the information flow of two units when the feature maps is changed. These problems make it impossible to improve the representation ability effectively. With respect to the above descriptions, it can be concluded that how to combine the two is urgent in the presence of 2D and 3D unit. Motivated by this, we design a new structure to explore hyperspectral image SR in this paper. The model alternately employ 2D and 3D units (see Fig. 2(c)), which greatly reduces the complexity of feature learning within 3D unit and effectively promotes the optimization of the whole network.

III. PROPOSED METHOD

In this section, we describe the proposed method in detail from the following aspects, including network structure, enhanced hybrid convolution module, and 2D/3D unit.

A. Network Structure

As shown in Fig. 3, the proposed ERCSR mainly contains three parts: feature extraction, image reconstruction, and residual skip connection. For hyperspectral image SR, let $I_{LR} \in R^{L \times W \times H}$ and $I_{SR} \in R^{L \times rW \times rH}$ denote the input LR hyperspectral image and reconstructed SR hyperspectral image, respectively, where W and H are the width and height of each band in the spatial domain, L represents the total number of the bands. The scale factor r is scale that specifies the desired size of the generated HR image for the LR image. As described in Section I, hypersepctral image SR requires attention to the reconstruction of space and spectrum. Therefore, we utilize separable 3D convolution (it is defined as SConv in Fig. 3) that has been proved to be comparable to the performance of regular 3D convolution [40] (see Fig.

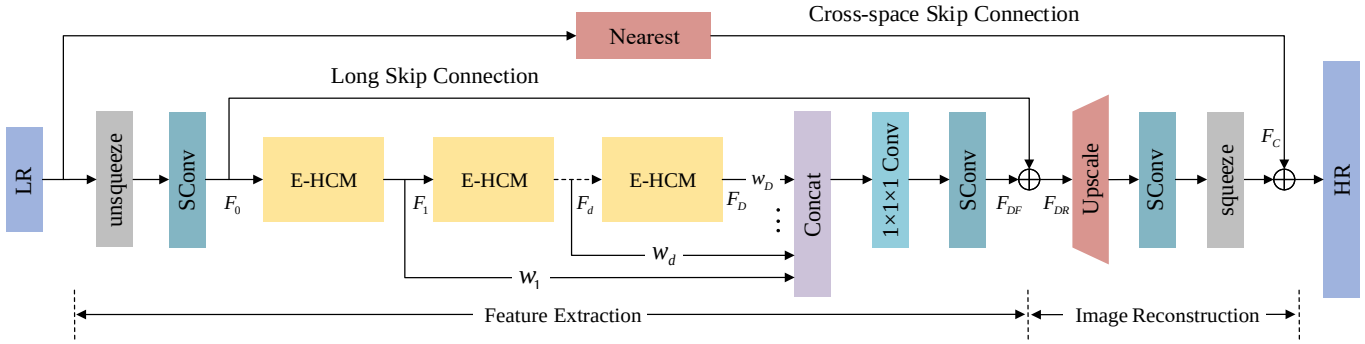


Fig. 3. Overall architecture of our proposed ERCSR.

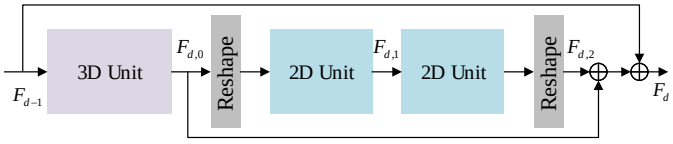


Fig. 4. Architecture of enhanced hybrid convolution module (E-HCM).

5(a)) to extract shallow spectral information after reshaping I_{LR} into four dimensions ($1 \times L \times W \times H$), i.e.,

$$F_0 = f_{sconv3D}(unsqueeze(I_{LR})), \tag{1}$$

where $unsqueeze(\cdot)$ is used to expand I_{LR}, and $f_{sconv3D}(\cdot)$ denotes separable 3D convolution operation. Then, these initial features are fed into enhanced hybrid convolution module (E-HCM). Assuming that we have D E-HCMs in network, the output F_D are denoted as

$$F_D = J_D(J_{D-1}((...J_1(F_0)...))), \tag{2}$$

where $J_d(\cdot)$ denotes the d-th E-HCM. After obtaining hierarchical features by D E-HCMs, they are concatenated together to enable the network to learn more effective information. Through $1 \times 1 \times 1$ convolution and separable 3D convolution, we finally acquire the output F_{DR} for feature extraction part after long skip connection, i.e.,

$$F_{DR} = F_{DF} + F_0. \tag{3}$$

With respect to the part of image reconstruction, we upsample F_{DR} in HR space by transposed convolution according to r, which is followed by a separable 3D convolution. As the input and output images are largely similar, an additional cross-space residual F_C is introduced to upsample the input LR image to HR space by *nearest*. It can greatly alleviate the burden on the model. After the feature maps are squeezed in three dimensions ($L \times W \times H$), the output I_{SR} is finally obtained by

$$I_{SR} = squeeze(f_{sconv3D}(up(F_{DR}))) + F_C, \tag{4}$$

where $up(\cdot)$ represents 3D transposed convolution layer, and $squeeze(\cdot)$ is squeeze function.

B. Enhanced Hybrid Convolution Module

Now we present the proposed E-HCM, whose structure is depicted in Fig. 4. The module is composed of 3D unit, 2D unit, and two reshape operations. First, we utilize 3D unit to analyze the relationship of spectrum and either horizontal or vertical direction in space. To increase the spatial exploration of image under the condition that the spectral content can be obtained, the feature maps after 3D unit are reshaped in four dimensions to perform 2D convolution. Concretely, assume that the size of feature maps is $N \times C \times L \times W \times H$ when the batch size N is considered, where C is the number of filters. To transform it, we treat each band separately in our work. The channel L and N are integrated together, i.e., $N * L \times C \times W \times H$. By doing so, there are two benefits. On the one hand, the design of 2D convolution in the network is beneficial to easily optimize network in contrast to 3D convolution. On the other hand, the whole network can more focus on spatial information to improve the spatial learning ability, when the spectral information can be extracted. We also adopt two local residual connections at the end of this module. It not only facilitates information fusion within the 3D unit, but also makes it easier for the network to study 2D features in 2D unit. Both of them improve the optimization of the whole model. As for the output of d-th module, it can be represented as

$$F_d = F_{d,2} + F_{d,0} + F_{d-1}. \tag{5}$$

Our proposed module appears alternately through 2D and 3D units, which greatly reduces the complexity of 3D feature learning. It solves the problem of redundancy caused by parallel structure in MCNet [30]. Importantly, compared with the network using 3D unit, i.e., 2D unit is replaced by 3D unit, it can not only reduce the size of the model, but also improve the performance of the model.

