Hyperspectral Image Super-Resolution Using Spectrum and Feature Context

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

***Abstract*—Deep learning-based hyperspectral image super-resolution (SR) methods have achieved great success recently. However, most methods utilize 2D or 3D convolution to explore features, and rarely combine the two types of convolution to design networks. Moreover, when the model only contains 3D convolution, almost all the methods take all the bands of hyperspectral image as input to analyze, which requires more memory footprint. To address these issues, we explore a new structure for hyperspectral image super-resolution using spectrum and feature context (SFCSR). Inspired by the high similarity among adjacent bands, we design a dual-channel network through 2D and 3D convolution to jointly exploit the information from both single band and adjacent bands, which is different from previous works. Under the connection of depth split (DS), it can effectively share spatial information so as to improve the learning ability of 2D spatial domain. Besides, our method introduces the features extracted from previous band, which contributes to the complementarity of information and simplifies the network structure. Through feature context fusion (FCF), it significantly enhances the performance of the algorithm. Extensive evaluations and comparisons on three public datasets demonstrate that our approach produces the state-of-the-art results over the existing approaches.**

***Index Terms*—Hyperspectral image, super-resolution (SR), context information, depth split (DS), dual-channel fusion**

I. Introduction

HYPERSPECTAL imaging system collects surface information from tens, hundreds or even thousands of continuous bands to obtain hyperspectral image. It can provide rich spatial and spectral information about objects, and reflect the subtle spectral characteristics of the measured objects in detail, greatly improving the ability of detection, recognition and other tasks. Therefore, it is widely used in some industrial fields, such as welding defect detection in aerospace and shipbuilding [1], optical gas imaging [2], mineral exploration via remote sensing satellite [3], etc.

In hyperspectral imaging system, spatial and spectral resolution are two important indices to measure image quality [4]. Spatial resolution means that the sensor can distinguish the minimum target size. It reflects the level of detail of the spatial information of the image. With respect to spectral resolution, it refers to the resolution of the spectral detail information of the ground object in the image. Hyperspectral image usually improve the spectral resolution of the image at the cost of reducing the spatial resolution, i.e., dividing more bands within the spectral range of the image. Due to the electronic hardware of hyperspectral imaging system, it results in that the spatial resolution of the hyperspectral image is lower than that of the natural or multispectral image [5]. In practice, the possible objects are usually detected by specific spectrum, and then further accurate analysis is made by spatial information. However, the spatial resolution has become the main constraint of hyperspectral image in some situations, where high accuracy is required. Under the limitation of the production process of hyperspectral imaging equipment, researchers [6]–[10] use signal processing technology to improve the spatial resolution of hyperspectral image, i.e., hyperspectral image super-resolution (SR). It can get better image content so as to obtain more accurate results in subsequent interpretation and application.

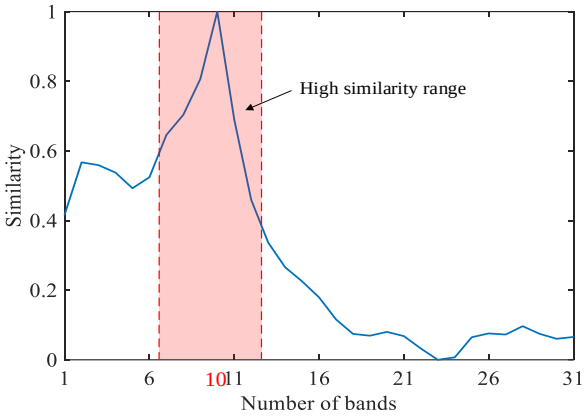


Fig. 1. The similarity for current band (band index 10) with other bands for image *balloons* on CAVE dataset.

Due to the powerful representation ability of convolution neural networks (CNNs), deep learning-based methods provide great success in some tasks [11], [12]. Recently, the SR using it to deal with natural image has been widely studied [13]–[16]. Typical networks contain SRCNN [15], SRGAN [17], etc. Those methods provide an important reference for the research of hyperspectral image SR. For hyperspectral image SR, previous works adopt either 3D or 2D convolu-

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: crabwq@gmail.com, liqmges@gmail.com, xuelong_li@nwpu.edu.cn) (*Corresponding author: Xuelong Li.*).

tion to explore the features. When the model only contains 2D convolution [18], [19], it is helpful to explore spatial information, but it ignores spectral knowledge, which leads to poor reconstruction performance. As for the networks using 3D convolution [20], [21], it is based on the fact that the spectral information can improve the spatial resolution of hyperspectral image. Compared with the networks with 2D convolution, these algorithms significantly improve the performance. However, there are some limitations of these networks [18]–[21]. Those models often take all the bands of hyperspectral image as input to analyze. Obviously, these bands are processed synchronously, which requires more GPU memory footprint, making it impossible for the network to design deeper under limited hardware conditions. Importantly, it ignores a key problem that high similarity exits among adjacent bands [22], [23], which is shown in Fig. 1. From this perspective, it is not necessary to analyze all the bands for hyperspectral image SR task. With respect to the methods containing 2D or 3D convolution, most of them utilize one type of convolution to explore features, and rarely combine the two to design networks.

As we just mentioned, the adjacent bands have high correlation. If the designed network can effectively utilize adjacent bands when reconstructing single band, it is more beneficial to the complementarity of information. Besides, most methods rarely combine these two types of convolution to design networks. Motivated by these, in our paper, we propose a new structure for hyperspectral image SR via spectrum and feature context (SFCSR). Different from previous works, our method employs dual-channel CNNs to extract the information from single band and adjacent bands, respectively, which is based on the spectrum context. Additionally, the local fusion is implemented to retain informative features learned from both channels. After achieving the previous band reconstruction, some features are transferred to the reconstruction task of the current band (feature context). Experiments on public datasets demonstrate that the proposed SFCSR outperforms the state-of-the-art methods across all datasets and scale factors. In summary, our main contributions are follows:

1) A novel approach to input the network is developed. Different from previous works, we apply the current band and its two adjacent bands to perform single band SR. Through the way of recurrent, the reconstruction of the hyperspectral image is achieved.
2) The dual-channel network contained 2D CNN and 3D CNN is designed to jointly exploit the information of single band and adjacent bands by depth split (DS). Compared with the single 3D CNN, it can effectively share spatial information so as to enhance the learning ability of 2D spatial domain.
3) The features that have been extracted from the previous band SR task are fed into the network of current band. It not only facilitates the complementarity of information, but also simplifies the network structure so that there is no need to design more branches.

The remainder of this paper is organized as follows: Section II describes existing hyperspectral image SR with CNNs. Section III introduces our proposed SFCSR in details. Then, experiments on several datasets are performed to evaluate our method in Section IV. Finally, Section V gives the conclusion.

