

Received April 19, 2020, accepted May 3, 2020, date of publication May 6, 2020, date of current version May 20, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992862

A Frequency-Separated 3D-CNN for Hyperspectral Image Super-Resolution

LIGUO WANG^{1,2}, (Member, IEEE), TIANYI BI², AND YAO SHI²

¹College of Information and Communications Engineering, Dalian Nationalities University, Dalian 116600, China

²College of Information and Communications Engineering, Harbin Engineering University, Harbin 150001, China

Corresponding author: Ligu Wang (wangliguo@hrbeu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61675051.

ABSTRACT Considering the limitations such as cost, it is of great significance to use super-resolution methods to improve image spatial quality in the field of hyperspectral remote sensing. Due to little dependence on auxiliary information which is difficult to obtain, i.e., multispectral images and natural images, methods based on single-frame are generally considered to have good flexibility and application value. In this paper, a three-dimensional convolutional neural network with three branches combined with an analytical method is proposed, achieving better SR quality and suppressing spectral distortion as well. Firstly, the wavelet transformation is introduced to decompose the hyperspectral image into a variety frequency of components effectively and reversibly. Then, these components are fed into different three-dimensional convolutional branches respectively. Finally, hyperspectral images with high resolution are obtained by dimension amplification, detail reconstruction and inverse wavelet transformation. The presence of frequency separation and the architecture of our model having different branches designed according to frequency make it better than comparable approaches. The method proposed in this paper not only combines the high efficiency of analytical method and the flexibility of neural network, but inhibits the influence of spectral distortion as well. Compared with the state-of-art methods on real space-based hyperspectral image datasets, the effectiveness of the proposed method is demonstrated.

INDEX TERMS Hyperspectral remote sensing, super-resolution, wavelet transformation, neural network.

I. INTRODUCTION

With the development of computer technology and electronic information technology, the remote sensing information of natural resources has become one of the core needs of human beings. Remote sensing technology has a wide range of applications, such as earth resources exploration [1], marine protection, forest fire prevention [2]. However, by considering the limitation such as payload, transmission bandwidth, and power consumption of equipment, hyperspectral image (HSI) usually retain higher spectral resolution at the cost of lower spatial resolution, resulting in the contradiction between the demand for spatial quality and the performance of HSI acquisition platforms, promoting the development of HSI technology, especially the technology of super-resolution (SR). In recent years, a lot of SR methods for HSI have been

proposed, which makes this direction an important research focus in the field.

There are many methods to improve the spatial resolution of HSI, and they can be classified by whether to introduce auxiliary information or not. Multispectral images (MSI) and panchromatic images are widely used as auxiliary information which has high spatial resolution [3] and can be utilized as prior knowledge to improve the spatial quality of HSI. Pan-sharpening methods [4]–[6] utilize high resolution (HR) panchromatic images on the same region as auxiliary information to improve the spatial resolution of HSI by fusion methods. Combined with deep neural network, Masi *et al.* [7] proposed a convolution-based sharpening network PNN. In order to retain spatial details, Yang *et al.* [8] proposed a neural network based on high-pass filtering domain. By means of alternate non-negative matrix factorization of the spectral data, Yokoya *et al.* [9] obtained HR HSI by reconstruction of the MSI abundance and hyperspectral

The associate editor coordinating the review of this manuscript and approving it for publication was Wenming Cao¹.

end member. Zhu *et al.* [10] utilized HR multispectral data to learn space dictionaries and reconstructed sparse coding to get HR HSI. Akhtar *et al.* [11] utilized the low-resolution (LR) HSI to learn the spectral dictionary, and reconstructed the multispectral abundance to get HR HSI. Wei *et al.* [12] realized a fusion data between MSI and HSI by sparse representation. Simoes *et al.* [13] proposed a convex vector total variation method based on subspace regularization to realize the data fusion. Zhang *et al.* [14] realized the data fusion by the popular low-rank clustering structure. Yang *et al.* [15] proposed a double-branch convolutional neural network to fuse MSI and HSI data. Xie *et al.* [16] utilized a deep neural network based on low-rank prior to fuse MSI and HSI. Due to the limitation of the acquisition area and timeliness, as well as much difficulty in reprocessing of the data, it is very difficult to acquire the auxiliary information, which limits the application of such methods.

The single-frame based SR methods HSI obtain the prior information from the LR HSI to predict the HR HSI, which is very flexible because no auxiliary information is required. Zhao *et al.* [17] proposed a hyperspectral SR method of sparse representation and non-local correlation regularization. Li *et al.* [18] proposed a group sparse representation based HSI SR method utilizing spatial domain and spectral domain autocorrelation. Wang *et al.* [19] proposed a non-local low-rank tensor approximate SR method based on tensor representation. Yuan *et al.* [20] utilized convolutional neural network to realize preliminary SR level by level, and fine-tuned the preliminary results by collaborative matrix decomposition. Li *et al.* [21] proposed the spectral difference network (SDCNN) to learn the spectral difference mapping between LR and HR. Hu *et al.* [22] improved the SDCNN and integrated it with the spatial error correction model to correct the human error in HR HSI. Considering that three-dimensional convolution can simultaneously extract the spatial spectrum joint features of HSI, Mei *et al.* [23] proposed a three-dimensional hyperspectral super-resolution method, which utilizes three-dimensional feature representation, to learn the mappings between LR and HR HSI.

As the research on neural networks develops, the cognition of neural network is more comprehensive. Firstly, the analytical method relies on prior information such as statistical law, formula derivation, human experience, etc., whereas the neural network accumulates prior information from the data set in the recursive process and corrects the extraction and utilization of features. Therefore, the extraction and utilization of information are realized in different ways. Besides, a proper combination of two methods will help to obtain a more flexible, robust and quick model. Additionally, the idea of processing of different frequency components specifically has gradually drawn more and more attention in the past two years. Chen *et al.* [24] proposed a plug-and-play octave convolutional layer to process different frequency components in the image specifically. It is capable of improving the performance of the state-of-art models and reduce network parameters by means of directly replacing the convolutional

layers to octave convolutional layers. Yang *et al.* [25] proposed a three-dimensional convolutional network based on two-dimensional wavelet packet (MW-3D-CNN). In this model, the shallow features are extracted from the LR HSI cubes, and then the wavelet packet coefficients of the features are extracted. The author especially emphasizes the internal relationship between different wavelet packet coefficients.

In this paper, it is proposed that a frequency-separated three-dimensional neural network combined with an analytics-combined (FS-3DCNN) is designed for HSI SR. Firstly, by means of wavelet transformation the LR HSI is decomposed into three groups of wavelet coefficients, according to the frequency similarity. The three convolutional branches of FS-3DCNN designed according to the frequency can suppress the spectral distortion while protecting the high-frequency information. The feature cubes are up-sampled by three-dimensional deconvolution and reconstructed in detail by three-dimensional convolution. Finally, HR HSI is obtained by inverse wavelet transformation. Experimental results on four real HSI datasets show that compared with the state-of-art methods, our method has certain improvement in image quality and spectral distortion.

The structure of this paper is as follows. In Section II, the related works about HSI SR are presented. In Section III, the details of FS-3DCNN and valuation metrics are presented. In Section IV, the experimental settings and results are presented. In Section V, we propose additional analysis and discussion about the experiment. In Section VI, the conclusion as well as our expectation of potential tend of single-frame HSI SR methods is presented.

II. RELATED WORKS

In this section, the ideas and researches related to this paper are briefly reviewed. Firstly, we introduce the theory and the research state of SR methods based on wavelet, and list some typical SR methods of wavelet transformation. Then, we enumerate some representative single-frame based convolutional neural network (CNN) solutions for SR.

A. SR METHODS BASED ON WAVELET

It is accepted by the public that time-frequency analysis helps to analyze and understand the various frequency components of signals, which is also suitable for images. Wavelet transformation overcomes the shortage of short-time Fourier transformation and is widely deployed in time-frequency analysis. The wavelet transformation utilizes wavelet functions to decompose images from different directions, obtaining high-frequency and low-frequency components of each direction.

Anbarjafari *et al.* [26] proposed an SR method that generates HR images by inverse wavelet transformation from LR images combined with the high-frequency sub-bands, which are obtained by wavelet transformation and then interpolated. The method proposed in [27] obtains HR images via inverse wavelet transformation from high-frequency sub-bands decomposed from LR images by two types of wavelets.