C. 3D Unit

Previous methods [28], [30], [31], [33] lack the joint analysis between spectrum and horizontal or vertical dimension in space. It results in the poor representation ability of the network. To address this issue, the split adjacent spatial and spectral convolution (SAEC) is designed to parallelly handle the relationship of spectrum with either horizontal or vertical direction at the front end of the module. The architecture is

shown in Fig. 5(c). It is natural decomposition of a regular 3D convolution. Specifically, the input F_{d-1} is processed by a convolution layer with the filter $1 \times k \times k$, which is applied to explore spatial content. To study the relationship of spectral dimension and other dimensions, the filter $k \times k \times k$ is separated in two forms, i.e., $k \times 1 \times k$ and $k \times k \times 1$. Through an addition operation, we have

$$T = \sigma(f_{ss}(F_{d-1})), \tag{6}$$

$$F_{d,0} = f_{sh}(T) + f_{sv}(T), \tag{7}$$

where $f_{ss}(\cdot)$, $f_{sh}(\cdot)$, and $f_{sv}(\cdot)$ represent convolution operations for space, spectrum and horizontal, and spectrum and vertical, respectively. By separating the filter, it can jointly mine the information of spectrum and other two directions, which effectively alleviates the spectral distortion of the reconstructed image.

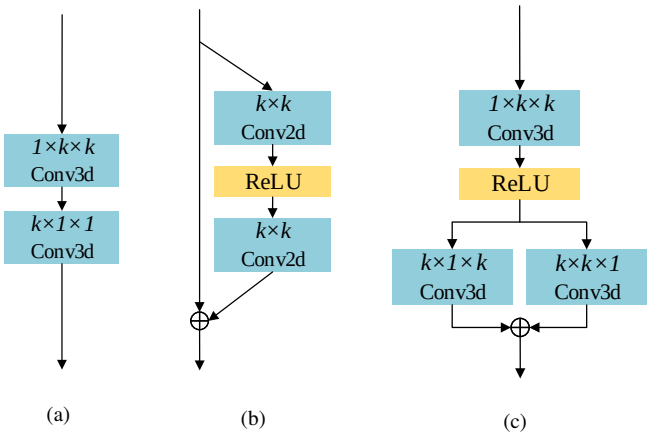


Fig. 5. Architecture of various convolutions. (a) Separable 3D convolution in SSRNet [33]. (b) 2D convolution in 2D unit. (c) SAEC in 3D unit.

D. 2D Unit

Due to the strong representation ability of CNNs, the performance of natural image SR has been greatly advanced recently [21], [36], [41]. As for the mainstream methods of natural image SR, the 2D convolution and residual connection are often employed as the main module. Therefore, we utilize the main module in natural image SR for reference to explore spatial features. In our paper, the 2D unit consists of two convolution layers, ReLU function, and residual connection, which is shown in Fig. 5(b). Supposing the reshaped results are still expressed as $F_{d,1}$ for d-th module, all processes for first unit are summarized as

$$F_{d,1} = f_{conv2D}(\sigma(f_{conv2D}(F_{d,0}))) + F_{d,0}, \tag{8}$$

where $F_{d,1}$ is the output of first 2D unit, and $f_{conv2D}(\cdot)$ denotes 2D convolution operation. To learn more spatial information, in our work, another 2D unit is added in this module. Similarly, the reshaped results in second 2D unit are still denoted as $F_{d,2}$. We finally obtain

$$F_{d,2} = f_{conv2D}(\sigma(f_{conv2D}(F_{d,1}))) + F_{d,1}. \tag{9}$$

Compared with all operations are done by 3D convolution, our designed network can significantly reduce the number of parameters. Furthermore, it can pay more attention to spatial resolution, thus dramatically improving the performance.

IV. Experiments

In this section, we evaluate our network both qualitatively and quantitatively. First, the benchmark datasets and implementation details are provided. Then, we analyze the effectiveness of model. Finally, we compare ERCSR to other state-of-the-art methods on benchmark datasets.

A. Datasets

1) CAVE: The CAVE dataset was collected by a tunable filter and a cooled CCD camera from range of 400 nm to 700 nm in steps of 10 nm. The 31-bands hyperspectral images contains wide scenes, such as skin, drinks, vegetables, etc. The spatial resolution of each hyperspectral image is 512×512 pixels.

2) Harvard: The Harvard dataset was obtained by Nuance FX, CRI Inc. camera from indoor or outdoor scenes under daylight illumination. Compared with CAVE dataset, it has more hyperspectral images (71 images). The size of each hyperspectral image cube is $31 \times 1040 \times 1392$.

3) Pavia Center: The Pavia Centre dataset was captured by the ROSIS sensor over Pavia, nothern Italy. Unlike the above datasets, it is hyperspectral remote sensing dataset and only contains one image. The image consists of 1096×715 pixels and 102 spectral reflectance bands.

B. Implementation Details

As we introduced in Section IV-A, these datasets are captured via different hyperspectral imaging cameras. It indicates that there is no the same attributes between them, which leads to training each dataset individually. With respect to CAVE and Harvard dataset, one can notice that they both consist of dozens of images, which is unlike to the Pavia Centre dataset. As for these two datasets, we randomly select 80% samples from each dataset for training and the rest for testing. The trained samples for CAVE and Harvard dataset are augmented by randomly choosing 24 patches from each image. With regard to Pavia Centre dataset, the top left 876×715 is selected to train, and the rest of image is used to test. 108 patches are randomly selected from the image to augment the training samples. After getting these patches for three datasets, each patch is scaled to 1, 0.75, and 0.5, respectively. Then, the scaled patch is rotated by $90°, 180°, 270°$ and horizontally flipped. Though bicubic interpolation, they are downsampled to acquire LR image with size of $L \times 32 \times 32$, where L is total number of band. Before feeding the mini-batch into the network, the average value of training images is subtracted.