II. Related Work

A. CNNs Using 2D Convolution

Recently, the development of deep learning has brought a new opportunity for the research of SR reconstruction, its performance is significantly better than the traditional methods. According to the characteristics of hyperspectral image, the scholars use the SR methods of natural image [24] for reference to design the model by 2D convolution [25]–[27]. For example, Li *et al.* [28] propose grouped deep recursive residual network (GDRRN). He *et al.* [29] propose a hypersepctral image SR method inspired by the deep Laplacian pyramid network [30]. Those designed networks cannot effectively utilize the information of spectral dimension, thus obtaining poor performance. Unlike natural image SR, hyperspectral image SR should avoid spectral distortion when reconstructing. Considering this issue, Li *et al.* [19] design a deep spectral difference convolutional neural network (SDCNN). After the hyperspectral image is reconstructed, the post-processing is adopted to avoid spectral distortion. Yuan *et al.* [18] use the knowledge in natural image to reconstruct hyperspectral image through transfer learning. Similarly, the spectral distortion is alleviated by the post-processing, i.e., collaborative nonnegative matrix factorization. This type of methods described above mainly contain two steps. It can be noticed that the spatial resolution of hyperspectral image may be changed in second step, while maintaining the spectral information. Later, Jiang *et al.* [27] propose a group convolution and progressive upsampling framework for hyperspectral image SR. The algorithm achieves the promising results. Nevertheless, due to the progressive architecture of the network, i.e., two-step upsampling, the model is unable to display smaller scale factors, such as $\times 2$ and $\times 3$.

B. CNNs Using 3D Convolution

Hyperspectral image not only has spatial information, but also contains abundant spectral information. For hyperspectral image SR, 2D convolution cannot effectively extract spectral information. To make use of the rich information of spectrum, there are many literatures on the hyperspectral image SR using 3D convolution [21], [31]–[33].

Mei *et al.* [21] first adopt 3D convolution to extract the spatial and spectral features simultaneously. Although the approach attains better performance compared with the network using 2D convolution, all convolutions are not padded during reconstruction. It leads to that the actual output of the network is smaller than the input. Yang *et al.* [33] design multi-scale wavelet 3D convolution neural network to achieve SR task. However, the model requires image data pre-processing and post-processing. Inspired by generative adversarial networks (GAN) [17], [34], the researchers develop hyperspectral image SR using GAN. For example, Li *et al.* [35] design 3D-GAN to conduct SR by regular 3D convolution. Generally speaking, GAN-based models are not easy to train. Additionally,

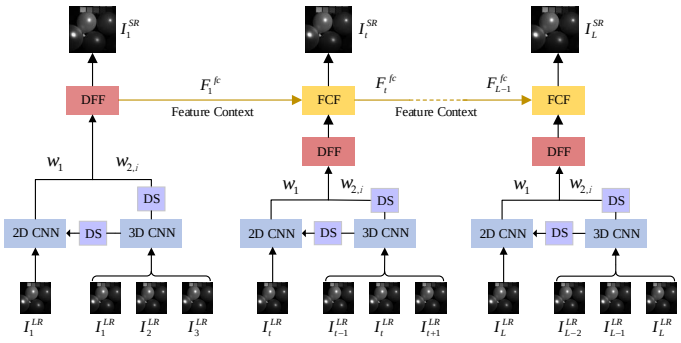


Fig. 2. Overall architecture of our proposed SFCSR. First, the single band is fed into 2D CNN to exploit the spatial information. The adjacent bands are input to 3D CNN to extract the spatial and spectral features simultaneously. Under depth split (DS), the network adaptively preserves the valid information learned from both channels with the help of dual-channel feature fusion (DFF). Then, the features learned in previous band are transferred to the reconstruction of the current band. Through feature context fusion (FCF), a single estimated band is obtained. After many such steps, we finally obtain the reconstructed hyperspectral image.

compared with network using 2D convolution, the number of parameters in this network increase significantly, which makes it impossible for the network to be designed deeper. Jiang *et al.* [36] propose 2D-1D GAN architecture that consists of two parts: spatial and spectral sub-network. The approach effectively reduces the number of parameter, but it cannot explore spectral and spatial information simultaneously. Later, Li *et al.* [20], [37] propose dual 1D-2D spatial-spectral CNN. The model uses 1D and 2D convolution to study the spatial and spectral features, respectively, and fuses them by changing the size of the feature maps. Similarly, Wang *et al.* [32] employ spatial and temporal separable 3D convolution to analyze spatial and spectral features. However, when extracting spectral information, it lacks the exploration of more spatial information. Considering this problem, Li *et al.* [9] propose mixed 2D/3D convolutional network for hyperspectral image SR. This model shares spatial information for 2D/3D feature maps so that the network can obtain more spatial exploration. The approach obtains the best performance compared with the existing works.

As we mentioned earlier, the main purpose of using spectral dimension information is to improve the performance about spatial SR. Moreover, as is displayed in Fig. 1, the current band is only closely correlated with adjacent bands. However, all the above methods take all the bands of hyperspectral image as input to analyze. From this point of view, it is not necessary to analyze all the bands synchronously. Thus, in our paper, we only utilize the current band and its two adjacent bands to implement SR task, which is different from the input mode of the existing algorithms.

III. Proposed Method

In this section, we detail the architecture of the proposed SFCSR, including network structure, dual-channel network, dual-channel feature fusion, and feature context.

A. Network Structure

Now we present the overall architecture of our SFCSR, which is shown in Fig 2. Different from previous approaches, our method deals with each band separately, and only the adjacent bands are utilized. The overall structure contains 2D CNN, 3D CNN, dual-channel feature fusion (DFF), and feature context fusion (FCF). Suppose I_t^{LR} represents the t-th band that currently needs to be reconstructed, and I_{t-1}^{LR} and I_{t+1}^{LR} denotes two adjacent bands for t-th band. With respect to I_t^{LR}, we only consider its spatial information, so only one band I_t^{LR} is fed into 2D CNN. Since the hyperspectral image has a remarkable characteristic that high similarity exits among adjacent bands [22], [23], [38], three bands I_t^{LR}, I_{t-1}^{LR}, and I_{t+1}^{LR} are input into 3D CNN. By doing so, it can make full use of the complementary information in the spatial dimension of the adjacent bands, as well as the information of the spectral dimension. Then, with the help of DFF, the output of two channels are concatenated by setting the weight so that the network can preserve the hierarchical features studied by both channels, which is defined as

$$F=\begin{cases} C[w_1 * f_X(I_1^{LR}), w_2 * f_Y(I_1^{LR}, I_2^{LR}, I_3^{LR})], & t=1 \\ C[w_1 * f_X(I_t^{LR}), w_2 * f_Y(I_{t-1}^{LR}, I_t^{LR}, I_{t+1}^{LR})], & 1<t<L \\ C[w_1 * f_X(I_L^{LR}), w_2 * f_Y(I_{L-2}^{LR}, I_{L-1}^{LR}, I_L^{LR})], & t=L \end{cases} \tag{1}$$