In [28], [29], the SR quality was improved by means of introducing edge prior into the prediction of high-frequency sub-bands. Additionally, Guo *et al.* [30] enhanced the details of images by combining wavelet transformation with CNN, learning mappings between LR and HR images in the wavelet domain. It is, however, quite inappropriate to migrate these methods of one-band-based fashion into HSI SR missions, on account of inevitable spectral distortion.

B. SR METHODS BASED ON CNN

As an effective operation to extract local information, convolution operation is capable of obtaining different feature representations by changing the convolutional kernel, which is trainable in neural network. Therefore, CNN is widely utilized in both image processing and natural language processing.

Dong *et al.* [31] proposed an SR convolutional neural network (SRCNN) for the first time. SRCNN is composed of only three convolutional layers, in which the outputs of the first two convolutional layers pass through nonlinear activation layers of rectified linear unit (ReLU). The LR image was amplified by bicubic interpolation and then fed to convolutional layers to fit the corresponding HR image. Dong *et al.* [32] then proposed FSRCNN, in which the complex bicubic interpolation was replaced by the deconvolutional layer, reducing the computational complexity of calculation to some extent. Shi *et al.* [33] proposed ESPCN, which proposed the sub-pixel convolutional operation for the first time. The sub-pixel convolutional operation amplifies the image scale by rearrange the multi-channels in order, reducing the computational complexity compared with the deconvolutional operation on the premise of ensuring the accuracy. Kim *et al.* [34] proposed VDSR with deeper network layers. Based on the architecture of VGG-Net [35], VDSR achieved a great depth and was able to learn local and global representations. The residual network (ResNet) [36] had outstanding performance not only in image classification tasks, but in SR tasks as well. ResNet learned the high-frequency portion, i.e., residuals, between the input image and the real image. Lim *et al.* [37] proposed EDSR based on single-scale amplification and MDSR based on multi-scale amplification. The main architecture of the two networks are roughly the same. These two networks adopt the strategy of sharing the parameters of the residual module to reduce the volume of parameters. Additionally, in order to increase the robustness of the model, residual proportional coefficient is added to the network. Lai *et al.* [38] proposed LapSRN based on Laplacian pyramid architecture, which is able to obtain SR images of three scales at one time. After extracting features from LR images, each level of the network utilizes deconvolution as amplification.

III. METHODOLOGY

In this section, we present a novel three-dimensional convolutional network architecture combined with wavelet transformation, i.e., FS-3DCNN. The network is composed of two

main parts, namely, feature extraction subnet and prediction subnet. For the purpose of reducing the noise introduced in the preprocessing, the strategy is adopted in FS-3DCNN that LR HSI is decomposed by wavelet transformation before it is fed to convolutional layers, and the network predicts the wavelet coefficients of HR HSI, reconstructing the HR HSI via the inverse wavelet transformation. We first introduce the basis of wavelet transformation and three-dimensional convolutional neural network, and then present the details of FS-3DCNN, including the architecture and loss function.

A. TWO-DIMENSIONAL WAVELET TRANSFORMATION

The two-dimensional wavelet transformation is able to decompose by the horizontal and vertical directions of images by a certain wavelet function, obtaining high-frequency and low-frequency components of each direction. Four components totally are obtained, namely, one approximate component of the low-frequency, two high-frequency components in the horizontal and vertical directions, and the highest frequency component in the diagonal direction, as shown in Figure 1.

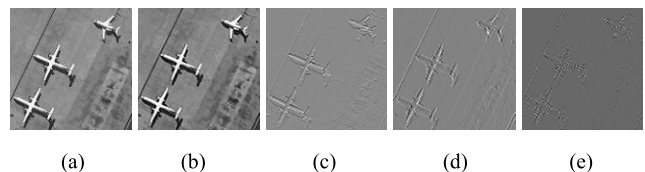


FIGURE 1. Schematic diagram of wavelet transform. The (a) original picture is decomposed into (b) approximation, (c) horizontal detail, (d) vertical detail, and (e) diagonal detail.

Given the scale function $\varphi(\cdot)$ and the wavelet function $\psi(\cdot)$, the approximate coefficient and the detail coefficient of the two-dimensional wavelet transformation are capable to be written as:

$$W_{\varphi}(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \varphi_{j_0, m, n}(x, y)$$

$$W_{\psi}^i(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j_0, m, n}^i(x, y),$$

$$i = \{H, V, D\} \quad (1)$$

where $f(x, y)$, $W_{\varphi}(j_0, m, n)$, and $W_{\psi}^i(j_0, m, n)$ present the image with the size of (M, N) , approximate coefficients and detail coefficients in three directions of level j_0 , respectively. Then the image is capable of being reconstructed by the coefficients:

$$f(x, y) = \frac{1}{\sqrt{MN}} \sum_m \sum_n W_{\varphi}(j_0, m, n) \varphi_{j_0, m, n}(x, y)$$

$$+ \frac{1}{\sqrt{MN}} \sum_{i=H, V, D} \sum_{j=j_0}^{\infty} \sum_m \sum_n$$

$$\times W_{\psi}^i(j, m, n) \psi_{j, m, n}^i(x, y) \quad (2)$$

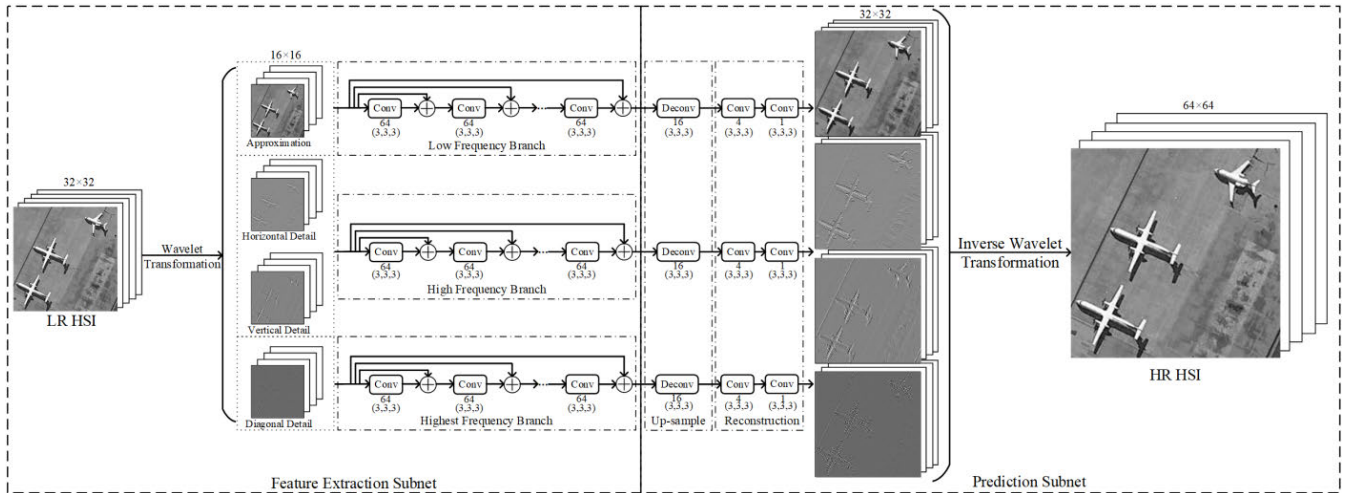


FIGURE 2. Flowchart of FS-3DCNN at the upscaling factor of two. In the flowchart, the spatial sizes of each image, filter depths, and filter sizes are presented, whereas the spectral dimension depends on the real datasets. At the factor of four, one more deconvolution layer with the same settings is added after the existing deconvolution layer. Besides, the spatial sizes of LR HSI and the four corresponding components at the factor of four are half of the former.

This process of decomposition and reconstruction following the one-to-one mapping principle, is linearly reversible, introducing minimal error. Besides, the spatial sizes of the coefficients are half of the size of the original image.