In our work, $L1$ loss function is employ to study the model. The parameter k of the kernel is set to 3, and the number of the filters is defined as 64. The optimizer ADAM (β_1 = 0.9, β_2 = 0.999) is adopted to learn designed network. The batch size is 12. The total number of training epochs is 200. We initialize the learning rate of all layers to 10^{-4}, which is halved by each 35 epochs. Our algorithm is conducted on PyTorch framework with NVIDIA GeForce GTX 1080 GPU.

TABLE II
STUDY OF THE NUMBER OF E-HCM MODULE.

Evaluation metric	2	3	4	5
PSNR	44.980	45.219	45.332	45.339
SSIM	0.9737	0.9740	0.9740	0.9740
SAM	2.220	2.210	2.208	2.209

TABLE III
PERFORMANCE OF THE INFLUENCE OF DIFFERENT 3D CONVOLUTION TYPES.

Type	PSNR	SSIM	SAM	Parameters
regular 3D convolution	45.109	0.9740	2.210	1.348M
separable 3D convolution	45.069	0.9739	2.208	1.102M
SAEC	45.332	0.9740	2.208	1.349M

C. Evalution Metrics

To qualitatively evaluate the performance of reconstructed image, in our paper, three methods are adopted. They are Peak Signal-to-Noise Ratio (PSNR), Spectral Angle Mapper (SAM), and Structural SIMilarity (SSIM), which are defined as

$$PSNR = \frac{1}{L}\sum_{l=1}^{L} 10log_{10}\left(\frac{MAX_l^2}{MSE_l}\right), \tag{10}$$

$$MSE_l = \frac{1}{WH}\sum_{w=1}^{W}\sum_{h=1}^{H}\left(I_{SR}(w,h,l) - I_{HR}(w,h,l)\right)^2, \tag{11}$$

where MAX_l is the maximal pixel value for l-th band, and I_{HR} denotes HR hyperspectral image.

$$SSIM = \frac{1}{L}\sum_{l=1}^{L}\frac{\left(2\mu_{I_{SR}}^l\mu_{I_{HR}}^l + c_1\right)\left(2\sigma_{I_{SR}I_{HR}}^l + c_2\right)}{m * n}, \tag{12}$$

$$m = (\mu_{I_{SR}}^l)^2 + (\mu_{I_{HR}}^l)^2 + c_1, \tag{13}$$

$$n = (\sigma_{I_{SR}}^l)^2 + (\sigma_{I_{HR}}^l)^2 + c_2, \tag{14}$$

where $\mu_{I_{SR}}^l$ and $\mu_{I_{HR}}^l$ represent the mean of I_{SR} and I_{HR} for l-th band. $\sigma_{I_{SR}}^l$ and $\sigma_{I_{HR}}^l$ denote the variance of I_{SR} and I_{HR} for l-th band. $\sigma_{I_{SR}I_{HR}}^l$ is the covariance between I_{SR} and I_{HR} for l-th band. c_1 and c_2 are the constants to avoid

$$SAM = arccos\left(\frac{< I_{SR}, I_{HR} >}{||I_{SR}||_2||I_{HR}||_2}\right), \tag{15}$$

where $arccos(\cdot)$ is arccos function. $< \cdot, \cdot >$ is dot product. $||\cdot||_2$ denotes the L2 norm.

D. Model Analysis

In this section, we investigate the proposed model on CAVE dataset in details from four aspects, including the number of module D, the study of SAEC and E-HCM, and ablation study.

TABLE IV
PERFORMANCE ANALYSIS OF THE COMBINATION OF DIFFERENT UNITS. THE BOLD INDICATES THE APPROACH USED IN THIS PAPER.

Type	PSNR	SSIM	SAM	Parameters
2D unit	45.129	0.9738	2.217	1.201M
3D unit	44.991	0.9737	2.220	1.645M
2D unit and 3D unit	45.125	0.9739	2.227	1.053M
two 2D units and 3D unit	45.332	0.9740	2.208	1.349M
three 2D units and 3D unit	45.308	0.9738	2.228	1.645M
2D unit and two 3D units	45.283	0.9740	2.223	1.497M
2D unit and three 3D units	45.301	0.9738	2.232	1.941M

TABLE V
ABLATION STUDY ABOUT THE COMPONENTS.

Component	Different combinations of components			
2D unit	√	×	√	√
3D unit	×	√	√	√
residual connection	×	×	×	√
PSNR	44.974	44.778	45.195	45.332
SSIM	0.9737	0.9736	0.9739	0.9740
SAM	2.216	2.227	2.210	2.208

1) Study of Module D: To determine that how many E-HCM modules are appropriate, in our study, we set the parameter D from 2 to 5 to analyze the influence for this part, whose results are shown in Table II. As seen from this table, this parameter has a significant impact on overall network performance. Besides, the performance of three indices varies greatly from 2 to 3, especially for PSNR. However, when D is set to 4 and 5, the growth rates of PSNR, SSIM, and SAM basically keep unchanged. It indicates that the performance of the designed network tends to saturation. If the depth of the network is further increased, its performance is not significantly improved. Therefore, we empirically choose $D = 4$ to implement the following experiments.

2) Study of SAEC: In our work, we design split adjacent spatial and spectral convolution (SAEC) to parallelly explore the relationship of spectral dimension with others. To verify the effectiveness of the proposed convolution, other two convolution ways are introduced to replace SAEC. Table III describes the performance for different convolution types. Overall, SAEC produces superior results. As seen from this table, the number of parameters using SAEC is basically the same as that of regular 3D convolution. Under this case, there is certain gap in the values of PSNR, and other indices almost keep unchanged. Unlike the above two convolution ways, the number of parameters adopting separate 3D convolution has smaller than that of other types. However, its performance is relatively poor. The reason for this is that there are few parameters or no effective use of spectral information. By contrast, our designed SAEC achieves the best performance, which is mainly due to the joint analysis between spectrum and other dimensions. It is beneficial to the mining of potential information.