where C represents concatenation operation, $f_X(\cdot)$ and $f_Y(\cdot)$ denote the operation of two channels, w_1 and w_2 are the weight of two channels, and L is the total number of band in hyperspectral image. Since previous band in space is extremely correlated with current band, to obtain features from previous band I_{t-1}^{LR}, the features F_{t-1}^{fc} are introduced to the reconstruction of current band I_t^{LR}, which is based on feature context. Finally, the model achieves single estimated band I_t^{SR}. Note that when performing the first band task, there is no FCF module in the network, only the DFF module, i. e.,

$$I_t^{SR}=\begin{cases} upscale(f_{DFF}(F)), & t=1 \\ upscale(f_{FCF}(f_{DFF}(F), F_{t-1}^{fc})), & t\neq 1 \end{cases} \tag{2}$$

where f_{DFF} and f_{FCF} are DFF and FCF operation, respectively, and $upscale(\cdot)$ is the function for upsampling.

Through multiple such steps, we acquire the reconstructed hyperspectral image. Unlike the existing networks, the proposed SFCSR applies several bands to achieve SR task. Importantly, the spectrum and feature context are employed, thus significantly improving the performance of the algorithm.

B. Dual-Channel Network

Previous works adopt either 2D or 3D convolution to explore the features, there are few ways to design the network by combining two types of convolution. Inspired by the existence of high similarity among adjacent bands, we introduce dual-channel network by combining two types of convolution, which is shown in Fig. 3. The architecture contains 3D CNN and 2D CNN. Next, we will present the details about the proposed modules.

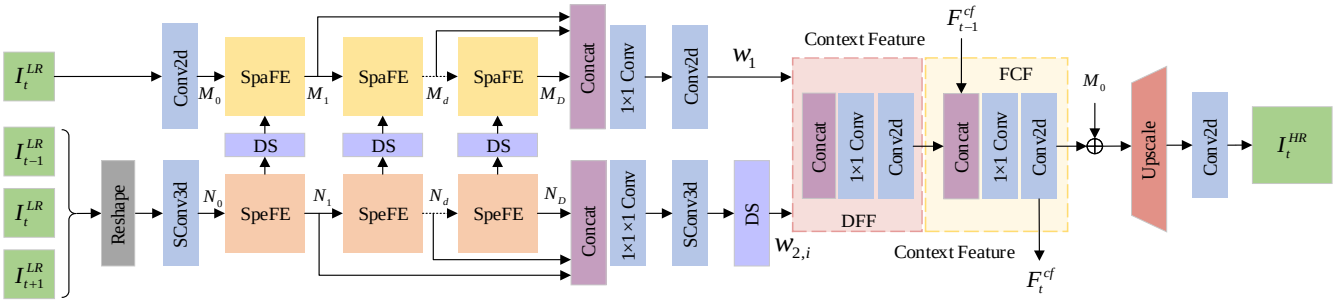


Fig. 3. Overall architecture of the proposed dual-channel network.

1) 3D CNN: The hyperspectral image has a remarkable characteristic that high similarity exits among adjacent bands [22], [23]. To improve the spatial resolution, we exploit two adjacent bands in hyperspectral image. As mentioned in Section II-B, 3D convolution can analyze information other than spatial dimensions. Thus, to produce spatial and spectral features synchronously, in our paper, we design a CNN using 3D convolution. Unlike 2D convolution, 3D convolution operation is conducted by convolving a 3D kernel with feature maps. If the regular 3D convolution is applied, obviously, the parameters of the network will increase significantly. Therefore, we adopt a separable 3D convolution instead of a regular 3D convolution, i.e, the filter $k \times k \times k$ is modified as $k \times 1 \times 1$ and $1 \times k \times k$, which has been proven to achieve the same effect [32], [39]–[41].

Now we introduce 3D CNN. To apply 3D convolution, we expand three bands (I_t^{LR}, I_{t-1}^{LR}, and I_{t+1}^{LR}) into four dimensions ($1 \times 3 \times W \times H$) at the beginning of the network, where W and H denote the width and height of band. By a separable 3D convolution (it is represented by *SConv3d* in Fig. 3), the shallow features N_0 about three bands are acquired. Similar to 2D CNN, several spatial and spectral feature extraction (SpeFE) modules are employed to further explore the potential information. Under skip connections, the output N_D is formulated as

$$N_D = Y_D(Y_{D-1}(...Y_1(N_0) + N_0...) + N_0) + N_0, \quad (3)$$

where $Y_d(\cdot)$ denotes the operation of the d-th SpeFE module. Likewise, the output from different SpeFE modules are concatenated. Moreover, we also add a convolution layer with the filter $1 \times 1 \times 1$ to control valid flow, which is then followed by a separable 3D convolution.

In each SpeFE module (see Fig. 4(a)), the spatial and spectral features are extracted using the filter $1 \times k \times k$ and $k \times 1 \times 1$, respectively. Through an addition operation, the information between them is effectively fused, which is conducive to the spatial exploration for each band. Furthermore, we also add the local residual connection in the module. Similarly, the initial features N_0 are attached to the end of each SpeFE module. Compared with 2D CNN, 3D CNN can use the spectral and spatial information from two adjacent bands to promote the reconstruction performance of single band. Meanwhile, it enables the network to effectively alleviate the spectral distortion of the reconstructed image.

2) 2D CNN: For hyperspectral image SR, its aim is to improve the spatial resolution while keeping the spectral resolution unchanged. Therefore, we do 2D convolution analysis of one of the channels to improve the spatial resolution. Specifically, the shallow features M_0 of single band I_t^{LR} are first extracted by 2D convolution. Then, this data is entered into spatial feature extraction (SpaFE) module. Under D SpaFE modules and skip connections, the output M_D is represented as

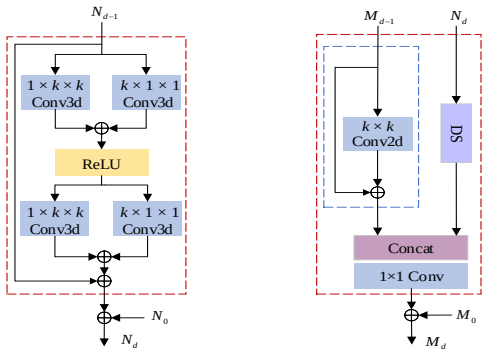
$$M_D = X_D(X_{D-1}(...X_1(M_0) + M_0...) + M_0) + M_0, \quad (4)$$

where $X_d(\cdot)$ stands for the operation of the d-th SpaFE module. After generating spatial features with SpaFE modules, our sub-network further fuses the different hierarchical states from all preceding layers in a concatenation way, which improves the network representation ability. Additionally, we add a convolution layer with the filter 1×1, then use 2D convolution to study more potential information after concatenation.