B. THREE-DIMENSIONAL CONVOLUTION

In the early stage, many HSI SR methods based on neural network are the migration and improved from related technologies in the field of computer vision. Ignoring the differences between natural image and HSI, these methods usually cause serious spectral distortion. To suppress the spectral distortion, some methods deploy additional strategies such as dictionary learning [39] and non-negative matrix decomposition [40]. However, it is more effective to utilize three-dimensional (3D) convolution to extract spatial-spectrum features of HSI, which is able to reduce probability of occurrence of spectral distortion instead of suppressing after occurrence. The output of the k -th feature cube in the d -th layer following formulation in [41] can be written as:

$$F_{(x,y,z)}^{d,k} = h \left(b^{d,k} + \sum_c \sum_{u=0}^{U-1} \sum_{v=0}^{V-1} \sum_{w=0}^{W-1} \omega_{(u,v,w)}^{d,k,c} F_{(x+u,y+v,z+w)}^{d-1,c} \right) \tag{3}$$

where c indicates feature cubes of the $(d-1)$ -th layer connecting to the k -th feature cube of the d -th layer, $\omega_{(u,v,w)}^{d,k,c}$ indicates the convolutional result at (u, v, w) with the k -th feature cube with the size of $U \times V \times W$, $F_{(x,y,z)}^{d,k}$ indicates convolutional result at (x, y, z) with the k -th feature cube of the d -th layer, $h(\cdot)$ indicates a non-linear activation function, e.g., ReLU and Hyperbolic Tangent. Therefore, 3D convolution is capable to extract spatial-spectral features directly [42], [43], reducing the occurrence of spectral distortion.

C. ARCHITECTURE OF FS-3DCNN

HSI is constituted of a variety of different frequency components. Theoretically, super-resolution tasks for different components are different. Thus, dealing these components in one particular network, which is utilized in most methods, can be regarded as a multi-task learning. Compared with dealing the components separately, more parameters and more complex network architecture are needed because the complexity of task is increased [24]. Then, the computation complexity and volume of the network will be increased, the performance will be limited in the same conditions of computation, and the difficulty and uncertainty of training will be greatly increased. Therefore, HSI is decomposed before it is fed to convolutional layers in FS-3DCNN, then extracted via 3D convolutional layers to obtain representations of different components specifically, effectively suppressing spectral distortion. The flowchart of the network is shown in Figure 2, and the number of layers will be introduced in Section IV.

Firstly, the two-dimensional wavelet transformation is applied to LR HSI, and the introduced noise and spectral distortion are small enough to be negligible. The four obtained components are divided into three groups according to frequency similarity, namely, {Approximation}, {Horizontal detail, Vertical detail}, and {Diagonal detail}, which are exploited as the inputs of one low-frequency branch and two high-frequency branches of FS-3DCNN, respectively.

Secondly, based on the existing researches of CNN, we found that convolution not only extracts neighborhood information, but also leads to the diffusion of information, and the higher the frequency is, the severer the diffusion of information is. As the times of convolutions, i.e., the depth of convolutional layers, increases, the information energy spreads out over a larger range. However, if convolution is performed multiple times with a smaller convolutional kernel,

the energy of information will be more concentrated than the saturation with a bigger convolutional kernel. In SR tasks, high-frequency information is of great significance. Therefore, the two high-frequency branches of FS-3DCNN exploit less convolutional layers with small convolutional kernels, whereas in the low-frequency branch, considering the size of actual samples, more convolutional layers with small convolutional kernels are utilized to obtain better representations of low-frequency features.

In order to further protect the evanescent high-frequency information, a large number of skip connections are added in each branch to transmit the original high-frequency information with rich details directly to the deep layers, which also reduces the gradient disappearance and increases the robustness of the network.

Thirdly, after the joint feature representations of spatial spectrum of different frequency components are obtained by three branches of FS-3DCNN, up-sampling and detail reconstruction are carried out. Since bicubic interpolation simply exploits spatial information, ignoring the interaction between spatial and spectral information, there is great chance to cause spectral distortion while up-sampling, let along the unobtrusive performance. Therefore, FS-3DCNN utilizes three-dimensional deconvolution for up-sampling. In the meantime, two three-dimensional convolutional layers are deployed in the tail of each branch to fine-tune both spatial and spectral details. Finally, the predicted HR wavelet coefficients are obtained through three branches of FS-3DCNN, corresponding to four wavelet coefficients of HR HSI. Finally, HR HSI is obtained via inverse wavelet transformation.

We believe that the novel network architecture of FS-3DCNN takes the advantages as follows. Firstly, the complex learning task of mappings between LR HSI and HR HSI is decomposed into four relatively simple tasks, that is, the frequency components of each sub-band of wavelet coefficients is relatively simple, which simplifies the overall architecture of the network, reduces the quantity of network parameters, decreases difficulty of training, and boosts converging. Secondly, since the wavelet transformation is a linear, reversible and lossless decomposition, the spatial and spectral and errors in the domains in the processing phase can be considered negligible, improving the accuracy of the network. Thirdly, the network is designed to extract and utilize the spatial-spectral joint features of HSI, which can inhibit the generation of spectral distortion before occurrence, contributing to certain advantages in spectral accuracy.

D. TRAINING OF FS-3DCNN

In the forward propagation, the network learns the mappings between LR wavelet coefficients and corresponding HR wavelet coefficients. The input LR HSI of FS-3DCNN, represented by $X_l \in \mathbb{R}^{2r \times 2r \times L}$, is then decomposed into wavelet coefficients $C_A, C_H, C_V, C_D \in \mathbb{R}^{r \times r \times L}$, where $2r$ and L indicate the length of spatial side and spectral bands,

respectively. Outputs of each branch are as follows:

$$\{m_1, m_3, m_3 \dots m_k \dots m_d\}, \quad k = 1, 2, 3 \dots d$$

$$\{m_k\} = F_j (C_i \oplus F_{j-1} (C_i \oplus F_{j-2} (\dots))),$$

$$i \in \{A, H, V, D\} \quad (4)$$

where m_k indicates the k -th feature cube with the same size as the wavelet coefficients, d indicates the depth of filters and differs in different branches, \oplus indicates merging of two matrixes, and F_j indicates the j -th convolutional layer. Then the predicted wavelet coefficients are as follows:

$$\hat{C}_V, \hat{C}_H, \hat{C}_V, \hat{C}_D \in \mathbb{R}^{Sr \times Sr \times L}$$

$$\hat{C}_i = F^{R2} \left(F^{R1} \left(F^{De} (C_i \oplus m_d) \right) \right), \quad i \in \{A, H, V, D\} \quad (5)$$

where S indicates the upscaling factor, $\hat{C}_V, \hat{C}_H, \hat{C}_V, \hat{C}_D$ indicate the predicted wavelet coefficients, F^{R1} and F^{R2} indicate the detail reconstruction, and F^{De} indicates the 3D deconvolution, related to the upscaling factor S . Ultimately, the predicted HR HSI $\hat{Y}_h \in \mathbb{R}^{2Sr \times 2Sr \times L}$ is obtained by inverse wavelet transformation, as follows:

$$\hat{Y}_h = IWT \left(\hat{C}_V, \hat{C}_H, \hat{C}_V, \hat{C}_D \right) \quad (6)$$

where $IWT(\cdot)$ indicates the inverse wavelet transformation.

In back propagation, all trainable parameters of the network are computed and iteratively optimized by Charbonnier loss function, which is a variant of l_1 -norm, strictly convex and infinitely differentiable. As Lai *et al.* proved in [44], Charbonnier loss function is capable of robustness, better handling outliers, and improving the performance over the l_2 -norm loss functions. The overall loss of the network can be written as:

$$Loss \left(\hat{Y}, Y; \theta \right) = \frac{1}{N} \sum_{i=1}^N \rho \left(\hat{Y}_s^{(i)} - Y_s^{(i)} \right)$$

$$\rho(x) = \sqrt{x^2 + \varepsilon^2} \quad (7)$$

where N indicates the batch size, θ indicates the current parameters, and $\rho(\cdot)$ indicates Charbonnier penalty function. Empirically, ε is settled to $1e - 3$.

IV. EXPERIMENT RESULTS

In this section, in order to prove the effectiveness and extensibility of the proposed method, we experiment on four real HSI datasets produced by four different HSI imagers, and compare FS-3DCNN with the state-of-art methods, including methods from computer vision and remote sensing. Secondly, we conduct comparative experiments on the upscaling factors of two and four to illustrate the performance of FS-3DCNN at different magnification scales. Thirdly, the experiment covers multiple cases, including cases with small sample size to prove the extensibility and robustness of the model.

A. EXPERIMENT SETTINGS

Firstly, we shortly review the real HSI datasets utilized: Pavia University scene (PaviaU), Kennedy Space Center (KSC),

TABLE 1. Experimental sampling and batch settings.