3) Study of E-HCM: As for E-HCM module, it consists of one 3D unit, two 2D units, and two reshape operations. Here, we replace these units in the module with the same type. Table IV exhibits the experimental performance under the same unit. Specifically, the 2D unit in each module is first replaced

TABLE VI
QUANTITATIVE EVALUATION OF STATE-OF-THE-ART SR ALGORITHMS BY AVERAGE PSNR/SSIM/SAM FOR DIFFERENT SCALE FACTORS ON CAVE DATASET. THE RED AND BLUE INDICATE THE BEST AND SECOND BEST PERFORMANCE, RESPECTIVELY.

Scale factor	Evalution metric	Bicubic	GDRRN [24]	3D-FCNN [28]	EDSR [36]	SSRNet [33]	MCNet [30]	ERCSR (ours)
×2	PSNR ↑	40.762	41.667	43.154	43.869	44.991	45.102	45.332
	SSIM ↑	0.9623	0.9651	0.9686	0.9734	0.9737	0.9738	0.9740
	SAM ↓	2.665	3.842	2.305	2.636	2.261	2.241	2.218
×3	PSNR ↑	37.562	38.834	40.219	40.533	40.896	41.031	41.345
	SSIM ↑	0.9325	0.9401	0.9453	0.9512	0.9524	0.9526	0.9527
	SAM ↓	3.522	4.537	2.930	3.175	2.814	2.809	2.789
×4	PSNR ↑	35.755	36.959	37.626	38.587	38.944	39.026	39.224
	SSIM ↑	0.9071	0.9166	0.9195	0.9292	0.9312	0.9319	0.9322
	SAM ↓	3.944	5.168	3.360	3.804	3.297	3.292	3.243

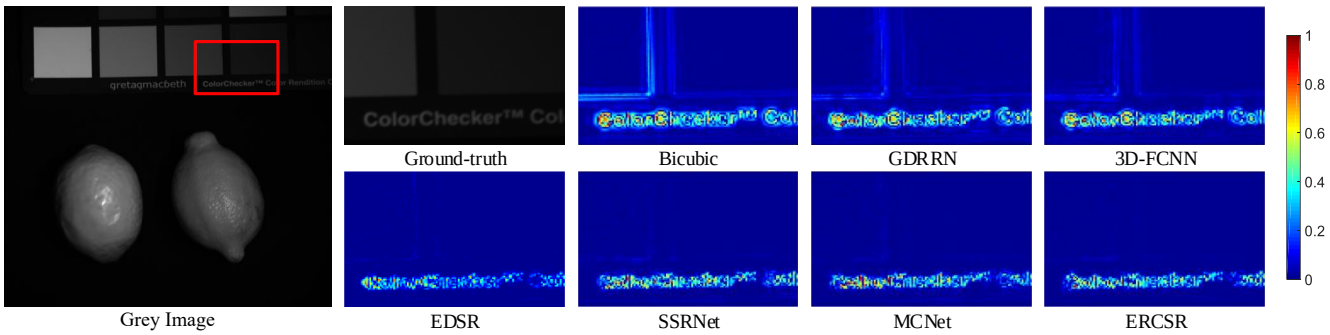


Fig. 6. Absolute error map comparisons for image *fake_and_real_lemons* at 660 nm on CAVE dataset.

with 3D unit, and two reshape operations are removed. The whole network stacked by 3D unit yields the worse results. In the case of this unit alone, the network mainly does not pay too much attention to spatial information mining, i.e., the number of feature extraction layers is the same for each dimension in E-HCM. When all units are taken to be 2D unit, it attains relatively better results. However, this way ignores the exploration of spectral information. For hyperspectral image SR, the aim of adopting spectral information is to enhance the performance of spatial resolution. As for the network with 2D and 3D units, when the spectral information can be extracted, the design of 2D units is helpful to improve the spatial learning ability of the whole network. Therefore, the combinations of different 2D and 3D units overall produce the better results in contrast to single unit. Among these combinations, we can find that both one 3D unit and two 2D units generate the best performance, and the number of network parameters is small. The combination not only considers the spectral information, but also can design more 2D convolution layers to extract features. It makes the performance of the three metrics better than that of other combinations.

4) Ablation Study: The proposed model mainly has three parts: feature extraction, image reconstruction, and residual skip connection. E-HCM in these parts is the main module of the whole network. In this section, we investigate the influence of different combinations about E-HCM on the performance of the model. Table V provides the ablation study about these combinations. Specifically, the E-HCM only has 2D or 3D unit, and the other components are removed. Their results are poor, particularly when the module exists 3D unit. As we introduced in Section I, the purpose using spectrum is to improve the performance of spatial reconstruction. From this point of view, the above situation is caused by paying too much attention to spectral information, while ignoring spatial information mining. When there are both 2D and 3D units without residual connection, the overall performance is obviously better than that of single unit. It indicates the structure that appears alternately through both units can improve the representation ability in space through spectral knowledge. It reveals that these components are an indispensable part for studying model. Finally, all components are attached to the module. We can notice that all results in three aspects are superior to any other combinations. Through these analyses, it can be concluded that each component contributes to network learning and optimization.

TABLE VII
COMPARISON OF THE NUMBER OF PARAMETERS OF THE ALGORITHM.

Method	×2	×3	×4
Bicubic	—	—	—
GDRRN [24]	219k	219k	219k
3D-FCNN [28]	39k	39k	39k
EDSR [36]	1404k	1589k	1552k
SSRNet [33]	830k	941k	1076k
MCNet [30]	1928k	2039k	2174k
ERCSR (ours)	1349k	1459k	1595k

E. Comparisons with the State-of-the-art Methods

In this section, we make a comprehensive comparison of the six existing methods with the proposed ERCSR. They include Bicubic, GDRRN [24], 3D-FCNN [28], EDSR [36], SSRNet [33], and MCNet [30]. Three benchmark datasets, CAVE, Harvard, and Pavia Centre, are employed to verify