For natural image SR, the main module of the network mainly contains two parts: several convolutions and a residual connection, which has been verified to achieve better performance [24], [42], [43]. Thus, we refer to the module from the natural image SR to design SpaFE module. As is shown in Fig. 4(b), we utilize one convolution layer with the filter $k \times k$ to design the module. Moreover, the local residual connection is attached to the module, which greatly alleviates the burden on SpaFE module. To enhance the learning ability of space domain, the current output from d-th SpeFE module is input to d-th SpaFE module by depth split (DS) that is introduced in Section III-C. Similarity, we also add a convolution layer with the filter 1×1 to achieve dimension reduction. Since the shallow network can retain more edge or texture, the features M_0 are fed into the end of each SpaFE module. Such local skip connection allow more abundant low-frequency information to be bypassed during training, which enhances the performance of the entire network.

C. Dual-channel Feature Fusion

After learning the features from single band and adjacent bands, we need to fuse them. Since the size of the output about two channels is different, we reshape one of them before fusion. In our paper, we choose the results of 3D CNN to execute by DS before fusion. With respect to DS, suppose the size of the current output is $N \times C \times 3 \times W \times H$ after conducting



(a) SpeFE module (b) SpaFE module

Fig. 4. Architecture of the proposed module.

SConv3d operation for 3D CNN, where N is the batch size, and C represents the number of channel. Through splitting, the original output of the 3D CNN is changed into three feature maps with the size of $N \times C \times W \times H$. This can be seen as processing each band of hyperspectral image separately. Then, under the action of DFF, the features of the two channels are fused by setting weights (w_1 and $w_{2,i}$, $i = 1, 2, 3$), which is shown in the light red region in Fig. 3. By doing so, it can make full use of the spectral and spatial information of adjacent bands, which is more beneficial to the reconstruction of single band. Finally, we add a 2D convolution operation in the end of DFF.

D. Feature Context

In the existing works [18], [28], there is almost no use of a single band to implement hyperspectral image SR. As mentioned earlier, there is a strong correlation between adjacent bands. Inspired by this, we can take advantage of context to improve the performance. In our paper, we propose for the first time to utilize the output information after DFF to establish context, which is different from the existing works. As shown in the light yellow area in Fig. 3, the features F_{t-1}^{fc} from previous band I_{t-1}^{LR} are chosen and introduced with the features of current band in a concatenation way. Besides, we set two weights (w_3 and w_4) to dynamically adjust its weight, i.e.,

$$C[w_3 * f_{DFF}(F), w_4 * F_{t-1}^{cf}]. \tag{5}$$

By this way, it helps the network to modulate the discriminative representations so as to focus on more useful information. Overall, this approach mainly has two advantages. On the one hand, due to the high correlation between adjacent bands, it is more favorable to the complementarity of information by combining with other bands. On the other hand, by storing the features extracted from previous band, the network structure is simplified so that there is no need to design more branches. As for the influence of this part, we will prove its effectiveness in Section IV-B3.

IV. Experiments

In this section, to verify the effectiveness of the proposed method, we will elaborate from the following aspects. First, we briefly introduce the benchmark datasets, implementation details. Then, we verify the effectiveness of each module through model analysis. Finally, we assess the performance of our SFCSR by comparisons to the state-of-the-art methods.

TABLE I
Analysis of the influence of the number of the module D in two sub-networks on the performance.

Evaluation metric	3	4	5	6	7
PSNR	44.937	45.144	45.300	45.340	45.401
SSIM	0.9736	0.9738	0.9739	0.9739	0.9739
SAM	2.233	2.225	2.217	2.216	2.215

A. Implementation Details

In our work, the CAVE [44], Harvard [45], and Foster [46] dataset are employed to evaluate the algorithm. Since they are collected by different hyperspectral cameras, which results in no identical attributes between datasets. This makes it necessary to train and test each dataset separately, which is different the natural image SR. Therefore, in our paper, we randomly select 80% samples from dataset for training and the rest for testing. Since the training data contains fewer hyperspectral images, we augment them by randomly selecting 24 patches. Each patch is scaled (1, 0.75, and 0.5), and these patches are rotated by 90° and horizontally flipped, respectively. Then, we downsample these patches by bicubic interpolation to yield low-resolution hyperspectral images with the size of $L \times 32 \times 32$.

For the training stage, we adopt $L1$ loss function to learn the model. With respect to the parameters of the network, the parameter k of the filter for all convolution layers is set to 3, and the number of their filters is set to 64. We adopt ADAM optimizer ($\beta_1 = 0.9, \beta_2 = 0.999$) to optimize the designed model. The learning rate of all layers is initialized 10^{-4}, which decreases by a half at every 35 epochs. Due to the limitation of hardware conditions, in our work, we only select the top left 512×512 for each test image to evaluate the performance of the algorithm using PSNR, SSIM, and SAM. Our method is performed using the PyTorch framework with NVIDIA GeForce GTX 1080 GPU.

B. Model Analysis

1) Study of Module D: In our proposed network structure, the SpaFE and SpeFE module of 2D CNN and 3D CNN affects the performance of the network. Thus, in this section, we set different numbers of modules (3 to 7) to analyze the impact for scale factor $\times$ 2 on CAVE dataset. The results are shown in Table I. One can observe that with the increase of parameter D, the overall results of the network is improved, especially for PSNR. It indicates that more modules are helpful to fully mine the potential information of hyperspectral image, which can further enhance the performance of the algorithm. But when D is set to 6 and 7, the growth rates of PSNR and SAM are obviously smaller, and the value of SSIM keeps unchanged. In our opinion, there are two main reasons for this phenomenon. The one is the increase of network parameters caused by the

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

Type	PSNR	SSIM	SAM	Params
regular 3D convolution	45.290	0.9736	2.228	1698k
separable 3D convolution	45.300	0.9739	2.218	1085k

TABLE III
ABLATION STUDY FOR DIFFERENT COMPONENTS.

Commponent	Different combinations of components					
2D CNN	√	×	√	√	√	√
3D CNN	×	√	√	√	√	√
DS between channels	×	×	×	√	×	√
DFF	×	×	√	√	√	√
FCF	×	×	×	×	√	√
PSNR	42.223	40.448	45.138	45.237	45.276	45.300
SSIM	0.9699	0.9710	0.9737	0.9738	0.9739	0.9739
SAM	2.568	5.996	2.230	2.223	2.219	2.217

use of more 3D convolutions in network. The other is that the network becomes deeper. All these may make the network difficult to train. Therefore, we empirically set the number of SpaFE and SpeFE module in dual-channel network to 5 in the following experiment.