Dataset	Training size	Testing size	Validating size	Total size	Batch size	Step per epoch
Botswana	74	9	9	92	4	19
Chikusei	980	122	122	1224	14	70
KSC	58	7	7	72	4	15
PaviaU	37	4	4	45	2	19

Botswana, and one newer dataset Chikusei Hyperspectral Data [45] (Chikusei). The PaviaU dataset is collected by the German space-based reflective spectral imager ROSIS-03, with a spectral range from 0.43 μm to 0.86 μm , spatial resolution of 1.3 m, spatial size of (610, 340), and spectral bands of 103 totally. The KSC dataset is collected by NASA AVIRIS, with spectral bands from 0.4 μm to 2.5 μm , spatial resolution of 18 m, spatial size of (512, 614), and spectral bands of 176 totally. The Botswana dataset is collected by NASA EO-1, with a spectral range from 0.4 μm to 2.5 μm , spatial size of (1476, 256), spatial resolution of 30 m, and spectral bands of 145 totally. The Chikusei dataset is collected by the Headwall Hyperspec-VNIR-C HSI imager, with a spectral range from 0.363 μm to 1.018 μm , spatial resolution of 2.5 m, spatial size of (2517, 2335), and spectral bands of 128 totally. These four datasets cover different HSI imagers, different ground distribution characteristics and different spatial and spectral resolutions.

Secondly, considering that the spatial sizes of wavelet coefficients are half of the LR HSI, besides, the experiment upscaling factors are two and four, the spatial size of each sample, therefore, should be settled to a certain size. Additionally, the number of samples should be kept to a certain extent to ensure the performance of the model as well. Considering the above factors, spatial size of HR HSI samples is set to (64, 64). Besides, Bicubic interpolation and Gaussian down-sampling are widely accepted as the methods of down-sampling. As discussed in [9], [46], the Gaussian down-sampling simulates the real situation better, FS-3DCNN uses a Gaussian filter for down-sampling, with a Gaussian kernel of zero mean and standard deviation of 0.8493 when the magnification is (2, 2), and when the magnification is (4, 4), the Gaussian kernel with zero mean and standard deviation of 1.6986 is used instead of the interpolation sampling method used in other studies. The above parameters are also provided in [9], [46].

After browsing the datasets, the useless parts of the datasets are removed, e.g., the black edges of zero padding, which reduces the noise that might be introduced. A linear function is utilized to compress the data within the range of [0,1), facilitating the training of the neural network, which can be written as:

$$I_{processed} = \frac{10 \times \lg(I_{raw} + 1)}{\lceil \max(\log_{10}(I_{raw} + 1) \times 10) \rceil} \quad (8)$$

where I_{raw} and $I_{processed}$ indicate the raw and compressed HSI, and $\lceil \cdot \rceil$ indicates ceiling. Because the HSI data can vary

from tens to thousands, the purpose of utilizing this logarithm trick of processing is to compress the hyperspectral data into a certain range while avoiding too low values to vanish in the training phase, which is capable of protecting high-frequency details and reducing training error as well. At the meantime, the restoration of data is also simpler compared with commonly used normalization. The noise introduced by the function can be considered negligible. Therefore, the HSI is able to be easily restored backwards by the function.

Thirdly, in terms of parameter settings, the numbers of 3D convolutional layers in the feature extraction subnet are 7, 5, and 3 in order, which are determined by a set of enumeration experiments. We believe that the more complex the frequency components of the branch are, the more convolution layers are demanded to extract features. The kernel size and filter depth of each 3D convolutional layer are (3, 3, 3) and 64. And the stride is (1, 1, 1). Each layer is padding with zeros to maintain the same shape. And ReLU is chosen as the activation function of each layer.

In the prediction subnet, the up-sampling module consists of 3D deconvolutional layers according to the upscaling factor, that is, when the upscaling factors are two and four, the numbers of convolutional layers are 1 and 2, respectively. The kernel size and stride of these 3D deconvolutional kernels are (3, 3, 3) and (2, 2, 1). And filter depth is 16. The reconstruction module consists of two 3D convolutional layers with kernel size of (3, 3, 3), stride of (1, 1, 1), filter depths of 4 and 1 in order, activation function of ReLU, and padding with zeros. In addition, in order to sufficiently shuffle the sample pairs at the beginning of each epoch, and learn the entire training set, the batch and step per epoch settings are presented in Table 1.

Before each experiment, the sample pairs are fully shuffled, with 10% drawn out for verification, 10% for testing, and the rest for training. The sampling and batch settings of the datasets are shown in Table 1. It can be inferred from Table 1 that the experiment covers both large and small sample sizes.

Finally, in order to iterate sufficiently and learn the training set sufficiently, the training epoch is settled to 800, and an early-stop strategy is utilized, which will stop training when the performance on the validation set does not improve. Before each experiment, sample pairs are randomly grouped. A total of 10 experiments on each dataset are conducted, and the final result is obtained by averaging the results of the testing sets.

TABLE 2. Experimental results at an upscaling factor of two.

Dataset	Indices	Bicubic	SRCNN	VDSR	3D-FCNN	FS-3DCNN (ours)
Botswana	PSNR (dB)	26.0969	28.7964	34.4028	37.1935	38.6145
	SSIM	0.6463	0.7986	0.9515	0.9456	0.9684
	SAM	4.4586°	4.0406°	3.3821°	3.1284°	3.0133°
Chikusei	PSNR (dB)	28.1341	31.5769	36.5989	38.0173	38.5655
	SSIM	0.7563	0.9235	0.9396	0.9678	0.9759
	SAM	4.0575°	3.6151°	3.1191°	3.0027°	2.9600°
KSC	PSNR (dB)	26.1866	28.2687	32.0116	33.7073	34.0422
	SSIM	0.6746	0.8347	0.8979	0.9110	0.9275
	SAM	4.5407°	4.2062°	3.7144°	3.5276°	3.4929°
PaviaU	PSNR (dB)	26.4534	28.3378	35.4788	34.0004	36.7992
	SSIM	0.7282	0.8771	0.9183	0.8936	0.9510
	SAM	4.5279°	4.2268°	3.3760°	3.5228°	3.2549°

TABLE 3. Experimental results at an upscaling factor of four.

Dataset	Indices	Bicubic	SRCNN	VDSR	3D-FCNN	FS-3DCNN (ours)
Botswana	PSNR (dB)	24.0422	25.3255	30.5455	31.0108	36.3517
	SSIM	0.4805	0.5237	0.8793	0.9078	0.9369
	SAM	5.2532°	5.2241°	4.2903°	3.7306°	3.4001°
Chikusei	PSNR (dB)	23.8683	24.9855	32.3083	33.4324	34.9351
	SSIM	0.6112	0.6719	0.8310	0.8933	0.9191
	SAM	5.6374°	5.7741°	4.0025°	3.8827°	3.6072°
KSC	PSNR (dB)	24.6429	25.8057	29.0800	30.2054	30.9391
	SSIM	0.5701	0.6418	0.7668	0.8121	0.8490
	SAM	5.1273°	5.0475°	4.5011°	4.3929°	4.2286°
PaviaU	PSNR (dB)	22.7032	23.8266	30.9226	31.6029	33.1933
	SSIM	0.4900	0.6153	0.7786	0.8091	0.8708
	SAM	6.1473°	5.9788°	4.4442°	4.2954°	4.0005°

B. COMPARED WITH THE STATE-OF-ART METHODS

The method proposed in this paper is compared with several state-of-art SR methods, including computer vision and remote sensing SR methods, i.e., Bicubic, SRCNN [47], VDSR [34], and 3D-FCNN [23]. SRCNN, VDSR, and 3D-FCNN followed the default settings as described in [23], [34], [47]. In order to ensure the fairness of the experiment, the three neural networks and FS-3DCNN are trained on the same datasets, and stopped until the indicator of the validating set does not improve.

Evaluation indicators of the experiment are peak-signal-noise-ratio (PSNR, dB), structural similarity index measurement (SSIM) [48], and spectral angle mapper (SAM), evaluating image quality, image similarity, and vector similarity, respectively. PSNR and SSIM take the average of indicators on each band of HSI, whereas SAM takes the average of indicators on each pixel of HSI. The experimental results at an upscaling factor of two is presented in Table 2.

It can be found that our FS-3DCNN is superior to other methods in evaluation indicators. VDSR performs well in SR quality of natural images, but introduces too much spectral distortion, resulting in relatively poor performance on HSI. Compared with SRCNN and Bicubic, because of its three-dimensional convolutional layer, 3D-FCNN can directly extract the spatial spectral joint information of HSI, it performs well in both SR quality and suppressing spectral distortion. However, compared with our method, 3D-FCNN is less capable of SR quality. On account of the novel network architecture and data processing fashion, FS-3DCNN has outstanding performance in three evaluation indices, which means that FS-3DCNN has outstanding performance not only in spatial accuracy, but also in spectral accuracy. To explore the robustness of our method, we also carry out experiments at an upscaling factor of four, and the experimental results are shown in Table 3.