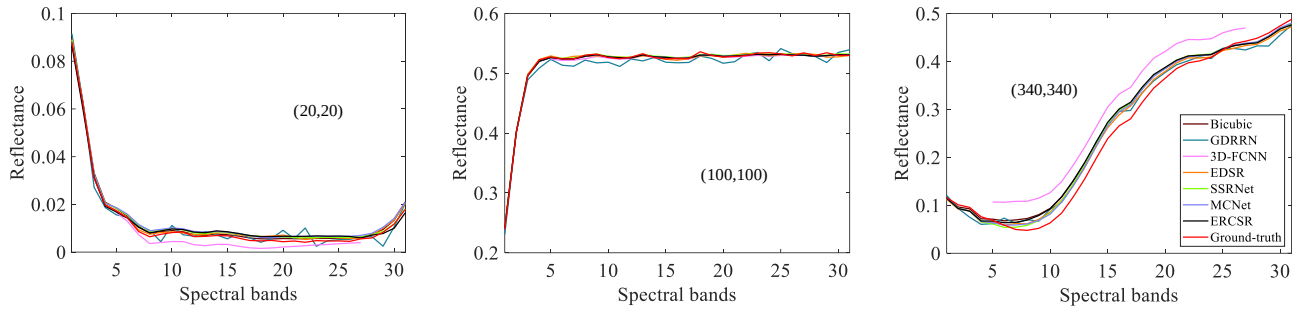


Fig. 7. Visual comparison of spectral distortion for image *fake_and_real_lemons* at pixel position (20, 20), (100, 100), and (340, 340) on CAVE dataset.

TABLE VIII
QUANTITATIVE EVALUATION OF STATE-OF-THE-ART SR ALGORITHMS BY AVERAGE PSNR/SSIM/SAM FOR DIFFERENT SCALE FACTORS ON HARVARD DATASET. THE RED AND BLUE INDICATE THE BEST AND SECOND BEST PERFORMANCE, RESPECTIVELY.

Scale factor	Evalution metric	Bicubic	GDRRN [24]	3D-FCNN [28]	EDSR [36]	SSRNet [33]	MCNet [30]	ERCSR (ours)
×2	PSNR ↑	42.833	44.213	44.454	45.480	46.247	46.263	46.372
	SSIM ↑	0.9711	0.9775	0.9778	0.9824	0.9825	0.9827	0.9832
	SAM ↓	2.023	2.278	1.894	1.921	1.884	1.883	1.875
×3	PSNR ↑	39.441	40.912	40.585	41.674	42.650	42.681	42.783
	SSIM ↑	0.9411	0.9523	0.9480	0.9592	0.9626	0.9627	0.9633
	SAM ↓	2.325	2.623	2.239	2.380	2.209	2.214	2.180
×4	PSNR ↑	37.227	38.596	38.143	39.175	40.001	40.081	40.211
	SSIM ↑	0.9122	0.9259	0.9188	0.9324	0.9365	0.9367	0.9374
	SAM ↓	2.531	2.794	2.363	2.560	2.412	2.410	2.384

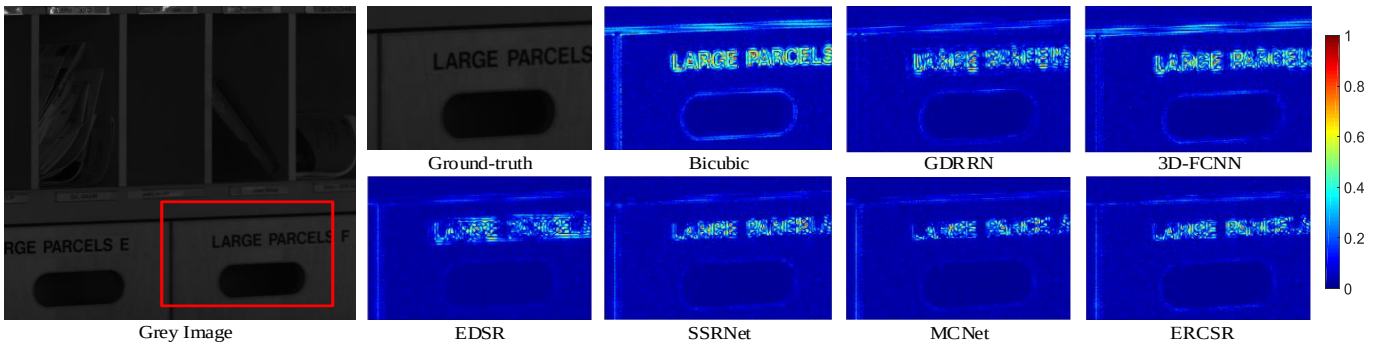


Fig. 8. Absolute error map comparisons for image *imgd5* at 680 nm on Harvard dataset.

the effectiveness of the proposed ERCSR for different scale factors. Note that unlike other two datasets, the Pavia Centre dataset is hyperspectral remote sensing dataset.

1) CAVE Dataset: Table VI depicts the quantitative evaluations of the state-of-the-art SR algorithms for different scale factors on CAVE dataset. One can observe that our ERCSR outperform best performance than other competitors in three metrics. Among these algorithms, GDRRN and EDSR adopt 2D convolution to conduct SR task, while other several deep learning methods utilize 3D convolution. We can find that the overall performance of the network using 3D convolution is better than that of using 2D convolution network. This is due to the fact that 3D convolution can effectively utilize the spectrum, thus improving feature exploration. Note that since the output size of the 3D-FCNN is changed, the results are actually accurate. Compared with the second best algorithm, MCNet, our method attains excellent performance. In particular, the proposed model is higher than MCNet in terms of three metrics for scale factor $\times 4$ (+0.130 dB, +0.0007, and -0.026). Moreover, the number of parameters of designed ERCSR is also lower than that of MCNet, which is shown in Tbale VII.

We adopt qualitative way to further analyze our method. To simply do comparison, in our paper, only one reconstructed hyperspectral image in this dataset for scale factor $\times 4$ is displayed, and the image in one band is presented. Since the ground-truth is grey image, to show some edge information clearly, the absolute error map between ground-truth and reconstructed hyperspectral image is employed. Fig. 6 provides the visual results of several algorithms. We can notice that our proposed ERCSR generates shallow edges or no edges in some regions, while other algorithms exhibit obvious texture information. Finally, we also visualize the spectral distortion of reconstructed image by selecting three pixels (see Fig. 7). As mentioned above, the output size of 3D-FCNN becomes small, so only part of the band is depicted. Among these methods, they obtain almost the same results at different pixels. Moreover, all of them can maintain the spectrum of the reconstructed image.