2) Study of Different 3D Convolution Types: In 3D CNN sub-network, the separable 3D convolution instead of regular 3D convolution is adopted to extract the spatial and spectral features from hyperspectral image. To verify the impact of the two convolution types on the network, in this section, we replace the separate 3D convolution and use the regular 3D convolution to conduct SR task for scale factor $\times$ 2 on CAVE dataset. The comparison results are shown in Table II. We can notice that the performance of regular 3D convolution is lower than that of separable 3D convolution. The main reason for this phenomenon is the lack of the fusion between space and spectrum for regular 3D convolution. In our paper, the spatial and spectral features are extracted in each SpeFE module using the filter $1 \times k \times k$ and $k \times 1 \times 1$, respectively. Through an addition operation, the information between them is effectively fused, which is conducive to the spatial exploration for each band. Besides, the separable 3D convolution can greatly reduce parameters. For regular 3D convolution, it obviously can't achieve the same result as separable 3D convolution, such as non-fusion. Through the above description, it can be concluded that separable 3D convolution is more conducive to the exploration of information.

3) Ablation Study: The proposed network is mainly composed of four components: 2D CNN, 3D CNN, DS, DFF, and FCF. They determine the network structure. In this section, we analyze the impact of each component on the performance of the network by setting different combinations. To simply do fair comparison, the ablation investigation is implemented for scale factor $\times 2$ on CAVE dataset.

Table III shows the ablation study about the these components. Concretely, the network only first contains 2D CNN or 3D CNN without other components. Their performance is worst, especially when the network only contains 3D CNN. The main reason for this phenomenon is that there is no global skip connection. This may make it difficult for effective shallow information to be transmitted to deeper layers of the network. Note that DFF is a module that combines two channels of feature maps, and DS is used to transmit spectral and spatial information from 3D CNN to 2D CNN. They can work when both 2D CNN and 3D CNN has. Therefore, we does not show the ablation experiment about these. Then, both 2D CNN and 3D CNN exist, the overall performance obviously enhances by adding DFF into the network. It verifies DFF is an indispensable part for improving performance. When FCF is added to the network with 2D CNN, 3D CNN, and DFF, the values of three evaluation metrics achieve better results than the above combinations. It exhibits FCF plays an important role in exploring hyperspectral image features. With respect to DS between dual-channel network, we notice that it also show a significant improvement after joining the network. Finally, all components are attached to the network. It can be seen that the results of all components are better than the performance of other combinations of components in three aspects, which reveals the effectiveness and benefits of the proposed each component.

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

In this section, we conduct a comprehensive comparison between the proposed SFCSR and the seven existing approaches, including Bicubic and several deep learning algorithms. In our work, the three datasets are adopted to investigate the performance of the proposed method by quantitative evaluation and qualitative comparison. Note that all convolution operations for 3D-FCNN are not padding during reconstruction, thus its output changes the size of reconstructed hyperspectral image. With respect to EDSR algorithm, the number of residual blocks and feature maps are 16 and 64, respectively. As for SSPSR method, the hyperparameters, i.e., the spectral band number of each group and the overlap between neighboring groups, are set to 4 and 1 on CAVE and Harvard dataset. These two hyperparameters on Foster dataset are fixed to 5 and 1. Moreover, due to the progressive upsampling architecture, the model is unable to display smaller scale factors, such as $\times 2$ and $\times 3$.

1) Quantitative Evaluation: Table IV provides the results of evaluation of our SFCSR by three commonly-used evaluation metrics for different scale factors. As shown in this table, our method outperforms other competitors on CAVE dataset. Specifically, the Bicubic produces the worst performance, particularly for PSNR and SSIM. As for GDRRN approach, it extracts the features of hyperspectral image via 2D convolution. Although its results are better than Bicubic in two aspects, it is obviously low than that of other algorithms conducted by 2D convolution. The main reason for this phenomenon is that the algorithm adds SAM to the loss function. The purpose is to avoid reconstructed spectral distortions. As a result, the loss function can not effectively optimize the network, thus obtaining poor results. As we mentioned above, 3D-FCNN changes the size of the reconstructed hyperspectral image, so its results are actually not very accurate. Among

TABLE IV
QUANTITATIVE EVALUATION OF STATE-OF-THE-ART SR ALGORITHMS BY AVERAGE PSNR/SSIM/SAM, PARAMETERS, AND RUNNING TIME FOR DIFFERENT SCALE FACTORS ON THREE DATASETS. THE RED AND BLUE INDICATE THE BEST AND SECOND BEST PERFORMANCE, RESPECTIVELY.