It could be inferred from Table 3 that the overall trend remains the same, although the performance in this case

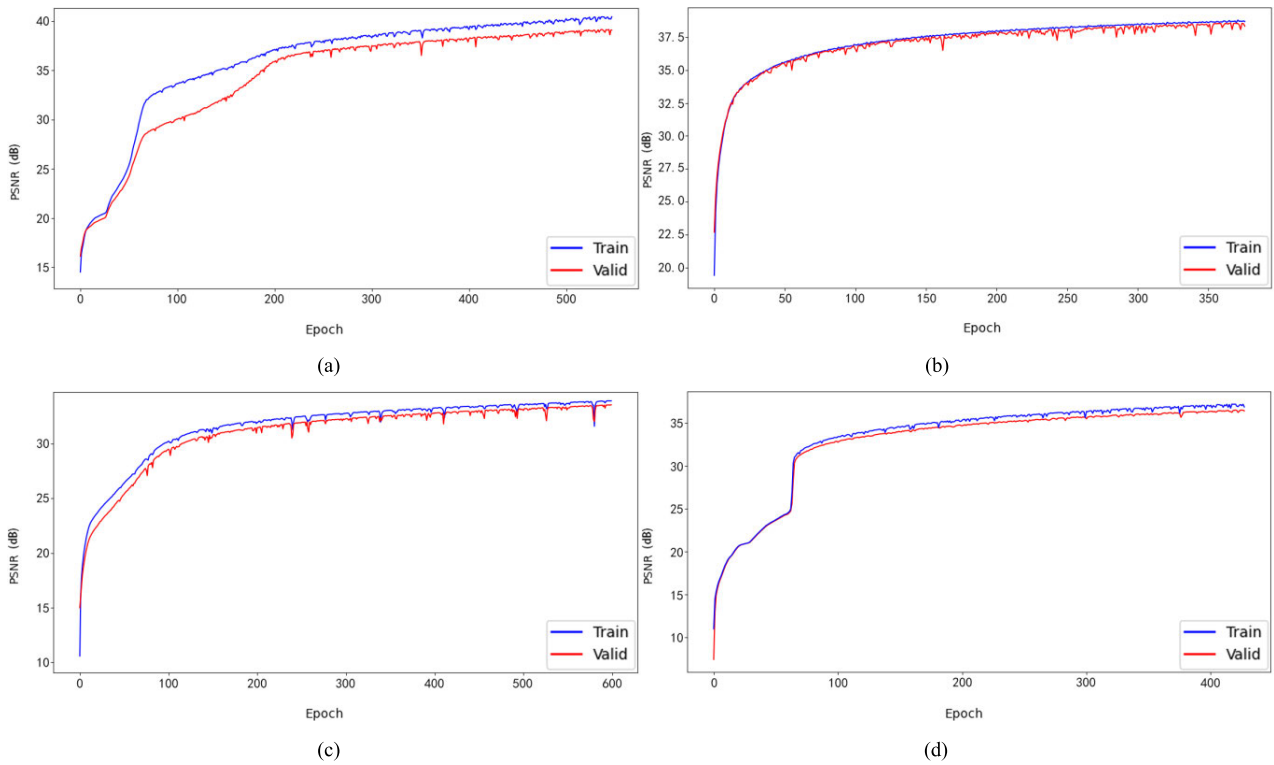


FIGURE 3. Change of PSNR with epoch during one training period at an upscaling factor of two on each dataset. (a) Botswana, (b) Chikusei, (c) KSC, and (d) PaviaU.

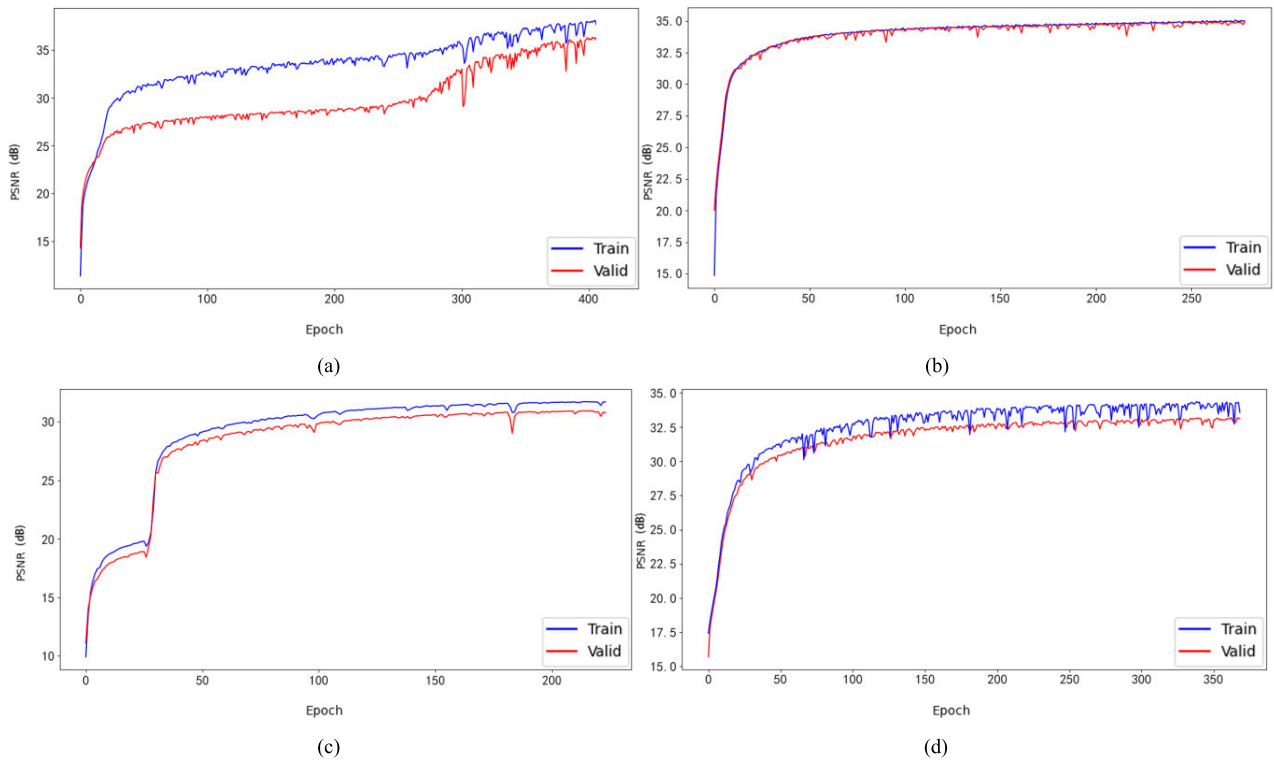


FIGURE 4. Change of PSNR with epoch during one training period but by an upscaling factor of four on each dataset. (a) Botswana, (b) Chikusei, (c) KSC, and (d) PaviaU.

decreases slightly. In summary, we can give the point that our method has a certain improvement in SR quality and spectral accuracy. Additionally, we plot the change of PSNR

metrics with epoch during one training period of FS-3DCNN at upscaling factors of two and four on each dataset, as shown in Figure 3 and Figure 4. Obviously, in the training process of

FS-3DCNN, it is trapped in local optimal conditions rarely, and can jump out after dozens of epochs. We believe it is because of the characteristics of Carboniferous loss function.

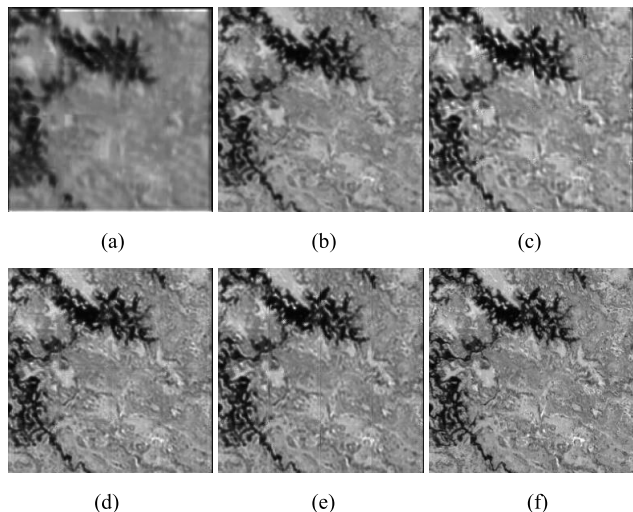


FIGURE 5. SR results on Botswana at an upscaling factor of two, i.e., (a) Bicubic, (b) SRCNN, (c) VDSR, (d) 3D-FCNN, (e) FS-3DCNN, and (f) original HR HSI.