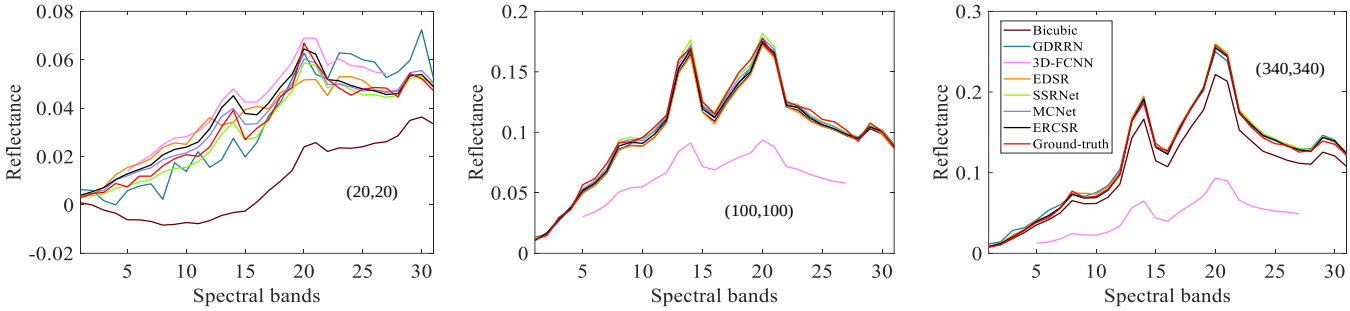


Fig. 9. Visual comparison of spectral distortion for image *imgd5* at pixel position (20, 20), (100, 100), and (340, 340) on Harvard dataset.

TABLE IX
QUANTITATIVE EVALUATION OF STATE-OF-THE-ART SR ALGORITHMS BY AVERAGE PSNR/SSIM/SAM FOR DIFFERENT SCALE FACTORS ON PAVIA CENTRE DATASET. THE RED AND BLUE INDICATE THE BEST AND SECOND BEST PERFORMANCE, RESPECTIVELY.

Scale factor	Evalution metric	Bicubic	GDRRN [24]	3D-FCNN [28]	EDSR [36]	SSRNet [33]	MCNet [30]	ERCSR (ours)
×2	PSNR ↑	32.383	33.762	34.540	35.515	35.397	35.404	35.422
	SSIM ↑	0.9020	0.9280	0.9427	0.9500	0.9493	0.9493	0.9498
	SAM ↓	4.059	4.317	3.472	3.437	3.448	3.445	3.435
×3	PSNR ↑	29.343	30.369	30.519	31.222	31.214	31.203	31.230
	SSIM ↑	0.7982	0.8407	0.8503	0.8701	0.8685	0.8679	0.8690
	SAM ↓	5.060	5.662	4.239	4.708	4.659	4.689	4.650
×4	PSNR ↑	27.672	27.988	28.494	28.684	28.902	28.907	28.912
	SSIM ↑	0.7080	0.7301	0.7621	0.7730	0.7802	0.7796	0.7786
	SAM ↓	5.776	5.988	4.950	5.654	5.577	5.587	5.534

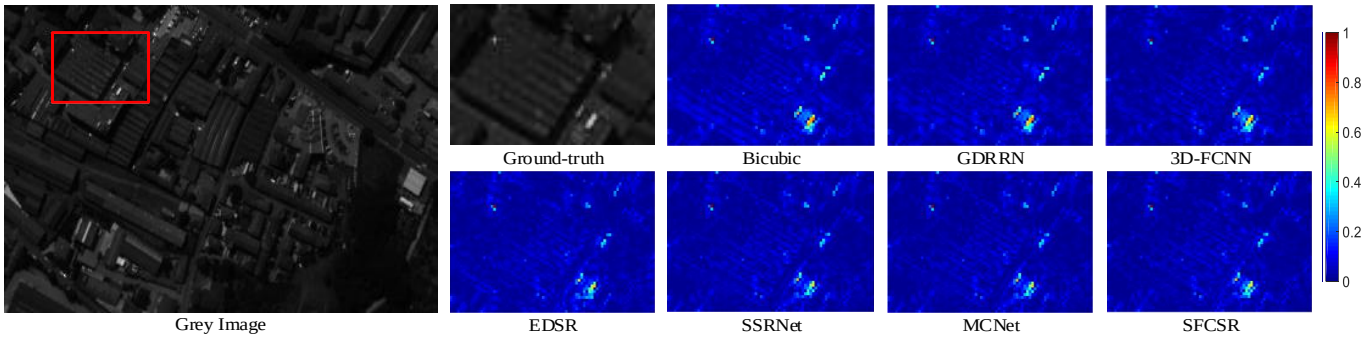


Fig. 10. Absolute error map comparisons for 20-th band image on Pavia Centre dataset.

2) Harvard Dataset: Similar to the results on CAVE dataset, Table VIII also displays that our ERCSR can outperform well in every aspect. With respect to GDRRN algorithm, it achieves the lowest results due to the design of a shallower network. Due to the deep design, the EDSR, which also adopts 2D convolution, attains very good performance. Nevertheless, there is still some performance gap compared with those networks that exploit 3D convolution. As for SSRNet algorithm, it focuses too much on the spectral dimension and ignores the spatial dimension, which leads to the poor performance. Although MCNet builds the network aiming at the problems existing in SSRNet, the results obtained by the two are basically the same. From our point of view, it is caused by the failure to make full use of the output of 2D unit. Considering this limitation, our proposed model utilizes two types of units to perform alternately, thus obtaining significant superiority.

We also illustrate a visual example for scale factor ×4, which is presented in Fig. 8. We can find that our method can also obtain relatively low error values, and the edge information of some objects is particularly weak. Unlike the visual result for image *fake_and_real_lemons*, the spectral curves yield relatively large distinctions, particularly for 3D-FCNN and Bicubic (see Fig. 9). Although there is a certain deviation between the spectral curves acquired by these methods and the corresponding ground-truth (such as at pixel position (20, 20)), in most cases, the spectral curves generated by ERCSR are closer to the ground-truth. It demonstrates the proposed ERCSR attains better spectral fidelity, which makes it possible to use spectral band analysis in practical application.