Scale	Method	Params	CAVE		Harvard		Foster	
			PSNR / SSIM / SAM	Time	PSNR / SSIM / SAM	Time	PSNR / SSIM / SAM	Time
×2	Bicubic	-	40.762 / 0.9623 / 2.665	-	42.833 / 0.9711 / 2.023	-	55.155 / 0.9881 / 4.391	-
	GDRRN [28]	219k	41.667 / 0.9651 / 3.842	0.193	44.213 / 0.9775 / 2.278	0.132	53.527 / 0.9963 / 5.634	0.217
	3D-FCNN [21]	39k	43.154 / 0.9686 / 2.305	1.138	44.454 / 0.9778 / 1.894	0.858	60.242 / 0.9987 / 5.271	1.201
	EDSR [14]	1404k	43.869 / 0.9734 / 2.636	0.098	45.480 / 0.9824 / 1.921	0.040	57.371 / 0.9978 / 5.753	0.104
	SSPSR [27]	10515k	-	-	-	-	-	-
	SSRNet [32]	830k	44.991 / 0.9737 / 2.261	1.231	46.247 / 0.9825 / 1.884	1.051	58.852 / 0.9987 / 4.064	1.179
	MCNet [9]	1928k	45.102 / 0.9738 / 2.241	2.096	46.263 / 0.9827 / 1.883	2.061	58.878 / 0.9988 / 4.061	2.322
	SFCSR (ours)	1085k	45.300 / 0.9739 / 2.217	2.280	46.342 / 0.9830 / 1.880	2.017	58.859 / 0.9988 / 4.052	2.764
×3	Bicubic	-	37.532 / 0.9325 / 3.522	-	39.441 / 0.9411 / 2.325	-	50.964 / 0.9943 / 5.357	-
	GDRRN [28]	219k	38.834 / 0.9401 / 4.537	0.183	40.912 / 0.9523 / 2.623	0.119	50.464 / 0.9926 / 6.833	0.193
	3D-FCNN [21]	39k	40.219 / 0.9453 / 2.930	1.128	40.585 / 0.9480 / 2.239	0.846	55.551 / 0.9958 / 6.303	1.189
	EDSR [14]	1589k	40.533 / 0.9512 / 3.175	0.091	41.674 / 0.9592 / 2.380	0.040	52.983 / 0.9956 / 7.716	0.100
	SSPSR [27]	10515k	-	-	-	-	-	-
	SSRNet [32]	941k	40.896 / 0.9524 / 2.814	0.951	42.650 / 0.9626 / 2.209	0.867	54.937 / 0.9967 / 5.134	0.923
	MCNet [9]	2039k	41.031 / 0.9526 / 2.809	1.330	42.681 / 0.9627 / 2.214	1.277	55.017 / 0.9970 / 5.126	1.313
	SFCSR (ours)	1270k	41.198 / 0.9524 / 2.794	1.246	42.778 / 0.9632 / 2.203	1.034	54.575 / 0.9967 / 5.046	1.462
×4	Bicubic	-	35.755 / 0.9071 / 3.944	-	37.227 / 0.9122 / 2.531	-	48.281 / 0.9880 / 5.993	-
	GDRRN [28]	219k	36.959 / 0.9166 / 5.168	0.169	38.596 / 0.9259 / 2.794	0.114	47.836 / 0.9877 / 7.696	0.196
	3D-FCNN [21]	39k	37.626 / 0.9195 / 3.360	1.114	38.143 / 0.9188 / 2.363	0.854	52.188 / 0.9918 / 7.798	0.145
	EDSR [14]	1552k	38.587 / 0.9292 / 3.804	0.084	39.175 / 0.9324 / 2.560	0.038	50.362 / 0.9915 / 7.103	0.097
	SSPSR [27]	12875k	38.366 / 0.9727 / 3.484	0.205	39.293 / 0.9333 / 2.448	0.154	50.320 / 0.9923 / 5.995	0.233
	SSRNet [32]	1076k	38.944 / 0.9312 / 3.297	0.717	40.001 / 0.9365 / 2.412	0.680	52.210 / 0.9939 / 5.702	0.712
	MCNet [9]	2174k	39.026 / 0.9319 / 3.292	0.913	40.081 / 0.9367 / 2.410	0.841	52.225 / 0.9941 / 5.685	0.987
	SFCSR (ours)	1233k	39.192 / 0.9321 / 3.221	0.864	40.077 / 0.9373 / 2.407	0.792	52.215 / 0.9939 / 5.618	1.036
×8	Bicubic	-	31.805 / 0.8485 / 5.291	-	33.275 / 0.8518 / 2.884	-	43.117 / 0.9595 / 7.855	-
	GDRRN [28]	219k	32.825 / 0.8518 / 7.582	0.161	33.711 / 0.8557 / 3.668	0.107	42.148 / 0.9548 / 13.219	0.182
	3D-FCNN [21]	39k	32.956 / 0.8600 / 4.384	0.987	33.363 / 0.8448 / 2.726	0.835	45.634 / 0.9667 / 8.827	1.016
	EDSR [14]	1700k	31.554 / 0.8233 / 10.670	0.050	34.068 / 0.8598 / 3.051	0.050	43.669 / 0.9655 / 9.374	0.112
	SSPSR [27]	15236k	32.043 / 0.8423 / 7.190	0.198	34.435 / 0.8676 / 2.899	0.149	44.100 / 0.9693 / 8.071	0.222
	SSRNet [32]	1863k	35.271 / 0.8832 / 4.439	0.532	35.042 / 0.8719 / 3.065	0.479	45.123 / 0.9730 / 7.512	0.579
	MCNet [9]	2961k	35.320 / 0.8833 / 4.423	0.575	34.927 / 0.8732 / 2.957	0.527	45.045 / 0.9726 / 7.516	0.641
	SFCSR (ours)	1381k	35.294 / 0.8828 / 4.378	0.652	35.097 / 0.8737 / 2.911	0.610	45.083 / 0.9721 / 7.506	0.701

TABLE V
SPECTRAL DISTORTION COMPARISONS IN THREE SCENES BY SAM. THE RED AND BLUE INDICATE THE BEST AND SECOND BEST PERFORMANCE, RESPECTIVELY.

Dataset	Image	Pixel Position	Bicubic	GDRRN	3D-FCNN	EDSR	SSPSR	SSRNet	MCNet	SFCSR
CAVE	fake_and_real_lemons	(10,70)	0.624	3.274	0.951	0.807	0.700	0.614	0.612	0.580
		(350,320)	0.724	1.316	1.315	0.730	0.585	0.611	0.541	0.533
Harvard	imgd5	(45,20)	1.679	1.778	1.814	1.731	1.691	1.687	1.690	1.662
		(260, 150)	3.620	3.524	17.343	3.103	2.703	2.901	2.331	2.324
Foster	Lillies_Closeup	(90,100)	14.739	21.881	42.050	32.226	21.840	17.090	13.577	14.209
		(200,50)	5.470	7.040	12.591	7.085	5.690	5.645	5.690	5.426

these approaches utilizing 2D convolution to conduct SR task, SSPSR gets the best performance. Unfortunately, the network requires too many parameters. Furthermore, the model is unable to display smaller scale factors due to the progressive upsampling. Except for the three methods (GDRRN, EDSR, and SSPSR), the other competitors explore the features by 3D convolution. As shown in the table, in the case of a small number of parameters, the results of the networks designed with 3D convolution is significantly higher than that of the networks with 2D convolution. It helps to obtain low spectral distortion results. This is mainly due to the full analysis of spectral information. Compared with previous five methods, SSRNet algorithm attains higher spectral fidelity. Among the above state-of-the-art approaches, they utilize 2D or 3D convolution to explore features, without combining the two types of convolution to design networks. Although MCNet integrates two convolution ways to design the network, the output results of some modules are not used effectively. As a result, the performance of the model is lower than that of our proposed SFCSR in the aspect of three evaluation metrics and parameters, except for running time.

Similarly, the proposed SFCSR overall obtains the best results than other methods on Harvard dataset across scale factors. As depicted in this table, the performance of the algorithms using 3D convolution is obviously higher than that of the algorithms using 2D convolution. Unlike on CAVE dataset, SSRNet, MCNet, and SFCSR achieve approximately the results, but our method still attains the best performance