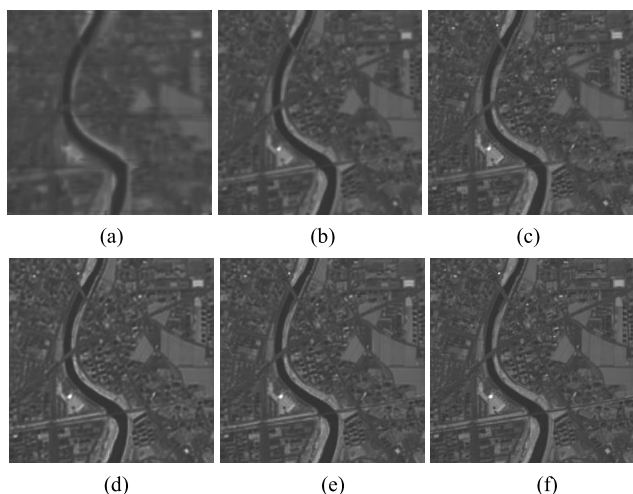


FIGURE 6. SR results on Chikusei at an upscaling factor of two, i.e., (a) Bicubic, (b) SRCNN, (c) VDSR, (d) 3D-FCNN, (e) FS-3DCNN, and (f) original HR HSI.

As shown in Figure 5 to Figure 12, we select a certain band from SR results, namely, the 75th band of Botswana, the 80th band of Chikusei, the 85th band of KSC, and the 20th band of PaviaU, and present a typical region with spatial size of (256, 256), in order to provide intuitive performances of each method at upscaling factors of two and four.

It can be found from the above figures that our method is the closest to the original HR HSI at both upscaling factors.

Table 4 presents the average training time of SRCNN, VDSR, 3D-FCNN, and FS-3DCNN on four datasets. It could be inferred from Table 4 that parameter optimization of FS-3DCNN requires more iterations and time than the others.

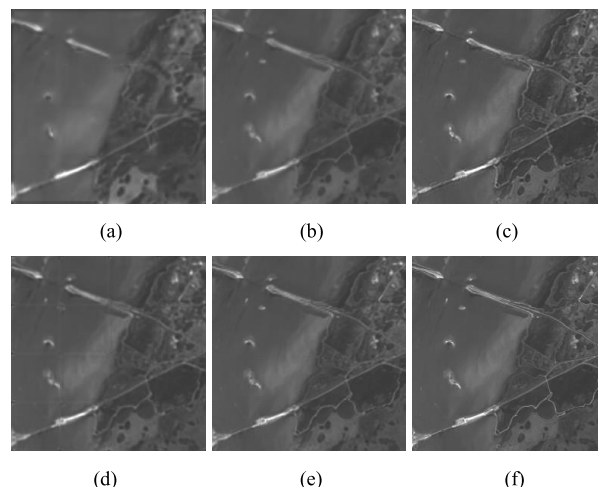


FIGURE 7. SR results on KSC at an upscaling factor of two, i.e., (a) Bicubic, (b) SRCNN, (c) VDSR, (d) 3D-FCNN, (e) FS-3DCNN, and (f) original HR HSI.

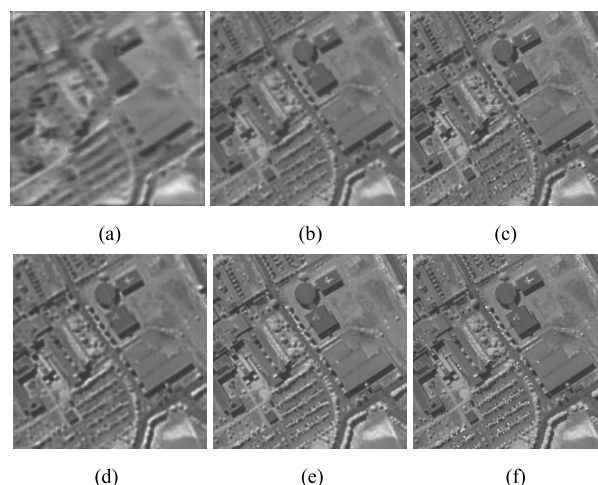


FIGURE 8. SR results on PaviaU at an upscaling factor of two, i.e., (a) Bicubic, (b) SRCNN, (c) VDSR, (d) 3D-FCNN, (e) FS-3DCNN, and (f) original HR HSI.

TABLE 4. Average training time of CNN based methods in seconds on four datasets at upscaling factors of two and four.

Method	SRCNN	VDSR	3D-FCNN	FS-3DCNN
Time at upscaling factor of 2 (second)				
Botswana	8.4	154.6	713.2	6072.2
Chikusei	59.4	565.0	1586.8	13714.0
KSC	14.5	254.0	611.4	4060.8
PaviaU	8.7	61.4	625.4	2857.9
Time at upscaling factor of 4 (second)				
Botswana	8.3	49.2	478.4	2109.7
Chikusei	44.9	375.5	956.5	4049.2
KSC	12.7	130.7	491.3	1085.3
PaviaU	8.8	57.6	461.8	1123.6

V. ANALYSIS AND DISCUSSIONS

A. SENSITIVITY ANALYSIS ON PARAMETERS

In order to explore the influence of the convolutional kernel size, and filter depth on the performance of FS-3DCNN,

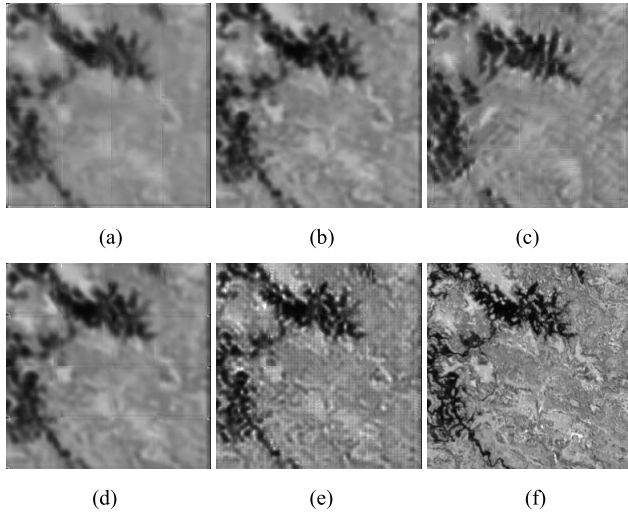


FIGURE 9. SR results on Botswana at an upscaling factor of four, i.e., (a) Bicubic, (b) SRCNN, (c) VDSR, (d) 3D-FCNN, (e) FS-3DCNN, and (f) original HR HSI.

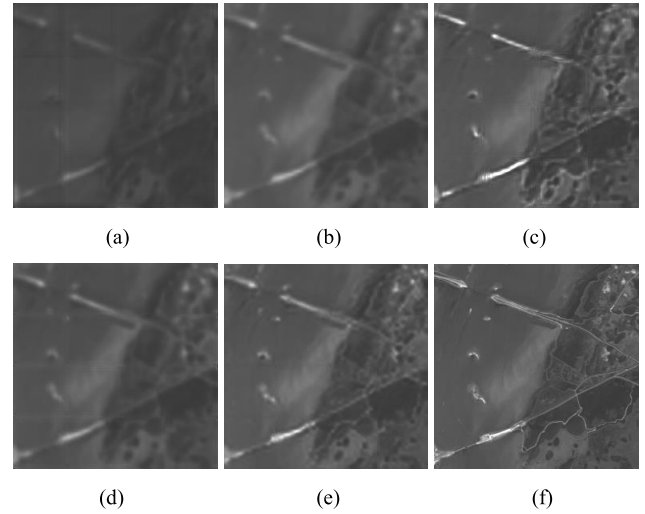


FIGURE 11. SR results on KSC at an upscaling factor of four, i.e., (a) Bicubic, (b) SRCNN, (c) VDSR, (d) 3D-FCNN, (e) FS-3DCNN, and (f) original HR HSI.