3) Pavia Centre Dataset: The above datasets are not hyperspectral remote sensing data. To comprehensively illustrate the effectiveness of the proposed algorithm, the Pavia University dataset that belongs to hyperspectral remote sensing image is utilized for further evaluation. As shown in Table IX, compared with other two datasets, our approach is not that superior in this dataset. As can seen from this table, almost all the models achieve good result only on some

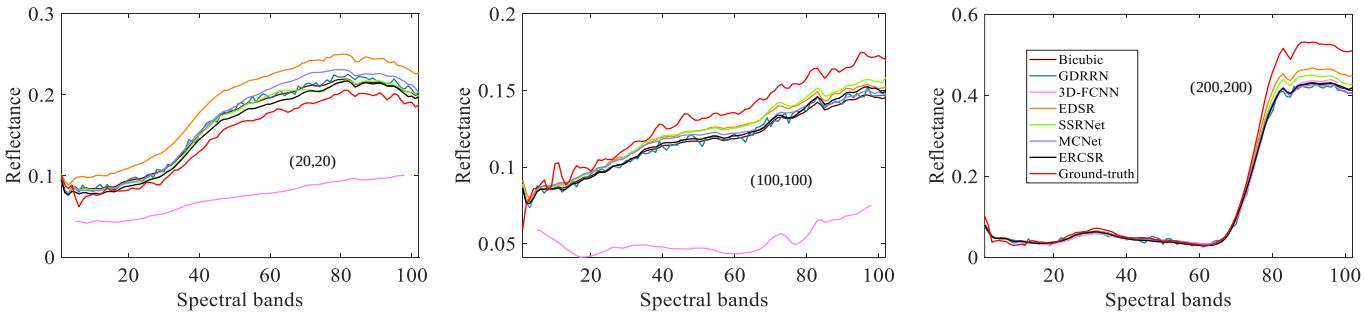


Fig. 11. Visual comparison of spectral distortion at pixel position (20, 20), (100, 100), and (200, 200) on Pavia Centre dataset.

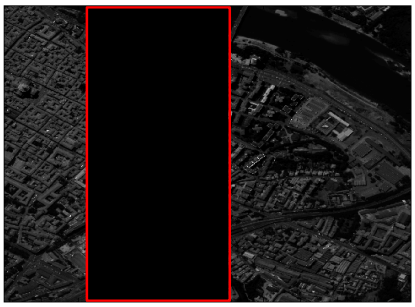


Fig. 12. Illustration of Pavia Centre hyperspectral image. The red box indicates that the pixel value in the area is 0.

evaluation metrics of some scale factors. The main reason for this phenomenon is that there is no data in the part of the image, i.e., the pixel value is 0 (see Fig. 12). It prevents these algorithms from effectively optimizing the network. As a whole, EDSR has better performance on small scale factor. As for large scale factor, the number of parameters of our method is approximately the same as that of EDSR (see Table VII), but the proposed ERCSR can get satisfactory results. Similarly, it produces comparable performance in visual comparisons about absolute error map and spectral distortion in Figs. 10 and 11. According to the above investigations, that is enough to verify our approach presents remarkable performance both in quantity and quality.

V. Conclusion

In our paper, we develop a new structure for hyperspectral image super-resolution by exploring the relationship between 2D/3D convolution (ERCSR). To learn more spatial information when spectral features are extracted, our method alternately employs 2D and 3D units to analyze during reconstruction. It greatly reduces the complexity of feature learning within the 3D unit, so as to improve the efficiency of the model. Different from previous work, our method adopts a novel way, namely split adjacent spatial and spectral convolution (SAEC), to parallelly study the features between spectrum and other directions. Extensive experiments on widely used benchmark datasets demonstrate that our ERCSR attains better performance against the state-of-the-art methods in terms of PSNR, SSIM, and SAM. In particular, as for Harvard dataset, compared with the second best method, the accuracies of the proposed ERCSR in PSNR and SSIM increase by 0.130 dB and 0.0367 for scale factor $\times 4$, and its SAM decreased by 0.026.

In the future, we intend to extend the proposed method from two aspects. As for enhanced hybrid convolution module (E-HCM), how to combine 2D/3D convolution is better? We can use network architecture search (NAS) by setting rules to get the optimal structure. Moreover, in our proposed SAEC, is it better to do the addition operation directly or other similar concatenation operations? All of the above can be optimized to further optimize the network structure and thus improve the performance of the whole model.

References

[1] F. F. Sabins, "Remote sensing for mineral exploration," *Ore Geol. Rev.*, vol. 14, no. 3-4, pp. 157–183, 1999.

[2] J. Lin, N. T. Clancy, J. Qi, Y. Hu, T. Tatla, D. Stoyanov, L. Maier-Hein, and D. S. Elson, "Dual-modality endoscopic probe for tissue surface shape reconstruction and hyperspectral imaging enabled by deep neural networks," *Med. Image Anal.*, vol. 48, pp. 162–176, 2018.

[3] W. Sun, G. Yang, J. Peng, and Q. Du, "Lateral-slice sparse tensor robust principal component analysis for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 107–111, 2020.

[4] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 57, no. 2, pp. 911–923, 2019.

[5] Q. Wang, Z. Yuan, and X. Li, "GETNET: A general end-to-end two-dimensional cnn framework for hyperspectral image change detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 57, no. 1, pp. 3–13, 2019.

[6] X. Zhang, G. Wen, and W. Dai, "A tensor decomposition-based anomaly detection algorithm for hyperspectral image," *IEEE Trans. Geosci. Remote Sensing*, vol. 54, no. 10, pp. 5801–5820, 2016.

[7] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral image super-resolution using deep feature matrix factorizationk," *IEEE Trans. Geosci. Remote Sensing*, vol. 57, no. 8, pp. 6055–6067, 2019.

[8] W. Dong, F. Fu, G. Shi, X. Gao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, 2016.

[9] 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, 2019.

[10] R. Dian and S. Li, "Hyperspectral image super-resolution via subspace-based low tensor multi-rank regularization," *IEEE Trans. Image Process.*, vol. 28, no. 10, pp. 5135–5146, 2019.