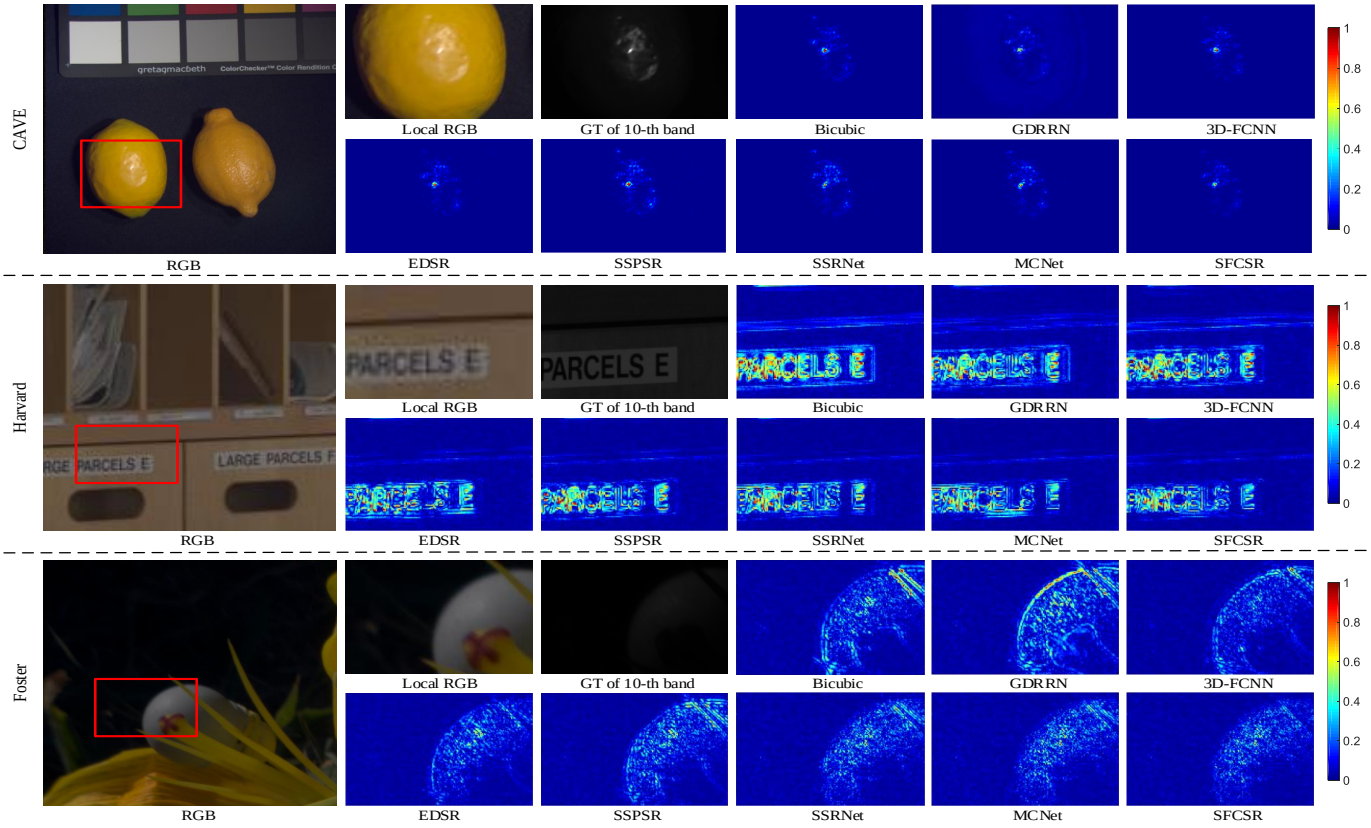


Fig. 5. Absolute error map comparisons for reconstructed hyperspectral image. The first to third lines represent the visual results of the image fake_and_real_lemons on CAVE dataset, *imgd5* on Harvard dataset, and *Lillies_Closeup* on Foster dataset. The bluer the absolute error map is, the better the visual quality is.

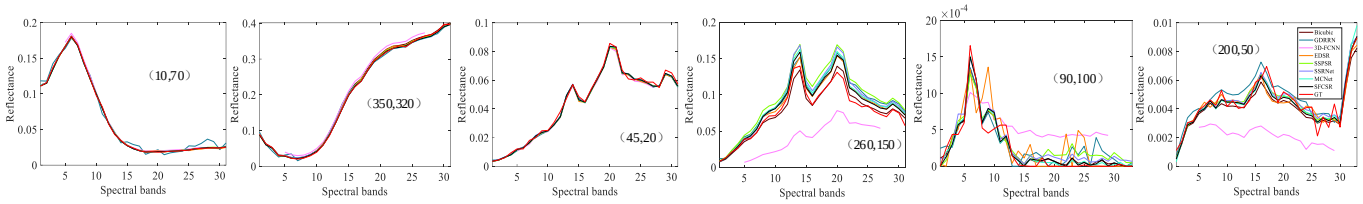


Fig. 6. Visual comparison of spectral distortion by selecting two pixel positions for each image. The images *fake_and_real_lemons*, *imgd5*, and *Lillies_Closeup* are displayed from left to right, respectively.

against the state-of-the-art methods. As for Foster dataset, our method achieves comparable results. Although MCNet obtains the superior results, its network requires more parameters. Moreover, the running time of the proposed SFCSR is not much different from that of the method for large scale factors. Through the above description, it demonstrates our approach can achieve the excellent performance over each dimension across all datasets and scale factors. Meanwhile, it proves the designed dual-channel network can effectively enhance the learning ability of 2D spatial domain by DS module. It solves the lack of the combination of two kinds of convolution in the existing methods. In addition, several bands are used as input to reduce GPU memory footprint.

2) Qualitative Evaluation: To further analyze the proposed SFCSR with other state-of-the-art SR approaches in qualitative way, several visual examples for scale factor $\times 4$ are shown in Fig. 5. Note that only one hyperspectral image is displayed in each dataset, and the 10-th band in corresponding image is presented. It can be seen from figures that the ground-truth is grey. To observe the difference between the reconstructed hyperspectral image and ground-truth more clearly, the absolute error map between them is given. As shown in these figures, our method describes the result of low absolute error. In particular, the proposed SFCSR exhibits shallow edges or no edges in some regions. It indicates our method can achieve best spatial reconstruction against other approaches, which is consistent with our analysis in Table IV. It also reveals that the two types of convolution are more helpful to enhance the ability of feature representation.

The above visualization just verifies the spatial reconstruction results of the image. In our work, we also show the spectral distortion of the reconstructed image for the above three images. Specifically, two pixel positions are selected to analyze the spectrum difference between the reconstructed

hyperspectral image and the ground-truth by calculating SAM. As shown in Fig. 6 and Table V, the values of SAM in our method are better than that of other algorithms in most cases. It illustrates our SFCSR obtains higher spectral fidelity. To sum up, the proposed SFCSR overall outperforms the state-of-the-art methods across all datasets and scale factors in terms of spatial reconstruction and spectral fidelity.

V. Conclusions

In this paper, we develop a new deep learning-based framework for hyperspectral image SR. Different from previous work, our method employs dual-channel CNNs to analyze the information from both single and adjacent bands, which mainly utilizes spectrum context. To facilitate the complementarity of information and simplify the network structure, the features from previous band are combined with that of current band, achieving the use of context. Extensive experiments on benchmark datasets demonstrate that the proposed model outperforms the state-of-the-art methods both quantitatively and qualitatively.