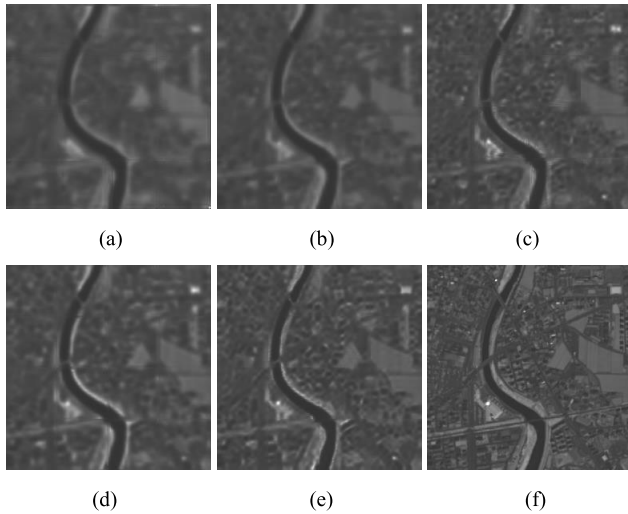


FIGURE 10. SR results on Chikusei at an upscaling factor of four, i.e., (a) Bicubic, (b) SRCNN, (c) VDSR, (d) 3D-FCNN, (e) FS-3DCNN, and (f) original HR HSI.

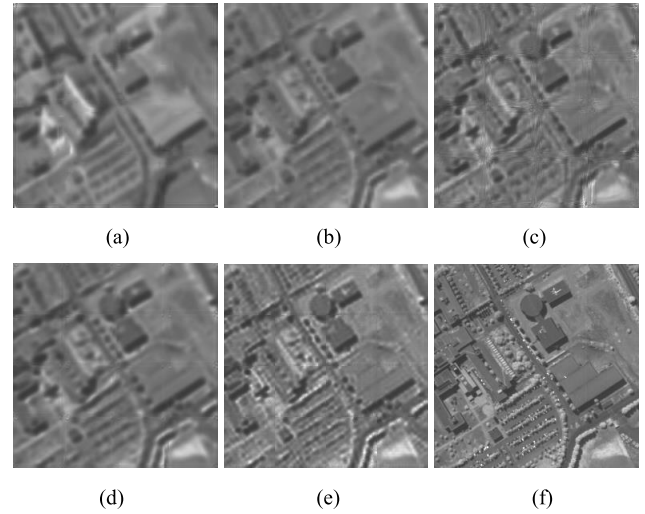


FIGURE 12. SR results on PaviaU at an upscaling factor of four, i.e., (a) Bicubic, (b) SRCNN, (c) VDSR, (d) 3D-FCNN, (e) FS-3DCNN, and (f) original HR HSI.

we give sensitivity analysis of FS-3DCNN over the network parameters. Table 5 presents PSNR of FS-3DCNN on four datasets in three conditions of convolutional kernels: (1, 1, 1), (3, 3, 3), and (5, 5, 5).

It can be inferred from Table 5 that the performance of the method is best when the size of the convolutional kernel is (3, 3, 3). This result confirms our analysis of convolutional characteristics above, that is, the energy of convoluted information is more concentrated, which is more conducive to the extraction and utilization of high-frequency information. In addition, it limits the model performance when the kernel size is too small.

Table 6 shows the PSNR performance of FS-3DCNN on the four datasets when the filter depths of the low-frequency branch are 32, 48, 64, 80, and 96, respectively. We found

that the performance is the highest when the filter depth is 64. However, when the filter depth continues to increase, the performance decreased slightly. We believe that the filter depth can affect the capacity of the model. However, when the capacity reaches a certain threshold, it will hardly improve the performance by increasing the filter depth continuously, and might even lead to the performance degradation.

B. ROBUSTNESS OVER WAVELET FUNCTIONS

In order to explore the possible impact of different wavelet functions on model performance, we compare three wavelet functions, namely Haar wavelet, Debaucheries wavelet (db2), and Symlets wavelet (sym2), whose support lengths are 1, 3, and 3, respectively. Table 7 shows the PSNR performances

TABLE 5. PSNR performance of FS-3DCNN on four datasets at different convolutional kernels.

Kernel size	(1, 1, 1)	(3, 3, 3)	(5, 5, 5)
PSNR at upscaling factor of 2 (dB)			
Botswana	26.1347	38.6145	29.2789
Chikusei	30.8277	38.5655	29.8009
KSC	30.0237	34.0422	28.2519
PaviaU	22.3428	36.7992	21.6414
PSNR at upscaling factor of 4 (dB)			
Botswana	27.7475	36.3517	29.1077
Chikusei	27.1276	34.9351	26.5856
KSC	22.1129	30.9391	24.3323
PaviaU	30.2351	33.1933	25.9076

TABLE 6. PSNR performance of FS-3DCNN on four datasets at different filter depths.

Filter depth	32	48	64	80	96
PSNR at upscaling factor of 2 (dB)					
Botswana	34.5203	36.6386	38.6145	37.6550	37.6430
Chikusei	34.6062	36.6237	38.5655	37.6009	37.5694
KSC	31.5128	32.5270	34.0422	33.5692	32.5375
PaviaU	34.5728	34.6692	36.7992	35.1718	34.6445
PSNR at upscaling factor of 4 (dB)					
Botswana	32.1581	34.2467	36.3517	35.8922	35.3802
Chikusei	30.9359	32.9657	34.9351	33.9705	33.9390
KSC	26.8824	28.8959	30.9391	30.4661	29.4344
PaviaU	30.8678	31.4320	33.1933	32.3307	32.2531

TABLE 7. PSNR performance of FS-3DCNN in three wavelet functions.

Wavelet function	Haar	db2	sym2
PSNR at upscaling factor of 2 (dB)			
Botswana	38.6145	38.6340	38.6430
Chikusei	38.5655	38.5723	38.5874
KSC	34.0422	34.0415	34.0739
PaviaU	36.7992	36.8315	36.7977
PSNR at upscaling factor of 4 (dB)			
Botswana	36.3517	36.3649	36.3962
Chikusei	34.9351	34.9632	34.9338
KSC	30.9391	30.9435	30.9602
PaviaU	33.1933	33.2248	33.1892

on four data sets and three wavelet functions. It can be considered that the slight differences of experiment results on each data are within the error range. Therefore, a reasonable choice of wavelet function will not affect the performance of the model fundamentally.

The experimental platform of this paper is as follows: Debian 10 operating system, Tensor Flow 2.1 [49] environment with python3.7, 16G RAM, NVIDIA GTX 1080Ti graphic card, with an SSD as storage device.

VI. CONCLUSIONS

In this paper, a novel network for HSI SR, i.e., FS-3DCNN, is proposed. Firstly, compared with the network which requires feature embedding before feature extraction, this processing method effectively reduces the noise introduced in the spatial domain and spectrum domain, increases the controllability of the network, and reduces the difficulty of training. Secondly, the network exploits 3D convolution and 3D deconvolution to extract and utilize the joint features of spatial-spectral information directly, effectively suppressing the occurrence of spectral distortion. Thirdly, each branch utilizes a different number of convolutional layers according to the frequency characteristics of the components, and adds a large number of skip connections to transmit the high-frequency information to the deep layers, increasing the robustness of the network and reducing the possibility of gradient disappearance. The outstanding performance of FS-3DCNN has been proved by experiments on the condition of different HSI imagers, different data sizes, and different resolutions of space and spectrum. However, wavelet transform still faces its own limitations, that is, Heisenberg uncertainty principle. How to break through this limitation is the future work in this field. In addition, we expect that the combination of analysis and neural network methods will be a major focus in this field and will lead to the improvement of SR performance.