[11] C. Yi, Y. Zhao, and J. C. Chan, "Hyperspectral image super-resolution based on spatial and spectral correlation fusion," *IEEE Trans. Geosci. Remote Sensing*, vol. 56, no. 7, pp. 4165–4177, 2018.

[12] W. Wan, W. Guo, H. Huang, and J. Liu, "Nonnegative and nonlocal sparse tensor factorization-based hyperspectral image super-resolution," *IEEE Trans. Geosci. Remote Sensing*, pp. 1–11, 2020.

[13] S. Bajwa, P. Bajcsy, P. Groves, and L. Tian, "Hyperspectral image data mining for band selection in agricultural applications," *Transactions of the Asae*, vol. 47, no. 3, pp. 895–907, 2004.

[14] L. Paluchowski and P. Walczykowski, "Preliminary hyperspectral band selection for difficult object detection," pp. 1–4, 2009.

[15] S. S. Hashjin, A. D. Boloorani, S. Khazai, and A. A. Kakroodi, "Selecting optimal bands for sub-pixel target detection in hyperspectral images based on implanting synthetic targets," *IET Image Process.*, vol. 13, no. 2, pp. 323–331, 2018.

[16] W. Xie, J. Lei, J. Yang, Y. Li, Q. Du, and Z. Li, "Deep latent spectral representation learning-based hyperspectral band selection for target detection," *IEEE Trans. Geosci. Remote Sensing*, vol. 58, no. 3, pp. 2015–2026, 2019.

[17] H. Kwon and Y. Tai, "RGB-guided hyperspectral image upsampling," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 307–315.

[18] N. Akhtar, F. Shafait, and A. S. Mian, "Bayesian sparse representation for hyperspectral image super resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 3631–3640.

[19] "Camera spectral response," https://www.maxmax.com/, Accessed June 12, 2020.

[20] Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Hyperspectral image super-resolution with optimized RGB guidance," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11661–11670.

[21] S. Anwar, S. Khan, and N. Barnes, "A deep journey into super-resolution: A survey," *arXiv preprint arXiv:1904.07523*, 2019.

[22] J. Jia, L. Ji, Y. Zhao, and X. Geng, "Hyperspectral image super-resolution with spectral–spatial network," in *Int. J. Remote Sens.*, 2018, pp. 7806–7829.

[23] P. V. Arun, K. M. Buddhiraju, A. Porwal, and J. Chanussot, "CNN-based super-resolution of hyperspectral images," *IEEE Trans. Geosci. Remote Sensing*, pp. 1–16, 2020.

[24] Y. Li, L. Zhang, C. Ding, W. Wei, and Y. Zhang, "Single hyperspectral image super-resolution with grouped deep recursive residual network," in *Proc. IEEE Int. Conf. Multimed. Big Data*, 2018, pp. 1–4.

[25] R. 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, 2017.

[26] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, 2017.

[27] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3147–3155.

[28] 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, pp. 1139, 2017.

[29] J. Yang, Y. Zhao, J. C. Chan, and L. Xiao, "A multi-scale wavelet 3d-cnn for hyperspectral image super-resolution," *Remote Sens.*, vol. 11, no. 13, pp. 1557, 2019.

[30] Q. Li, Q. Wang, and X. Li, "Mixed 2D/3D convolutional network for hyperspectral image super-resolution," *Remote Sens.*, vol. 12, no. 10, pp. 1660, 2020.

[31] J. Li, R. Cui, B. Li, R. Song, Y. Li, Y. Dai, and Q. Du, "Hyperspectral image super-resolution by band attention through adversarial learning," *IEEE Trans. Geosci. Remote Sensing*, pp. 1–15, 2020.

[32] J. Li, R. Cui, Y. Li, B. Li, Q. Du, and C. Ge, "Multitemporal hyperspectral image super-resolution through 3d generative adversarial network," in *Proc. Int. Workshop Anal. Multitemporal Remote Sens. Images*, 2019, pp. 1–4.

[33] Q. Wang, Q. Li, and X. Li, "Spatial-spectral residual network for hyperspectral image super-resolution.," *arXiv: 2001.04609*, 2020.

[34] R. Jiang, X. Li, A. Gao, L. Li, H. Meng, S. Yue, and L. Zhang, "Learning spectral and spatial features based on generative adversarial network for hyperspectral image super-resolution," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process Proc.* IEEE, 2019, pp. 3161–3164.

[35] Q. Wang, Q. Li, and X. Li, "Hyperspectral image super-resolution using spectrum and feature context," *IEEE Trans. Ind. Electron.*, 2020.

[36] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1132–1140.

[37] Q. Li, Q. Wang, and X. Li, "An efficient clustering method for hyperspectral optimal band selection via shared nearest neighbor," *Remote Sens.*, vol. 11, no. 3, pp. 350, 2019.

[38] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sensing*, 2020.

[39] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, 2016.

[40] S. Xie, C. Sun, J. Huang, Z. Tu, and K. Murphy, "Rethinking spatiotemporal feature learning: Speed-accuracy trade-offs in video classification," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 318–335.

[41] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, "Image super-resolution by neural texture transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 7982–7991.

Qiang Li received the B.E. degree in measurement & control technology and instrument from Xi'an University of Posts and Telecommunications, Xi'an, China, in 2015, and the M.S. degree in communication and transportation engineering from Chang'an University, Xi'an, China, in 2018.

He is currently pursuing the Ph.D. degree with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning. His research interests include hyperspectral image processing and computer vision.

Qi Wang (M'15-SM'15) received the B.E. degree in automation and the Ph.D. degree in pattern recognition and intelligent systems from the University of Science and Technology of China, Hefei, China, in 2005 and 2010, respectively.

He is currently a Professor with the School of Computer Science and the Center for Optical Imagery Analysis and Learning, Northwestern Polytechnical University, Xi'an, China. His research interests include computer vision and pattern recognition.

Xuelong Li (M'02-SM'07-F'12) is a Full Professor with the School of Computer Science and the Center for OPTical IMagery Analysis and Learning (OPTIMAL), Northwestern Polytechnical University, Xi'an, China.