As for the proposed method, the designed dual-channel network can effectively share spatial information by information exchange to enhance the learning ability of 2D spatial domain. Moreover, the proposed method only utilizes the current band and its two adjacent bands as input. These novel approaches can inspire us to design networks in this way in the future work. Of course, the proposed method also needs to be improved in two aspects. First, the output of each SpeFE module is input to SpaFE module by depth split (DS). There is a lack of information exchange from SpaFE module to SpeFE module. Thus, we can improve performance by exchanging the information in each module of the two channels through DS operation. Second, the spatial resolution of some bands of hyperspectral image is poor, due to the influence of noise and other problems. Therefore, the band with rich spatial information (key band) can be used to improve the reconstruction performance of the band with poor resolution (non-key band).

References

[1] P. Sassi, P. Tripicchio, and C. A. Avizzano, "A smart monitoring system for automatic welding defect detection," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9641–9650, 2019.

[2] R. Olbrycht and M. Kaluza, "Optical gas imaging with uncooled thermal imaging camera - impact of warm filters and elevated background temperature," *IEEE Trans. Ind. Electron.*, 2019.

[3] B. Thai and G. Healey, "Invariant subpixel material detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 40, no. 3, pp. 599–608, 2002.

[4] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral image super-resolution by spectral mixture analysis and spatial-spectral group sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1250–1254, 2016.

[5] Y. Qu, H. Qi, and C. Kwan, "Unsupervised sparse dirichlet-net for hyperspectral image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 2511–2520, 2018.

[6] W. Xie, X. Jia, Y. Li, and J. Lei, "Hyperspectral image super-resolution using deep feature matrix factorization," *IEEE Trans. Geosci. Remote Sensing*, pp. 1–13, 2019.

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

[8] N. Akhtar, F. Shafait, and A. Mian, "Hierarchical beta process with gaussian process prior for hyperspectral image super resolution," in *Proc. Eur. Conf. Comput. Vis.*, pp. 103–120, 2016.

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

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

[11] N. Lv, C. Chen, T. Qiu, and A. K. Sangaiah, "Deep learning and superpixel feature extraction based on contractive autoencoder for change detection in SAR images," *IEEE Trans. Ind. Inform.*, vol. 14, no. 12, pp. 5530–5538, 2018.

[12] A. Kumar, S. K. Singh, S. Saxena, K. Lakshmanan, A. K. Sangaiah, H. Chauhan, S. Shrivastava, and R. K. Singh, "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," *Inf. Sci.*, vol. 508, pp. 405–421, 2020.

[13] W. Jia, Y. Zhao, R. Wang, S. Li, H. Min, and X. Liu, "Are recent sisr techniques suitable for industrial applications at low magnification?" *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9828–9836, 2019.

[14] 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.*, pp. 1132–1140, 2017.

[15] 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.

[16] J. Ma, X. Wang, and J. Jiang, "Image super-resolution via dense discriminative network," *IEEE Trans. Ind. Electron.*, vol. PP, no. 99, pp. 1–1, 2019.

[17] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, and Z. Wang, "Photo-realistic single image super-resolution using a generative adversarial network,," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 105–114, 2017.

[18] 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.

[19] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, 2017.

[20] J. Li, R. Cui, B. Li, R. Song, Y. Li, and Q. Du, "Hyperspectral image super-resolution with 1D-2D attentional convolutional neural network," *Remote Sens.*, vol. 11, no. 23, p. 2859, 2019.

[21] S. Mei, X. Yuan, J. Ji, Y. Zhang, S. Wan, and Q. Du, "Hyperspectral image spatial super-resolution via 3d full convolutional neural network," *Remote Sens.*, vol. 9, no. 11, p. 1139, 2017.

[22] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, 2019.

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

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

[25] J. Hu, Y. Li, and W. Xie, "Hyperspectral image super-resolution by spectral difference learning and spatial error correction," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1825–1829, 2017.

[26] 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.*, pp. 11 661–11 670, 2019.

[27] J. Jiang, H. Sun, X. Liu, and J. Ma, "Learning spatial-spectral prior for super-resolution of hyperspectral imagery," *IEEE Trans. Comput. Imaging*, vol. 6, pp. 1082–1096, 2020.

[28] 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*, pp. 1–4, 2018.

[29] Z. He and L. Lin, "Hyperspectral image super-resolution inspired by deep laplacian pyramid network," *Remote Sens.*, vol. 10, no. 12, p. 1939, 2018.

[30] W. Lai, J. Huang, N. Ahuja, and M. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 5835–5843, 2017.

[31] W. Liu and J. Lee, "An efficient residual learning neural network for hyperspectral image superresolution," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1240–1253, 2019.

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

[33] 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, p. 1557, 2019.
[34] M. Salvaris, D. Dean, and W. H. Tok, "Generative adversarial networks," *arXiv: Machine Learning*, pp. 187–208, 2018.
[35] 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*, pp. 1–4. IEEE, 2019.
[36] 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.*, pp. 3161–3164. IEEE, 2019.
[37] J. Li, R. Cui, B. Li, Y. Li, S. Mei, and Q. Du, "Dual 1D-2D spatial-spectral cnn for hyperspectral image super-resolution," in *Proc. Int. Geosci. Remote Sen.s Symp.*, pp. 3113–3116. IEEE, 2019.
[38] 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, p. 350, 2019.
[39] 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.*, pp. 318–335, 2018.
[40] Z. Hu, Y. Hu, J. Liu, B. Wu, D. Han, and T. R. Kurfess, "3D separable convolutional neural network for dynamic hand gesture recognition," *Neurocomputing*, vol. 318, pp. 151–161, 2018.
[41] Z. Qiu, T. Yao, and T. Mei, "Learning spatio-temporal representation with pseudo-3D residual networks," in *Proc. IEEE Conf. on Compu. Vis.*, pp. 5534–5542, 2017.
[42] Y. Zuo, Q. Wu, Y. Fang, P. An, L. Huang, and Z. Chen, "Multi-scale frequency reconstruction for guided depth map super-resolution via deep residual network," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 2, pp. 297–306, 2020.
[43] R. Lan, L. Sun, Z. Liu, H. Lu, Z. Su, C. Pang, and X. Luo, "Cascading and enhanced residual networks for accurate single-image super-resolution," *IEEE Trans. Syst. Man Cybern. -Syst.*, pp. 1–11, 2020.
[44] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar, "Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2241–2253, 2010.
[45] A. Chakrabarti and T. Zickler, "Statistics of real-world hyperspectral images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 193–200, 2011.
[46] S. M.Nascimento, K. Amanoand, and H. D. Foster, "Spatial distributions of local illumination color in natural scenes," *Vision Res.*, vol. 120, pp. 39–44, 2016.

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.

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.

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.