REFERENCES

- [1] N. Yokoya, J. Chan, and K. Segl, "Potential of resolution-enhanced hyperspectral data for mineral mapping using simulated EnMAP and Sentinel-2 images," *Remote Sens.*, vol. 8, no. 3, p. 172, 2016.
- [2] A. F. H. Goetz, G. Vane, J. E. Solomon, and B. N. Rock, "Imaging spectrometry for earth remote sensing," *Science*, vol. 228, no. 4704, pp. 1147–1153, 1985.
- [3] L. Loncan, L. B. de Almeida, J. M. Bioucas-Dias, X. Briottet, J. Chanussot, N. Dobigeon, S. Fabre, W. Liao, G. A. Licciardi, M. Simoes, J.-Y. Tourneret, M. A. Veganzones, G. Vivone, Q. Wei, and N. Yokoya, "Hyperspectral pansharpening: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 27–46, Sep. 2015.
- [4] M. Dalla Mura, G. Vivone, R. Restaino, P. Addesso, and J. Chanussot, "Global and local gram-Schmidt methods for hyperspectral pansharpening," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*. Milan, Italy: Springer, Jul. 2015, pp. 37–40.
- [5] H. R. Shahdoosti and H. Ghassemian, "Combining the spectral PCA and spatial PCA fusion methods by an optimal filter," *Inf. Fusion*, vol. 27, pp. 150–160, Jan. 2016.
- [6] S. Yang, K. Zhang, and M. Wang, "Learning low-rank decomposition for pan-sharpening with spatial-spectral offsets," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 8, pp. 3647–3657, Aug. 2018.
- [7] G. Masi, D. Cozzolino, L. Verdoliva, and G. Scarpa, "Pansharpening by convolutional neural networks," *Remote Sens.*, vol. 8, no. 7, p. 594, 2016.
- [8] J. Yang, X. Fu, Y. Hu, Y. Huang, X. Ding, and J. Paisley, "PanNet: A deep network architecture for pan-sharpening," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*. Venice, Italy: Springer, Oct. 2017, pp. 5449–5457.
- [9] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multi-spectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.
- [10] X. X. Zhu, C. Grohnfeldt, and R. Bamler, "Exploiting joint sparsity for pansharpening: The J-SparseFI algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 5, pp. 2664–2681, May 2016.
- [11] N. Akhtar, F. Shafait, and A. Mian, "Sparse spatio-spectral representation for hyperspectral image super-resolution," in *Proc. Eur. Conf. Comput. Vis.* Zurich, Switzerland: Springer, 2014, pp. 63–78.

- [12] Q. Wei, J. Bioucas-Dias, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral and multispectral image fusion based on a sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3658–3668, Jul. 2015.
- [13] M. Simoes, J. Bioucas-Dias, L. B. Almeida, and J. Chanussot, "A convex formulation for hyperspectral image superresolution via subspace-based regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3373–3388, Jun. 2015.
- [14] L. Zhang, W. Wei, C. Bai, Y. Gao, and Y. Zhang, "Exploiting clustering manifold structure for hyperspectral imagery super-resolution," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5969–5982, Dec. 2018.
- [15] J. Yang, Y.-Q. Zhao, and J. Chan, "Hyperspectral and multispectral image fusion via deep two-branches convolutional neural network," *Remote Sens.*, vol. 10, no. 5, p. 800, 2018.
- [16] Q. Xie, M. Zhou, Q. Zhao, D. Meng, W. Zuo, and Z. Xu, "Multispectral and hyperspectral image fusion by MS/HS fusion net," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Los Angeles, USA: Springer, Jun. 2019, pp. 1585–1594.
- [17] Y. Zhao, J. Yang, and J. C.-W. Chan, "Hyperspectral imagery super-resolution by Spatial-Spectral joint nonlocal similarity," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2671–2679, Jun. 2014.
- [18] J. Li, Q. Yuan, H. Shen, X. Meng, and L. Zhang, "Hyperspectral image super-resolution by spectral mixture analysis and Spatial-Spectral group sparsity," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 9, pp. 1250–1254, Sep. 2016.
- [19] Y. Wang, X. Chen, Z. Han, and S. He, "Hyperspectral image super-resolution via nonlocal low-rank tensor approximation and total variation regularization," *Remote Sens.*, vol. 9, no. 12, p. 1286, 2017.
- [20] Y. Yuan, X. Zheng, and X. Lu, "Hyperspectral image superresolution by transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 5, pp. 1963–1974, May 2017.
- [21] Y. Li, J. Hu, X. Zhao, W. Xie, and J. Li, "Hyperspectral image super-resolution using deep convolutional neural network," *Neurocomputing*, vol. 266, pp. 29–41, Nov. 2017.
- [22] 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, Oct. 2017.
- [23] 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.
- [24] Y. Chen, H. Fan, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, Y. Shuicheng, and J. Feng, "Drop an octave: Reducing spatial redundancy in convolutional neural networks with octave convolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*. Seoul, South Korea: Springer, Oct. 2019, pp. 3435–3444.
- [25] J. Yang, Y. Q. Zhao, J. C. W. Chan, and L. Xiao, "A multi-scale wavelet 3D-CNN for hyperspectral image super-resolution," *Remote Sens.*, vol. 11, no. 13, p. 1557, 2019.
- [26] G. Anbarjafari and H. Demirel, "Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image," *ETRI J.*, vol. 32, no. 3, pp. 390–394, Jun. 2010.
- [27] H. Demirel and G. Anbarjafari, "IMAGE resolution enhancement by using discrete and stationary wavelet decomposition," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1458–1460, May 2011.
- [28] H. Chavez-Roman and V. Ponomaryov, "Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1777–1781, Oct. 2014.
- [29] H. Demirel and G. Anbarjafari, "Discrete wavelet transform-based satellite image resolution enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 6, pp. 1997–2004, Jun. 2011.
- [30] T. Guo, H. S. Mousavi, T. H. Vu, and V. Monga, "Deep wavelet prediction for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*. Honolulu, HI, USA: IEEE, Jul. 2017, pp. 104–113.
- [31] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
- [32] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland: Springer, 2014, pp. 184–199.
- [33] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Las Vegas, NV, USA: Springer, Jun. 2016, pp. 1874–1883.
- [34] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Las Vegas, NV, USA: Springer, Jun. 2016, pp. 1646–1654.
- [35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Las Vegas, NV, USA: Springer, Jun. 2016, pp. 770–778.
- [37] 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. Workshops (CVPRW)*. Honolulu, HI, USA: Springer, Jul. 2017, pp. 136–144.
- [38] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA: Springer, Jul. 2017, pp. 624–632.
- [39] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3467–3478, Aug. 2012.
- [40] W. Dong, F. Fu, G. Shi, X. Cao, J. Wu, G. Li, and X. Li, "Hyperspectral image super-resolution via non-negative structured sparse representation," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2337–2352, May 2016.
- [41] S. Ji, W. Xu, M. Yang, and K. Yu, "3D convolutional neural networks for human action recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221–231, Jan. 2013.
- [42] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, p. 67, 2017.
- [43] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.
- [44] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," *IEEE Trans. Pattern. Anal.*, vol. 41, no. 11, pp. 2599–2613, Sep. 2018.
- [45] N. Yokoya and A. Iwasaki, "Airborne hyperspectral data over Chikusei," Space Appl. Lab., Univ. Tokyo, Tokyo, Japan, Tech. Rep. SAL-2016-05-27, 2016.
- [46] N. Yokoya, T. Yairi, and A. Iwasaki, "Coupled nonnegative matrix factorization unmixing for hyperspectral and multispectral data fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 528–537, Feb. 2012.
- [47] K. Umehara, J. Ota, N. Ishimaru, S. Ohno, K. Okamoto, T. Suzuki, N. Shirai, and T. Ishida, "Super-resolution convolutional neural network for the improvement of the image quality of magnified images in chest radiographs," in *Proc. Med. Imag., Image Process.*, vol. 10133. Orlando, FL, USA: SPIE, Feb. 2017, Art. no. 101331.
- [48] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [49] M. Abadi, "Tensorflow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Oper. Syst. Des. Implement. (OSDI)*, Savannah, GA, USA: USENIX, 2016, pp. 265–283.



LIGU WANG (Member, IEEE) received the M.S. and Ph.D. degrees in signal and information processing from the Harbin Institute of Technology, Harbin, China, in 2002 and 2005, respectively.

From 2006 to 2008, he was a Postdoctoral Researcher with the Department of Information and Communications Engineering, Harbin Engineering University, Harbin. He is currently a Professor with Harbin Engineering University and a Distinguished Professor with Dalian Nationalities University. He has authored one book named *Hyperspectral Image Processing* and more than 180 articles in journals and conference proceedings. His current research interests include remote sensing and machine learning.



TIANYI BI received the B.S. degree from the Department of Electronic Information Engineering, Beihang University, Beijing, China, and the M.S. degree from the Department of Information and Communications Engineering, Harbin Engineering University, Harbin, China. His current research interests include neural networks, computer vision, and natural language processing.



YAO SHI received the B.S. and M.S. degrees from the Department of Information and Communications Engineering, Harbin Engineering University, Harbin, China, where she is currently pursuing the Ph.D. degree. Her current research interests include machine learning and remote sensing imagery processing.

• • •